



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

**Análise da evasão em cursos superiores: o caso da
evasão em cursos superiores da área de Computação**

Raphael Magalhães Hoed

Dissertação apresentada como requisito parcial para conclusão do
Mestrado Profissional em Computação Aplicada

Orientador
Prof. Dr. Marcelo Ladeira

Brasília
2016

Ficha catalográfica elaborada automaticamente,
com os dados fornecidos pelo(a) autor(a)

HH693a Hoed, Raphael Magalhães
Análise da evasão em cursos superiores: o caso da
evasão em cursos superiores da área de Computação /
Raphael Magalhães Hoed; orientador Marcelo Ladeira.
- Brasília, 2016.
188 p.

Dissertação (Mestrado - Mestrado Profissional em
Computação Aplicada) -- Universidade de Brasília, 2016.

1. Evasão. 2. Computação. 3. Análise de
Sobrevivência. 4. Mineração de Dados. I. Ladeira,
Marcelo, orient. II. Título.



Universidade de Brasília

Instituto de Ciências Exatas
Departamento de Ciência da Computação

Análise da evasão em cursos superiores: o caso da evasão em cursos superiores da área de Computação

Raphael Magalhães Hoed

Dissertação apresentada como requisito parcial para conclusão do
Mestrado Profissional em Computação Aplicada

Prof. Dr. Marcelo Ladeira (Orientador)
CIC/UnB

Prof. Dr. Paulo Jorge Leitão Adeodato Prof. Dr. Sérgio Antonio Andrade de Freitas
CIn/UFPE FGA/UnB

Prof. Dr.^a Maria de Fátima Ramos Brandão
CIC/UnB

Prof. Dr. Marcelo Ladeira
Coordenador do Programa de Pós-graduação em Computação Aplicada

Brasília, 16 de Dezembro de 2016

Dedicatória

Dedico este trabalho primeiramente a Deus, por concessão da saúde, fé e perseverança indispensáveis à condução dos meus estudos. Ao meus pais, que sempre me apoiaram incondicionalmente e me ensinaram desde cedo o valor da educação. A minha namorada, pela paciência e compreensão nos momentos em que estive ausente, pelo carinho irrestrito nos momentos de cansaço e hesitação. A minha tia Arnaldina Hoed, por ter sido fundamental em minha criação e por ter tão entusiasticamente participado da minha alfabetização.

Agradecimentos

A Deus por ter me dado saúde e força para superar as dificuldades.

Ao meu orientador Marcelo Ladeira, pelo suporte irrestrito, no pouco tempo que lhe coube, pelas suas correções e incentivos, pela sua paciência, amizade e por sempre me estimular a buscar o aprimoramento em qualquer coisa que me propus a fazer.

Aos familiares e amigos que sempre me apoiaram, deram conselhos e me estimularam nessa jornada.

Aos colegas de mestrado da turma de 2014, pelo companheirismo, união no desenvolvimento dos trabalhos e pela ótima recepção desde o início do curso.

A todos os professores do mestrado que sem dúvidas nenhuma contribuíram para a minha formação.

Às professoras Maria de Fátima Ramos Brandão, Letícia Lopes Leite e Cibele Queiroz da Silva que me deram valiosos conselhos e orientações para aprimoramento do meu trabalho.

À Secretaria de Assuntos Acadêmicos da UnB que rapidamente e gentilmente forneceu dados valiosos para condução das minhas pesquisas.

Resumo

Esta dissertação apresenta um estudo sobre evasão nos cursos de graduação da área de Computação, com base em dados fornecidos pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) e com base em dados fornecidos pela Universidade de Brasília (UnB). Nos estudos quantitativos realizados, são obtidas as taxas anuais de evasão e aplicada a técnica estatística de Análise de Sobrevivência e mineração de regras de associação via algoritmo Apriori. É apresentada também uma análise qualitativa a partir de questionários aplicados a alunos evadidos em cursos superiores de Computação da UnB para verificação das causas de evasão. Os resultados obtidos com os estudos dos dados do INEP são apresentados por instituições públicas e privadas. São comparadas, inicialmente, as evasões nas grandes áreas do conhecimento de acordo com a classificação adotada pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE). Após a comparação da evasão nessas grandes áreas, é feito um detalhamento para a grande área de Ciências, Matemática e Computação e posteriormente para a área de Computação. Um estudo de caso das causas da evasão em cursos da área de Computação foi conduzido na UnB, onde foram estudados os cursos de Bacharelado em Ciência da Computação, Licenciatura em Computação, Engenharia de Software e Engenharia de Computação para levantamento de causas de evasão nesses cursos. Foram obtidas evidências de que a relação candidatos/vagas é inversamente proporcional à evasão, de que os cursos da grande área de Ciências, Matemática e Computação que requerem maior uso de conhecimentos matemáticos e de abstração algorítmica possuem maiores taxas de evasão e de que o sexo, a forma de ingresso na instituição e ser ou não cotista afetam as taxas de evasões nos cursos na área de Computação.

Palavras-chave: Evasão, Computação, Análise de Sobrevivência, Mineração de Dados

Abstract

This dissertation presents a study on dropout in undergraduate courses in Computing, based on data provided by the Brazilian National Institute for Educational Research Anísio Teixeira (INEP) and based on data provided by the University of Brasília (UnB). In the quantitative studies carried out, the annual rates of dropout are obtained and the statistical technique of Survival Analysis and association rules mining using Apriori algorithm are applied to the data. It is also presented a qualitative analysis from quiz applied to students that have dropped out in UnB Computing higher courses to verify the causes of school dropout. The results obtained with INEP data studies are separated by public and private institutions. First, dropout in large areas of knowledge are compared according to the classification adopted by the Organization for Economic Co-operation and Development (OECD). After the comparison of the dropout in the large areas, a detailing is made for the large area of Sciences, Mathematics and Computation and later for the area of Computation. A case study of the causes of dropout in courses in the Computing area was conducted at UnB in order to survey the causes of dropout on the courses of Bachelor in Computer Science, Degree in Computer Science, Software Engineering and Computer Engineering. Evidence was obtained that the candidate/vacancy ratio is inversely proportional to dropout, that the courses in the large area of Science, Mathematics and Computation that require greater use of mathematical knowledge and algorithmic abstraction present higher evasion rates and that gender, the form of entry into the institution and quotas for students affect the dropout rates in Computing courses.

Keywords: Dropout, Computation, Survival analysis, Data Mining

Sumário

1	Introdução	1
1.1	Definição do Problema	1
1.2	Justificativa	2
1.3	Contribuição Esperada	3
1.4	Hipóteses a Investigar	4
1.5	Objetivos	5
2	Fundamentação Teórica e Revisão do Estado da Arte	6
2.1	Conceitos e Cálculos de Evasão Escolar	6
2.1.1	Os Conceitos de Evasão	6
2.1.2	Calculando a Evasão	8
2.2	Evasão Escolar no Ensino Superior	13
2.2.1	Causas da Evasão nos Cursos Superiores	13
2.2.2	Causas da Evasão nos Cursos de Computação	20
2.3	Técnicas Estatísticas	26
2.3.1	Box-plot	26
2.3.2	Análise de Sobrevivência	29
2.4	Mineração de Dados	34
2.4.1	Conceito e Tarefas de Mineração de Dados	34
2.4.2	Metodologias de Mineração de Dados	35
2.4.3	Mineração de Regras de Associação - Algoritmo Apriori	36
2.5	Classificação dos Cursos Superiores Adotada pelo INEP	41
3	Metodologia	43
3.1	Tratamento e Análise dos Dados do INEP	43
3.2	Tratamento e Análise dos Dados da UnB	47
4	Estudos Realizados	50
4.1	Estudo dos Microdados do INEP	50
4.1.1	Evasão e Concorrência	50

4.1.2	Análise de Sobrevivência	65
4.2	Estudo dos Dados da UnB	75
4.2.1	Análise de Sobrevivência	76
4.2.2	Mineração de Regras de Associação	83
4.2.3	Causas de Evasão	88
5	Conclusões	106
	Referências	112
	Apêndice	120
A	Taxas de Evasões Anuais	121
B	Respostas dos Evadidos - Pesquisa de Evasão	125
B.1	Outros Motivos para ter Evadido	125
B.1.1	Ciência da Computação	125
B.1.2	Licenciatura em Computação	127
B.1.3	Engenharia de Computação	130
B.1.4	Engenharia de Software	134
B.2	Motivos para Escolha do Curso de Computação	135
B.2.1	Ciência da Computação	135
B.2.2	Licenciatura em Computação	140
B.2.3	Engenharia de Computação	145
B.2.4	Engenharia de Software	148
B.3	Disciplinas Consideradas mais Difíceis	150
B.3.1	Ciência da Computação	150
B.3.2	Licenciatura em Computação	154
B.3.3	Engenharia de Computação	158
B.3.4	Engenharia de Software	161
C	Artigo Apresentado na 4ª CIACA - Lisboa, Portugal	164

Lista de Figuras

2.1	Exemplo de Box-plot	27
2.2	Estrutura da caixa de um Box-plot	28
2.3	Fases do modelo de Referência CRISP-DM	36
4.1	Evasão nas grandes áreas do conhecimento (instituições privadas)	51
4.2	Evasão nas grandes áreas do conhecimento (instituições públicas)	52
4.3	Evasão nas grandes áreas do conhecimento (instituto Lobo)	53
4.4	Evasão vs Concorrência (instituições privadas)	54
4.5	Evasão vs Concorrência (instituições públicas)	55
4.6	Evasão por sexo nas grandes áreas do conhecimento (instituições privadas)	57
4.7	Evasão por sexo nas grandes áreas do conhecimento (instituições públicas)	58
4.8	Evasão na grande área de Ciências, Matemática e Computação (instituições privadas)	59
4.9	Evasão na grande área de Ciências, Matemática e Computação (instituições públicas)	60
4.10	Evasão nos cursos de Computação (instituições privadas)	62
4.11	Evasão nos cursos de Computação (instituições públicas)	63
4.12	Evasão por sexo nos cursos de Computação (instituições privadas)	64
4.13	Evasão por sexo nos cursos de Computação (instituições públicas)	65
4.14	Curvas de sobrevivência (sexo) - instituições privadas	66
4.15	Curvas de sobrevivência (sexo) - instituições públicas	67
4.16	Curvas de sobrevivência (cotista) - instituições privadas	68
4.17	Curvas de sobrevivência (cotista) - instituições públicas	69
4.18	Curvas de sobrevivência (ingresso) - instituições privadas	70
4.19	Curvas de sobrevivência (ingresso) - instituições públicas	71
4.20	Curvas de sobrevivência (pós-graduação em Computação) - instituições privadas	72
4.21	Curvas de sobrevivência (pós-graduação em Computação) - instituições públicas	73
4.22	Curvas de sobrevivência (sexo) - UnB	76

4.23	Curvas de sobrevivência (cotista) - UnB	78
4.24	Curvas de sobrevivência (ingresso) - UnB	79
4.25	Curvas de sobrevivência (disciplina Cálculo 1) - UnB	80
4.26	Curvas de sobrevivência (disciplinas de algoritmos) - UnB	81
4.27	Disciplinas mais complexas segundo os alunos consultados	104

Lista de Tabelas

2.1	Conceitos de evasão e suas amplitudes	7
2.2	Fórmulas para cálculo da evasão	9
2.3	Taxas de evasão no curso de Formação de Prof. da Educação Básica	12
2.4	Causas de evasão em cursos superiores	16
2.5	Média salarial nos cursos de Computação	23
2.6	Reprovações por disciplina - Licenciatura em Informática (UFMT)	24
2.7	Causas de evasão em cursos de Computação	25
2.8	Evasão anual para cursos da universidade X - dados fictícios	27
2.9	Categorias das técnicas de Análise de Sobrevivência	31
2.10	Etapas de execução do algoritmo Apriori	40
2.11	Grande área de Ciências, Matemática e Computação	42
3.1	Variáveis usadas nos estudos com dados do INEP	45
3.2	Variáveis usadas nos estudos com dados da UnB	48
4.1	Teste de correlação - evasão vs concorrência (instituições privadas)	56
4.2	Teste de correlação - evasão vs concorrência (instituições públicas)	56
4.3	Evasão na Computação em relação à media nacional (instituições privadas)	61
4.4	Evasão na Computação em relação à media nacional (instituições públicas)	61
4.5	Estatísticas - curvas de sobrevivência (sexo) - instituições privadas	66
4.6	Estatísticas - curvas de sobrevivência (sexo) - instituições públicas	67
4.7	Estatísticas - curvas de sobrevivência (cotista) - instituições privadas	68
4.8	Estatísticas - curvas de sobrevivência (cotista) - instituições públicas	69
4.9	Estatísticas - curvas de sobrevivência (ingresso) - instituições privadas	70
4.10	Estatísticas - curvas de sobrevivência (ingresso) - instituições públicas	71
4.11	Estatísticas - curvas de sobrevivência (pós-graduação em Computação) - instituições privadas	72
4.12	Estatísticas - curvas de sobrevivência (pós-graduação em Computação) - instituições públicas	74
4.13	Análise multivariada da sobrevivência (instituições privadas)	74

4.14	Análise multivariada da sobrevivência (instituições públicas)	75
4.15	Estatísticas - curvas de sobrevivência (sexo) - UnB	77
4.16	Estatísticas - curvas de sobrevivência (cotista) - UnB	78
4.17	Estatísticas - curvas de sobrevivência (ingresso) - UnB	79
4.18	Estatísticas - curvas de sobrevivência (disciplina Cálculo 1) - UnB	80
4.19	Estatísticas - curvas de sobrevivência (disciplinas de algoritmos) - UnB	82
4.20	Análise multivariada da sobrevivência (UnB)	83
4.21	Regras de associação (cursos de Computação da UnB)	86
4.22	Dimensionamento da amostra	88
4.23	Dados demográficos da amostra (% de respondentes)	90
4.24	Fatores relacionados à evasão	92
4.25	Alfa de Cronbach para cada classificação usada na pesquisa	93
4.26	Teste de Kruskal Wallis (Ciência da Computação)	94
4.27	Itens do maior subconjunto homogêneo (Ciências da Computação)	95
4.28	Teste de Kruskal Wallis (Licenciatura em Computação)	95
4.29	Itens do maior subconjunto homogêneo (Licenciatura em Computação)	96
4.30	Teste de Kruskal Wallis (Engenharia de Computação)	97
4.31	Itens do maior subconjunto homogêneo (Engenharia de Computação)	98
4.32	Teste de Kruskal Wallis (Engenharia de Software)	99
4.33	Itens do maior subconjunto homogêneo (Engenharia de Software)	100
4.34	Porcentagem das respostas conforme classificação adotada	101
4.35	Itens acrescentados pelos alunos como causas de evasão	102
5.1	Fatores relacionados à evasão nos cursos superiores de Computação	106
A.1	Evasão anual e concorrência nas grandes áreas do conhecimento	121
A.2	Evasão anual média nas grandes áreas do conhecimento (2001 a 2005)	122
A.3	Evasão anual por sexo nas grandes áreas do conhecimento (instituições privadas)	122
A.4	Evasão anual por sexo nas grandes áreas do conhecimento (instituições públicas)	123
A.5	Evasão anual na grande área de Ciências, Matemática e Computação	123
A.6	Evasão anual nos cursos de Computação	123
A.7	Evasão anual por sexo nos cursos de Computação (instituições privadas)	124
A.8	Evasão anual por sexo nos cursos de Computação (instituições públicas)	124

Lista de Abreviaturas e Siglas

- ANDIFES** Associação Nacional dos Dirigentes das Instituições Federais de Ensino Superior. 17
- CAPES** Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. 46, 47
- CRg** cobertura da regra mais geral. 85
- CRISP-DM** Cross Industry Standard Process for Data Mining. 35, 43
- EKM** Estimador de Kaplan-Meier. 32
- ENEM** Exame Nacional do Ensino Médio. 17, 70, 146
- FURG** Universidade Federal do Rio Grande. 17
- INEP** Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. viii, xii, 5–7, 10–13, 30, 41, 43–46, 50, 51, 58–61
- KDD** Knowledge Discovery in Databases. 34, 35
- MEC** Ministério da Educação. 13, 18, 19
- OCDE** Organização para a Cooperação e Desenvolvimento Econômico. 9, 41, 51
- PAS** Programa de Avaliação Seriada. 19, 48, 79, 83, 87
- PROUNI** Programa Universidade para Todos. 18, 70
- PUCRS** Pontifícia Universidade Católica do Rio Grande do Sul. 109
- REUNI** Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais. 16, 17
- SAA** Secretaria de Assuntos Acadêmicos. 47

SBC Sociedade Brasileira de Computação. 61

SEMESP Sindicato das Entidades Mantenedoras de Estabelecimentos de Ensino Superior no Estado de São Paulo. 1

SEMMA Sample, Explore, Modify, Model, Assessment. 35

SINE Sistema Nacional de Emprego. 23

SISU Sistema de Seleção Unificada. 17–19, 33, 48, 70, 72, 80

TCG Taxa de Conclusão de Graduação. 10

UEFS Universidade Estadual de Feira de Santana. 17

UERJ Universidade Estadual do Rio de Janeiro. 19

UFCG Universidade Federal de Campina Grande. 17

UFG Universidade Federal de Goiás. 19

UFMG Universidade Federal de Minas Gerais. 9

UFMS Universidade Federal de Mato Grosso do Sul. 18

UFMT Universidade Federal de Mato Grosso. 22, 24

UFPB Universidade Federal da Paraíba. 10

UFPE Universidade Federal de Pernambuco. 110

UFPR Universidade Federal do Paraná. 108

UFRGS Universidade Federal do Rio Grande do Sul. 37

UFRN Universidade Federal do Rio Grande do Norte. 17

UFSM Universidade Federal de Santa Maria. 17

UnB Universidade de Brasília. viii–xiii, 4, 5, 10, 17, 19, 30, 43, 47–50, 75–83, 86, 91, 104, 106–108, 110, 131, 134, 135

UNIFESP Universidade Federal de São Paulo. 10

UNIPAMPA Universidade Federal do Pampa. 17

UNIVASF Universidade Federal do Vale do São Francisco. 17

USP Universidade de São Paulo. 3

UTFPR Universidade Tecnológica Federal do Paraná. 18

Capítulo 1

Introdução

Neste Capítulo são apresentados a definição do problema, a justificativa para o estudo da evasão nos cursos superiores de Computação, a contribuição esperada desse trabalho, as hipóteses a serem investigadas e os objetivos geral e específicos.

1.1 Definição do Problema

A evasão de alunos nos cursos de graduação no Brasil é um problema já conhecido e de incidência histórica, que é explorado nesse trabalho. Costumeiramente, temos notícias sobre colações de grau em que o número de concluintes é consideravelmente menor do que o número de matriculados inicialmente no curso.

Em se tratando dos cursos superiores da área de Computação, existem números no Brasil e em outros países que apontam para taxas elevadas de evasão. Segundo levantamento realizado em 2012 pelo Sindicato das Entidades Mantenedoras de Estabelecimentos de Ensino Superior no Estado de São Paulo (SEMESP) [29], “a cada três alunos que entram no curso de sistemas de informação, apenas um recebe o diploma. Em Ciência da Computação, a cada quatro alunos que entram no curso, apenas um termina”. Na Irlanda, analisando-se os anos letivos de 2007/08 a 2008/09, verificou-se que os cursos de Ciência da Computação experimentaram a mais alta taxa de evasão, com 27% desses alunos saindo ainda no primeiro ano [67]. Também na Irlanda, analisando-se os anos letivos de 2012/13 a 2013/14, verificou-se que as áreas de Construção e Afins, Serviços, Ciência da Computação e Engenharia apresentaram as maiores taxas de evasão [62]. Foi verificado que os cursos da área de Ciência da Computação da Universidade de Tecnologia de Helsinki na Finlândia, tem de 500 a 600 inscrições anuais e as taxas de evasão tem variado entre 30 e 50% [55]. De acordo com Muratet et al.¹ (2009 apud Barcelos e Silveira,

¹Muratet, Mathieu et al. Towards a serious game to help students learn computer programming. International Journal of Computer Games Technology, v. 2009, p. 3, 2009.

2012) [9] “no período de 2005 a 2009 houve uma redução de cerca de 20% no número de estudantes de Ciência da Computação na França”. Crenshaw et al.² (2008, apud Barcelos e Silveira, 2012) [9] “reportam queda de 50% no interesse de estudantes americanos pela área ao longo dos anos 2000”.

A intenção do discente de evadir, muitas vezes, é descoberta tardiamente quando essa já é irreversível, após a saída do aluno ou quando esse já assumiu uma forte intenção de deixar o curso ou a instituição. Os gestores também podem ser mais capazes de lidar com a evasão discente, conhecendo melhor as causas que levam à evasão para assim tomar medidas mais eficazes no sentido de mitigar o problema. Os funcionários, principalmente aqueles que trabalham com gestão acadêmica, devem estar cientes dos fatores relacionados à evasão, para que se construa uma política unificada dentro da universidade e para que se tenha uma visão mais abrangente, sob diferentes pontos de vista, do problema da evasão. Sendo assim, não apenas professores mas também os seguintes profissionais da educação devem ser conscientizados da importância de se combater o problema: reitor, vice-reitor, pró-reitores, diretor, vice-diretor e coordenadores de cursos. Em um âmbito macro, as unidades acadêmicas, os departamentos, os núcleos de assistência e as pró-reitorias devem trabalhar se aproximando dos professores e dos alunos e contribuindo para a aproximação entre todas as partes que compoem a comunidade acadêmica. No presente trabalho, ao se investigar e procurar uma maior compreensão sobre o fenômeno de evasão nos cursos superiores de Computação, foram buscadas evidências para os seguintes questionamentos:

- Existe uma correlação entre taxas de evasões e a relação candidatos/vagas dos cursos nas grandes áreas do conhecimento?
- Cursos de graduação que requerem maior capacidade de abstração, conhecimento matemático e/ou físico possuem maiores níveis de evasão?
- Quais as áreas do conhecimento possuem cursos de graduação mais predispostos à evasão dos alunos?
- Quais são algumas causas de evasão em Computação citadas por alunos evadidos?

1.2 Justificativa

A realidade da evasão nos cursos de graduação no Brasil já foi elencada por diversos autores, conforme é descrito no Capítulo 2. O número de alunos matriculados contribui significativamente para compor a matriz orçamentária de uma instituição. Além de ser

²Crenshaw, Tanya L.; Chambers, Erin Wolf; Metcalf, Heather. A case study of retention practices at the University of Illinois at Urbana-Champaign. ACM SIGCSE Bulletin, v. 40, n. 1, p. 412-416, 2008.

uma obrigação de uma instituição de ensino buscar mecanismos que favoreçam a permanência do aluno na escola, a evasão também gera um impacto orçamentário. Nas instituições públicas, por exemplo, compreender esse fenômeno pode contribuir significativamente no sentido de diminuir o desperdício de recursos. Nas instituições privadas, há uma clara relação entre evasão e perda de receitas.

Além da perda de recursos decorrentes da evasão escolar, temos os problemas inerentes às frustrações dos discentes, que muitas vezes tem seus sonhos interrompidos ou completamente podados diante do abandono do curso. Quando um aluno ingressa em um curso, são investidos recursos financeiros, tempo é destinado para as atividades de graduação e é difundida a ideia de que aqueles que possuem curso superior são mais bem remunerados. “Assim, para o discente, começar e não terminar um curso de graduação pode acarretar uma frustração profissional que o acompanhará por toda a vida” [21]. De acordo com Tinto³ (1997 apud Lima Júnior, 2013) [60], temos os seguintes compromissos institucionais:

- (1) As instituições devem estar comprometidas com o bem-estar dos estudantes acima de quaisquer outras necessidades institucionais.
- (2) As instituições devem estar comprometidas em primeiro lugar com a educação de todos e não somente de alguns.
- (3) As instituições devem estar comprometidas com o desenvolvimento de comunidades sociais e educacionais capazes de dar suporte aos alunos e nas quais todos possam ser integrados como membros integrais.

Um estudo feito na Universidade de São Paulo (USP), revelou que, para o ano de 2010, a área de Ciências Biológicas mostrou-se a mais cara da Universidade (R\$ 45.4 Mil), seguida das áreas de Ciências Exatas (R\$ 29.5 Mil) e de Ciências Humanas (R\$ 16.4 Mil), isso considerando o custo per capita durante o ano [54]. Ressalta-se que a evasão na USP no período de 1998 a 2008 nas áreas de Ciências Exatas foi a mais alta se comparada com as demais áreas, sendo de 79,88% [54].

Desta forma, além de ser uma obrigação institucional zelar pela aprendizagem e pelo bem estar dos discentes, é também necessário conter a evasão pois isso está diretamente associado à perda de recursos financeiros.

1.3 Contribuição Esperada

A análise da evasão nos cursos superiores do Brasil é importante na busca de subsídios para se compreender esse fenômeno, visando mitigá-lo. Em um âmbito mais abrangente,

³Tinto, Vicent. Classrooms as communities: Exploring the educational character of student persistence. *The Journal of Higher Education*, v. 68, n. 6, pp. 599-623, 1997.

permite entender melhor a evasão nos cursos superiores da área de Computação e identificar aspectos que possam ser trabalhados no sentido de evitar que alunos em risco de evadir o façam. Este estudo também poderá beneficiar diretamente a UnB, favorecendo subsídios para a adoção de estratégias mais efetivas no sentido de mitigar a desistência de alunos em cursos de Computação ofertados pela instituição.

Em relação à melhoria da gestão escolar, ao se conhecer melhor o fenômeno da evasão, pode-se adequar os serviços educacionais às aspirações e peculiaridades dos discentes. Os núcleos de assistência estudantil, as coordenações e os professores poderão atuar preventivamente. Outra possibilidade, a partir do conhecimento prévio dos fatores que envolvem o problema, é qualificar melhor os professores, visando a melhoria da metodologia de ensino e o estreitamento dos laços entre os professores e alunos.

A abordagem usada nesse trabalho, utilizando as técnicas estatísticas de Análise de Sobrevivência e a mineração de regras de associação com o algoritmo Apriori, podem servir de inspiração para outros estudos envolvendo evasão discente, inclusive em áreas diferentes de Computação. Esta abordagem enfatiza o estudo da evasão levando-se em consideração o aspecto temporal (quando a evasão ocorre) e quais variáveis estão associadas a uma maior evasão, o que possibilita a obtenção de informações valiosas para lidar com o problema nos momentos mais oportunos e levando-se em consideração os fatores mais críticos. É importante ressaltar que nesse estudo não foi avaliado o impacto do desempenho do aluno sobre a evasão e nem da retenção negativa (permanência prolongada do estudante) no curso.

1.4 Hipóteses a Investigar

As hipóteses a seguir são investigadas ao longo desse trabalho:

1. A relação candidatos/vagas é inversamente proporcional à evasão, ou seja, há uma tendência de que cursos superiores mais concorridos tenham taxas de desistência inferiores aos cursos superiores com menor concorrência.
2. Os cursos da grande área de Ciências, Matemática e Computação que requerem maior uso de conhecimentos matemáticos e de abstração algorítmica possuem maiores taxas de evasão.
3. O sexo, a forma de ingresso na instituição e ser ou não cotista afetam as taxas de evasões nos cursos da área de Computação.

1.5 Objetivos

O objetivo geral deste trabalho consiste em:

Analisar a evasão nos cursos superiores do Brasil, com base em dados fornecidos pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP), e identificar causas de evasão nos cursos superiores de Computação, com base em um estudo de caso feito na UnB.

Os objetivos específicos consistem em:

- Verificar se existe correlação entre as taxas de evasões e a relação candidatos/vagas dos cursos nas grandes áreas do conhecimento.
- Comparar a evasão entre as grandes áreas do conhecimento.
- Comparar a evasão entre as áreas que compoem a grande área de Ciências, Matemática e Computação.
- Verificar para os cursos da área de Computação, se as taxas de evasão variam em função do sexo do aluno, da forma de ingresso na instituição, do fato de ser ou não ser cotista e do fato da instituição ofertar ou não curso de pós-graduação em Computação.
- Comparar a evasão entre os cursos de Computação.
- Realizar um estudo de caso da evasão nos cursos de Computação da UnB, visando identificar causas de evasão.

Nos capítulos posteriores são abordados a Fundamentação Teórica e Revisão do Estado da Arte, a Metodologia empregada nos estudos, os Estudos Realizados e as Conclusões, com discussões feitas em relação aos resultados obtidos.

Capítulo 2

Fundamentação Teórica e Revisão do Estado da Arte

Este capítulo aborda os conceitos de evasão escolar comumente utilizados e elenca formas de se calcular evasão. O problema da evasão nos cursos superiores, em especial nos cursos de Computação é elencado. Também são explicadas as técnicas estatística de Análise de Sobrevivência e o diagrama *Box-plot*. São abordados também o conceito e as tarefas comuns de mineração de dados, metodologias de mineração de dados e a mineração de regras de associação com o algoritmo Apriori. As classificações de curso adotadas pelo INEP também são explicadas nesse capítulo.

2.1 Conceitos e Cálculos de Evasão Escolar

O conceito de evasão escolar pode nos parecer à primeira vista bem sucinto e autoexplicativo, sendo explicado como o fenômeno de saída do aluno do curso antes da obtenção do diploma. No entanto, os conceitos que envolvem evasão escolar podem mudar de acordo com o autor. Conforme o estudo que está sendo conduzido, a abordagem utilizada por um autor para se calcular a evasão escolar pode diferir da abordagem usada por outro autor, levando a resultados bem diversos. Nas subseções a seguir os conceitos de evasão e as fórmulas empregadas no cálculo de evasão são exploradas.

2.1.1 Os Conceitos de Evasão

Nessa subseção, os diferentes conceitos de evasão são discutidos, é apresentada a abordagem usada nesse trabalho e a justificativa para o seu emprego.

Gaioso [45] apresenta, como definição para a evasão, a “interrupção no ciclo de estudo”. Segundo o autor, o aluno é tido como evadido quando “deixa o curso por qualquer

motivo que não seja a obtenção da titulação”, podendo ser enquadrado em uma das seguintes alternativas: não efetuar a matrícula no prazo estabelecido; transferência interna ou mudança de curso; transferência externa; matrícula em curso de outra instituição via aprovação em processo seletivo; desistência, re-opção ou jubramento.

O INEP diferencia ainda evasão e abandono: “O conceito técnico de abandono é diferente de evasão. Abandono quer dizer que o aluno deixa a escola num ano mas retorna no ano seguinte. Evasão significa que o aluno sai da escola e não volta mais para o sistema”[32].

Vargas¹ (2007 apud Almeida, 2007) [3] explora os conceitos de evasão de acordo com a sua amplitude, conforme pode ser visto na Tabela 2.1:

Tabela 2.1: Conceitos de evasão e suas amplitudes

Autor/Data	Definição	Amplitude do Conceito
Utiyama e Borba (2003)	Evasão é entendida como a saída definitiva do aluno de seu curso de origem, sem concluí-lo	Ampla. Não foi estabelecido nenhum critério de tempo no curso para a saída do aluno
Maia e Meireles (2005)	Evasão consiste em alunos que não completam cursos ou programas de estudo, podendo ser considerados como evadidos aqueles alunos que se matriculam e desistem antes mesmo de iniciar o curso	Específica que mesmo os alunos que nunca começaram o curso devem ser considerados no cálculo das taxas de evasão
Abbad, Carvalho e Zerbiani (2005)	Evasão refere-se à desistência definitiva do aluno em qualquer etapa do curso	Não deixa claro se evasão se aplicaria apenas aos alunos que chegaram a iniciar o curso ou se abrangeria também aqueles que apenas se matricularam e nunca iniciaram o curso

Fonte: Vargas¹ (2007 apud Almeida, 2007)[3]

Nesse estudo, são considerados evadidos todos os alunos desistentes com algum vínculo de matrícula, ainda que nunca tenham iniciado o curso. Em se tratando dos dados do INEP, que também são estudados nesse trabalho, não há como distinguir se os alunos chegaram a assistir alguma aula. Além disso, é interessante também saber o que motivou essa categoria de alunos a evadir do curso.

Deve-se levar em consideração também, que a evasão pode ser temporária ou definitiva. Sobre isso, Polydoro² (2000 apud Fialho, 2014) [43] diz o seguinte:

¹Vargas, Miramar Ramos Maia. Implantação de programas de educação a distância. Material Didático do Curso de Pós-Graduação em Educação a Distância. Centro de Educação a Distância, Universidade de Brasília, 2007.

²Polydoro, Soely Aparecida Jorge. O Trancamento de Matrícula na Trajetória Acadêmica do Universitário: condições de Saída e de Retorno à Instituição. 2000. 175 f. Tese (Doutorado em Educação) Faculdade de Educação - Universidade Estadual de Campinas. São Paulo. 2000.

Temos dois aspectos sob os quais a evasão pode ser analisada, a evasão temporária e a evasão definitiva. Na primeira, o aluno sai parcialmente do curso, mas com a intenção de voltar logo após a solução do problema. Neste espaço de tempo, entre o abandono, a solução do problema e a volta para o curso, tem-se uma grande probabilidade de o aluno não retornar para o curso no qual ingressou. Na evasão definitiva, o aluno se evade e não pretende voltar [43].

Lidando com dados agregados o pesquisador fica impossibilitado de definir se a evasão do discente é temporária ou definitiva, uma vez que não há dados individualizados para analisar cada aluno separadamente.

Podemos também lidar com evasão reversível ou irreversível [43]:

Por fim, conceitua-se modalidade reversível e irreversível. O primeiro caso vai depender se a instituição preservar o aluno que pretende se evadir, procurando uma maneira de mantê-lo na instituição mesmo que esteja com a intenção de sair. Por isso, é considerada reversível, pois o aluno quer sair, mas a instituição consegue mantê-lo, suprimindo as dificuldades e necessidades. A modalidade irreversível, a instituição pode até tentar, mas não consegue manter o aluno em seu estabelecimento educacional, provocando a evasão escolar.

Levando-se em conta que muitas evasões podem ser revertidas, torna-se ainda mais relevante conhecer as diversas variáveis que levam os alunos a evadirem e conhecer quando eles estão mais propensos a saírem do curso. Nessa dissertação, além de se analisar causas inerentes à evasão dos discentes, é analisado o aspecto temporal da evasão, via análise de sobrevivência. Conhecer bem o fenômeno de evasão é fundamental no sentido de se promover ações para que as evasões possam se enquadrar na categoria reversível.

O conceito de evasão pode se mostrar diversificado, de acordo com o objetivo do autor. Em estudo feito pela comissão especial de estudos sobre a evasão nas universidades públicas brasileiras [25] são apresentados os conceitos de evasão de curso (desligamento apenas do curso, sem desvínculo com a instituição); evasão da instituição (desligamento da instituição na qual está matriculado); evasão do sistema (abandono do ensino superior). Essa comissão afirma que a seleção do conceito mais adequado de evasão deve ser feito em função do objetivo pretendido no estudo a ser realizado.

Nesse estudo foi adotado o conceito de evasão de curso, pois pretende-se verificar se o aluno evadiu do curso de Computação, ainda que ele tenha permanecido na mesma instituição de ensino fazendo outro curso. Sendo assim, transferências para outros cursos de áreas diferentes de Computação, dentro da mesma instituição, serem consideradas evasões no âmbito desse estudo.

2.1.2 Calculando a Evasão

Ao se consultar a literatura é possível encontrar diferentes abordagens para cálculos de evasão. Algumas abordagens empregadas estão relacionadas na Tabela 2.2:

Tabela 2.2: Fórmulas para cálculo da evasão

Fórmula	Legenda da fórmula	Uso da fórmula
$E = [(Ni - Nd - Nr)/Ni] \times 100$ <p>(Fórmula 1)</p>	E = Percentual de evasão Ni = Número de ingressantes Nd = Número de diplomados Nr = Número de retidos	MEC [25]
$E = [(Ni - Nd) \times 100] / Ni$ <p>(Fórmula 2)</p>	E = Percentual de evasão Ni = Número de ingressantes Nd = Número de diplomados	Adachi [2]
$E = 1 - [C(t) / I(t-4)]$ <p>(Fórmula 3)</p>	E = Percentual de evasão C = Número de concluintes I = Número de ingressantes t = Ano avaliado (t-4) = Quatro anos anteriores	OCDE [83]
$E(n) = 1 - [M(n) - I(n)] / [M(n-1) - C(n-1)]$ <p>(Fórmula 4)</p>	E = Percentual de evasão M = Número total de matriculados C = Número de concluintes I = Número de ingressantes n = Ano em estudo (n-1) = Ano anterior	Silva Filho et al. [88]

Na Tabela 2.2, as fórmulas empregadas visam todas o cálculo da evasão de curso.

A Fórmula 1 é usada para acompanhamento de uma ou mais turmas de um determinado curso, onde é calculada a evasão levando-se em conta cada período de ingresso e o respectivo período máximo para conclusão do curso. Em relação a essa fórmula pode-se dizer que:

[...]Dessa forma, no levantamento de evasão de curso, considera-se a série histórica de dados sobre uma geração/turma de alunos ingressantes e o tempo máximo de integralização curricular. São identificados como evadidos do curso os alunos que não se diplomaram neste período e que não estão mais vinculados ao curso em questão. [25]

A Fórmula 2 apresentada é baseada na Fórmula 1, desconsiderando-se os alunos retidos no período e foi usada para estudo da evasão nos cursos de graduação na Universidade Federal de Minas Gerais (UFMG). Os alunos classificados como retidos são aqueles que após o prazo máximo de integralização curricular definido na instituição, não concluem o curso, mantendo ainda vínculo de matrícula na universidade [2]. Em se tratando da Fórmula 2, como o índice verificado desses alunos retidos foi bem pequeno, foram desconsiderados na respectiva fórmula [2].

A Fórmula 3 foi empregada pela Organização para a Cooperação e Desenvolvimento Econômico (OCDE) para o cálculo de evasão considerando-se cursos com períodos de

duração de 5 anos. Essa fórmula abrange, portanto, a taxa de evasão no período completo do curso. Seguem considerações sobre esse modelo:

Esse modelo utiliza parte da Taxa de Conclusão de Graduação (TCG) que é um indicador do REUNI. O TCG é calculado pela razão entre os concluintes do ano T e os ingressantes do ano T-4, para cursos de 5 anos de duração. Diferente dos dois modelos anteriores, esse modelo abarca um período completo do curso, que é conhecido como geração completa. Assim num curso de 5 anos o período entre ingresso e conclusão é de 5 anos. [83]

A Fórmula 4 é empregada para o cálculo de evasão anual. Ela consiste basicamente em verificar qual o percentual de alunos que progride de um ano para o ano posterior. Por exemplo, se a pretensão for saber a taxa de evasão em um determinado curso no ano de 2010, considera-se o número de matriculados em 2010 menos o número de ingressantes em 2010 e divide-se pelo número de matriculados em 2009 menos o número de concluintes em 2009. Assim, pode-se obter o número de matriculados em 2009 que continuaram a estudar em 2010. Os proponentes dessa metodologia de cálculo de evasão anual fundaram o instituto Lobo. Este instituto tem como objetivo principal:

Contribuir na solução dos problemas brasileiros nas áreas de educação, ciência e tecnologia. Para tanto, suas atividades são voltadas para a geração e divulgação de conhecimento, através de pesquisas e cursos desenvolvidos pelo seu corpo de gestores e consultores associados. Mais especificamente, as ações do Instituto Lobo visam reunir competência e experiência acadêmicas que possam ser transformadas em projetos de interesse do País. Todos os projetos precisam estar intimamente relacionados com as questões nacionais e, mais do que isso, devem contribuir diretamente na solução prática dos problemas enfrentados pelos gestores públicos e privados dos setores educacional, científico e tecnológico. [87]

A Fórmula 4 tem sido utilizada em diversos trabalhos para cálculos de taxas anuais de evasão. A UnB, por exemplo, tem feito uso dessa metodologia em alguns de seus estudos sobre trajetórias dos discentes para mostrar taxas anuais de evasão [59]. A Pró-Reitoria de graduação da Universidade Federal de São Paulo (UNIFESP) [37], ao analisar as taxas de evasão nos cursos de graduação das Universidades Federais em 2014 usando os microdados do censo da educação Superior – INEP, comparou as taxas anuais de evasão dos cursos ofertados pela UNIFESP em relação aos mesmos cursos ofertados por outras universidades federais, usando para o cálculo da evasão a metodologia do instituto Lobo, definida pela Fórmula 4. Estudo sobre evasão anual nos cursos de licenciatura em física no país, compreendendo o período de 2000 a 2012 [68], também usou a fórmula adotada pelo instituto Lobo como abordagem. Para os cálculos de evasão anual nos cursos de graduação presenciais da Universidade Federal da Paraíba (UFPB) no período de 2007 a 2012 [43], a metodologia do instituto lobo também foi utilizada. Estes são apenas alguns

dos exemplos que mostram o emprego desta fórmula em diversos trabalhos desenvolvidos em várias universidades.

Em documento disponibilizado pelo instituto Lobo [86], são comparadas duas Fórmulas consideradas clássicas para cálculo da evasão anual, chamadas de E1 (Evasão 1) e E2 (Evasão E2). A primeira Fórmula (E1) é a Fórmula 4, já discutida anteriormente, exibida na Tabela 2.2. A segunda fórmula [86] está representada a seguir:

$$E2 = 1 - [(M(n+1))/(M(n)-C(n)+I(n))]$$

Onde M corresponde ao número total de matrículas; n corresponde ao ano em avaliação; C corresponde ao número de concluintes; I corresponde ao número de ingressos. Em relação à essas duas fórmulas, pode-se dizer o seguinte:

E1 mede a perda de alunos veteranos de um ano para o outro em relação ao total de alunos veteranos que, potencialmente, deveriam se matricular, uma vez que não concluíram o curso. Por outro lado, E2 mede a perda de alunos veteranos, mas, agora, em relação à expectativa de matrículas no ano n+1, isto é, veteranos que deveriam se matricular, como no caso acima, porém, somadas às matrículas dos ingressantes no ano. Certamente $E2 < E1$, mas E1 é uma descrição mais correta da realidade, sendo a fórmula utilizada pelo Instituto Lobo! [86]

O estudo de evasão publicado pelo instituto Lobo em 2007 exhibe as taxas de evasões anuais nas áreas gerais do conhecimento e em diversos cursos, e “gerou matérias em diversos jornais e revistas de circulação nacional e publicações em revistas científicas indexadas nacionais e internacionais, com uma importante e nova abordagem do problema” [85].

Diante da grande aceitação e uso da metodologia do instituto lobo, vale a pena uma análise mais aprofundada sobre essa abordagem. A maioria dos estudos conduzidos pelo referido instituto lida com dados agregados, ou seja, só se tem disponíveis informações totalizadas de matriculados, concluintes e ingressantes, conforme trecho a seguir: “Como os dados do INEP são obrigatoriamente agregados, não é possível acompanhar a evasão por uma coorte representada pelo acompanhamento individual dos alunos” [88]. Essa metodologia se mostrou eficaz quando se precisa estimar a evasão e não se pode acompanhar o status de cada aluno no curso. No entanto, como se trata de uma estimação, os resultados podem apresentar, em algumas circunstâncias, inconsistências e valores que não retratam a evasão real. A Tabela 2.3 foi construída a partir de informações do instituto Lobo e retrata um problema que pode ocorrer ao lidar com dados agregados do INEP:

Tabela 2.3: Taxas de evasão no curso de Formação de Prof. da Educação Básica

Taxas de evasão - 2001 a 2005						
Curso	2001	2002	2003	2004	2005	Média
Formação de Prof. da Educação Básica/Normal Superior	-46	17	25	3	38	7

Fonte: Dados obtidos do Instituto Lobo, 2007 [88]

A Tabela 2.3 foi extraída a partir de um trabalho do instituto Lobo [88] sobre evasão no ensino superior no Brasil. Dados de outros cursos foram omitidos na Tabela 2.3. O que desperta a atenção na evasão para o curso de Formação de Professor da Educação Básica/Normal Superior, é a taxa de evasão correspondente ao ano de 2001, que foi de -46%. Em relação a esse valor atípico, tem-se o seguinte comentário:

Em 2001, os cursos de Formação de Professores da Educação Básica / Normal Superior apresentaram uma taxa de evasão anual negativa (-46%), reflexo, certamente, da busca por esse curso que teve um aumento significativo da oferta entre 2000 e 2001, por conta de uma exigência da Lei de Diretrizes e Bases da Educação – LDB – Lei n. 9.394/96, e que deve ter atraído alunos já formados em outros cursos – ou que mudaram de curso – que não foram considerados como ingressantes, mas apenas como matriculados. Após esse resultado atípico, o curso passou dois anos com taxas positivas e altas de evasão, para cair de novo em 2004 e chegar a 38% em 2005, mostrando um comportamento inconstante, de modo que a média no período não tem muito significado. [88]

É fácil perceber que o valor de evasão de -46% não condiz com a evasão real para o curso de Formação de Professores da Educação Básica no ano de 2001. No entanto, quando se depara com dados agregados fica difícil em certas circunstâncias obter uma precisão no cálculo da evasão. Taxas mais coerentes de evasão poderiam ser obtidas no período avaliado na Tabela 2.3 se os dados disponibilizados pelo INEP fossem individualizados, com a condição de matrícula informada para cada discente em cada curso (se cursando, se concludente, se desvinculado, se transferido etc).

Os dados do censo da educação superior fornecidos pelo INEP passaram a ser individualizados a partir de 2009, conforme o resumo técnico do Censo da Educação Superior de 2009 [34]. Anteriormente a essa data, para se estimar a evasão anual ou a evasão total (considerando-se o período total para conclusão do curso), era necessário usar uma abordagem como as mencionadas anteriormente. Contudo, vimos que podem surgir alguns problemas ao se usar a abordagem clássica para dados agregados do instituto Lobo para se calcular evasão usando os dados do INEP. Desta forma, nesse trabalho, o acompanhamento de cada aluno é feito individualmente, anualmente, para se obter as taxas de evasão anuais. A metodologia empregada para cálculo da evasão é detalhada no Capítulo 3 (Metodologia).

2.2 Evasão Escolar no Ensino Superior

O problema da evasão nos cursos superiores é multifacetado. Diferentes questões podem motivar a saída precoce do aluno do curso: questões sócio-econômicas, vocacionais, institucionais, pessoais, de saúde, familiares e outras questões. Políticas governamentais ligadas à educação, formas de ingresso, tipo de instituição (se privada ou pública), também podem influenciar a evasão. Diferentes fatores ligados à evasão nos cursos superiores são discutidos nas subseções a seguir.

2.2.1 Causas da Evasão nos Cursos Superiores

A evasão nos cursos de graduação do país é um problema constante, mas os dados provenientes dos alunos de nível superior são pouco explorados [82]. Essa realidade pode ser confirmada visto que não há um estudo em abrangência nacional feito pelo INEP, por exemplo, no sentido de se apurar evasão [31].

O problema de evasão nos cursos de graduação em território brasileiro é uma realidade já conhecida. Várias e frequentes causas que levam a esse quadro já foram mapeadas. Segundo a Comissão Especial do MEC para Estudo da Evasão (1996) citada por Adache [2] algumas causas que poderiam ser elencadas são:

As circunstâncias mais voltadas às características individuais dos alunos (como personalidade, dificuldades pessoais de adaptação à vida universitária, descoberta de novos interesses etc); Fatores internos às instituições (como currículos desatualizados, alongados, rígida cadeia de pré-requisitos, estrutura de apoio ao ensino insuficiente etc); Fatores externos às instituições (como questões relativas ao mercado de trabalho, falta de reconhecimento social da carreira, desvalorização da profissão etc), sendo as causas as mais variadas. Devido à diversidade de causas possíveis e caráter muitas vezes subjetivo do motivo que leva à desistência discente, constitui um trabalho não trivial tentar antever se o aluno se manterá ou não até o término do curso.

Gaioso [45], ao pesquisar a evasão em uma instituição de ensino superior no período de 2000 a 2003, elenca algumas causas de evasão mais recorrentes no meio universitário: a necessidade de trabalhar / horário de trabalho incompatível com o de estudo, problemas financeiros, casamento / nascimento de filhos, desconhecimento da metodologia do curso escolhido, deficiência da educação básica e reprovações sucessivas, falta de orientação vocacional / profissional e imaturidade, a ausência de perspectivas de trabalho, a falta de laços afetivos na universidade, busca de desafio a si mesmo (ou seja, buscar ser admitido em um curso sem a intenção de cursá-lo, apenas para demonstrar sua capacidade em ser admitido para aquele curso), a herança profissional (matricular-se em um curso escolhido pela família do discente, não por ele mesmo), a ausência de um referencial na família, mudança de endereço e concorrência entre as instituições de ensino superior.

Em um estudo realizado por Baggi e Lopes [6] sobre a relação entre evasão e avaliação no ensino superior, analisando a produção teórica entre 2008 e 2009, verificou-se que as causas para a evasão são muito diversas e dependendo de fatores sociais, culturais, políticos e econômicos em que a instituição se encontra. Uma causa possível seria a má qualidade do ensino ofertado pela instituição, levando à desmotivação e perda do aluno. Isso requer uma maior reflexão a respeito da relação entre a avaliação e critérios adotados pela instituição e a evasão, permitindo assim elaborar propostas mais eficazes de combate à evasão. Sendo assim a evasão pode estar, em muitos casos, associada diretamente a fatores institucionais.

De acordo com Souza et al. [90] que estudaram a evasão no Brasil no período de 10 anos em cursos superiores com base na produção acadêmica sobre o assunto durante o período, as seguintes causas de evasão aparecem com maior frequência: falta de condições financeiras, influência familiar, questão vocacional, reprovações em disciplinas que requerem conhecimento matemático, qualidade do curso, localização da instituição, trabalho, idade do aluno (evasão mais alta verificada entre os de maior idade). Verifica-se que a questão vinculada à aprendizagem em matemática figura entre as principais causas de evasão em cursos superiores .

Segundo Prediger e Luana [76] que conduziram uma pesquisa com alunos e professores das séries iniciais e séries finais do ensino fundamental procurando entender os motivos de desinteresse pela matemática, verificou-se que há uma queda de interesse por essa disciplina em séries finais do ensino fundamental em relação às iniciais e, segundo eles, “um dos fatores que se acredita causar o decréscimo é a passagem dos conteúdos trabalhados de forma mais concreta para conceitos, tidos como mais abstratos e lineares”.

Tal como a matemática, disciplinas físicas também se mostram complexas para os discentes. Em um trabalho em que se avalia se é possível automatizar a dedução de determinadas equações físicas, ou seja, se a partir dos dados experimentais é possível que um computador substitua o trabalho dos físicos teóricos e resolva o problema, Cubitt et al. [20] revelam que embora o desenvolvimento atual dos supercomputadores leve a acreditar que o trabalho humano possa ser substituído, no caso da física isso é matematicamente impossível, uma vez que esse tipo de problema se denomina, na computação, como NP, o que enfatiza a complexidade da disciplina de física. Machado e Cavalcanti [63] conduziram um trabalho com a finalidade de conhecer a influência que as reprovações nas disciplinas de Física exercem no sucesso/insucesso acadêmico, dos estudantes do curso de Química da Universidade Federal da Bahia relacionando as reprovações com a evasão universitária. Em entrevista realizada com os alunos de Química que cursaram disciplinas de Física no período de 2006 a 2010, 78% dos entrevistados não obtiveram sucesso acadêmico nessas disciplinas. Dentre os alunos que não obtiveram sucesso, 33% dos alunos consideram

disciplinas difíceis e 45% dos alunos as consideram disciplinas muito difíceis. Esses alunos apontaram como motivos relacionados às dificuldades o modo abstrato e conteudista das disciplinas e como os assuntos são abordados em sala de aula, bem como, o uso de conhecimentos matemáticos que estão além do exigido pela estrutura curricular do curso.

O estreitamento dos laços entre professores e discentes, também se mostra importante para diminuição da evasão do aluno. De acordo com Pascarella e Terenzini³ (2005 apud Mercuri e Fior, 2012) [66], quando o professor se aproxima do discente no sentido de interagir mais com o aluno não só dentro como fora da sala de aula, a probabilidade de permanência do aluno é ampliada. Sendo assim, a forma como o professor lida com o discente exerce um papel primordial não apenas para o seu aprendizado mas também para a sua permanência na instituição.

Segundo Barroso e Falcão [10] as condições que motivam a evasão escolar são classificadas sob três classes distintas: i) Questão econômica - impossibilidade de permanecer no curso por questões sócio-econômicas; ii) Questão vocacional – falta de identificação com o curso; iii) Questão institucional – abandono por fracasso nas disciplinas iniciais, deficiências prévias de conteúdos anteriores, inadequação aos métodos de estudo, dificuldades de relacionamento com colegas ou com membros da instituição.

Muitas causas de evasão em cursos de graduação citadas por diversos autores são recorrentes em instituições de ensino superior e podem ser classificadas em um dos 3 grandes grupos previstos por Barroso e Falcão [10]. A Tabela 2.4 contem algumas causas de evasão em cursos superiores segundo alguns autores: Adachi [2], Andriola et. al. [5], Muriel [69] e Giraffa e Mora [47]. Pode-se reparar que muitas causas descritas nos artigos são compatíveis entre si:

³Ernest T.. Pascarella, and Patrick T.. Terenzini. How college affects students: a third decade of research. Jossey-Bass, 2005.

Tabela 2.4: Causas de evasão em cursos superiores

Classificação de causas de evasão segundo Barroso e Falcão	Motivo de evasão	Autores
Questões sócio-econômicas	Incompatibilidade com o trabalho	Adachi
		Andriola et. Al.
		Muriel
	Questão financeira	Adachi
		Andriola et. Al.
		Muriel
Falta de tempo para estudar	Giraffa e Mora	
Questão vocacional	Mudança de curso	Andriola et. Al.
		Adachi
		Muriel
	Curso pouco atraente socialmente	Adachi
		Muriel
Questão institucional	Critérios de avaliação usados pela instituição	Adachi
		Muriel
	Dificuldades em disciplinas	Adachi
		Giraffa e Mora
		Muriel
		Adachi
	Falta de informações sobre o curso	Andriola et. Al.
		Muriel
		Adachi
	Base escolar deficiente	Andriola et. Al.
Muriel		
Adachi		
Inadequações do professor	Giraffa e Mora	

É natural que questões sócio-econômicas exerçam maior impacto sobre a evasão para alunos das universidades particulares do que para alunos das escolas públicas. Sendo assim, ao se levantar causas de evasão, a mantenedora institucional (se pública ou privada) deve ser levada em conta, sendo que, dependendo se a instituição onde se analisou as causas de evasão é pública ou privada, as causas podem diferir [58].

Políticas educacionais adotadas pelo governo também podem impactar a evasão nos cursos superiores. Algumas, dessas políticas são discutidas nessa subseção.

O Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais (REUNI) é um programa de expansão das universidades federais que objetiva ampliar o acesso e a permanência no ensino superior [22]. Dentre as ações do programa podemos destacar: “o aumento de vagas nos cursos de graduação, a ampliação da oferta de cursos noturnos, a promoção de inovações pedagógicas e o combate à evasão” [22].

O Governo Federal, através do REUNI relata que “os índices de evasão de estudantes

nos cursos de graduação atingem, em alguns casos, níveis alarmantes” [70]. A meta do governo é a elevação gradual da taxa de conclusão média dos cursos de graduação presenciais para noventa por cento. A relação entre o REUNI e a evasão nos cursos superiores merece ser melhor analisada e foge ao escopo desse trabalho.

Cabe analisar também a influência do Sistema de Seleção Unificada (SISU) sobre a evasão no ensino superior. Uma breve explicação sobre o SISU é apresentada a seguir:

O Sistema de Seleção Unificada consiste, como o próprio nome indica, em um sistema de seleção em escala nacional cujo objetivo é a alocação de candidatos para cursos oferecidos na rede federal de educação superior pública. As vagas das instituições federais de ensino superior (IFES) são oferecidas àqueles que realizaram o Exame Nacional do Ensino Médio (ENEM). O sistema foi estabelecido em 2009 e a primeira seleção unificada ocorreu em 2010. Antes do SiSU, a seleção para as Ifes era feita, basicamente, por meio de exames vestibulares de cada instituição, o que propiciava um emaranhado de datas sobrepostas de provas e de deslocamentos custosos para estudantes que pretendiam pleitear vagas em IES mais distantes. [46]

Em um estudo realizado durante o XXVI Encontro Nacional de Pró-reitores de Graduação desenvolvida pelo Grupo de Trabalho sobre Evasão e Retenção junto ao colegiado da ANDIFES [4], foi feita uma análise quantitativa da evasão em oito instituições nacionais entre o período de 2000 a 2011, sendo elas: UnB, a Universidade Federal do Vale do São Francisco (UNIVASF), a Universidade Federal de Campina Grande (UFCG), a Universidade Federal do Rio Grande do Norte (UFRN), a Universidade Estadual de Feira de Santana (UEFS), a Universidade Federal do Rio Grande (FURG), a Universidade Federal do Pampa (UNIPAMPA) e a Universidade Federal de Santa Maria (UFSM). O sistema SISU (implantado em 2009, mas avaliado a partir de 2010) apresentou no primeiro ano uma taxa correspondente ao dobro da evasão geral, de acordo com o levantamento. De acordo com o prof. Raul Ceretta Nunes (então pró-reitor adjunto de Graduação da UFSM), que apresentou o estudo no encontro nacional, pode-se supor que a elevada taxa corresponda mais a um excesso de mudança de instituições entre os alunos do que uma desistência do curso superior visto que muitos alunos que ingressam em universidades longe de casa, retornaram às proximidades de sua cidade. Em relação ao abandono entre cotistas e não cotistas, observou-se que no processo vestibular, a porcentagem de evasão dos cotistas foi inferior nos três primeiros anos levantados (2005, 2006 e 2007). Porém, a partir de 2009, a evasão de cotistas chegou a quase o dobro da taxa dos não cotistas (21,51%) e, mesmo com uma queda significativa nos anos seguintes, tem se mantido superior. No sistema Sisu, observou-se que no primeiro ano analisado a evasão de não cotistas foi quase o triplo dos cotistas (46,91% contra 17,40%). Número que teve redução em 2011 para 23,97%.

Em documento disponibilizado pela Universidade Tecnológica Federal do Paraná (UTFPR), o diretor do campus Apucarana - Paraná, Aloysio Gomes de Souza Filho também comenta sobre o impacto da forma de seleção por SISU sobre a evasão:

“O aluno pode optar por cursos em instituições diferentes, em cidades que nunca havia ouvido falar. Por isso, parte dos discentes encontra diversas dificuldades e não consegue se adaptar. Por exemplo, uma pessoa que morou a vida inteira em São Paulo, não vai encontrar o mesmo ritmo em Apucarana”, compara o diretor [41]

Carvalho e Oliveira⁴ (2014 apud Gilioli, 2016) [46] trazem uma observação referente a influência da forma de ingresso via SISU, ao relatarem uma análise feita do curso de Licenciatura em História da Universidade Federal de Mato Grosso do Sul (UFMS):

[...]a metodologia SISU/ENEM fez crescer a procura pelo Curso, [que] todavia não voltou aos patamares de ingresso e permanência dos anos anteriores à 2009. Um problema observado na metodologia SiSU/Enem é que ao preencher vagas em chamadas sucessivas [o que ocorreu até 2014 nesse sistema], com casos de até sexta chamada, os alunos que não fizeram a opção primeira pela Licenciatura em História e pela UFMS, parece estarem mais suscetíveis a desistência. Isso tem-se refletido na evasão em menor espaço de tempo, ainda no primeiro semestre letivo.

De acordo com Portilho et al.⁵(2015 apud Gilioli, 2016) [46], é difundido no meio universitário que o SISU aumentou a taxa de evasão nas instituições públicas, embora seja necessário estudos aprofundados sobre o fenômeno. A justificativa para isso seria os alunos pretenderem retornar para regiões mais próximas dos seus lares, o que seria reforçado se o aluno possuir condições de ingressar em outras universidades.

Este impacto da forma de ingresso via SISU sobre a evasão merece ser avaliada também em relação à mantenedora institucional (se pública ou privada). Visto que o Programa Universidade para Todos (PROUNI) também utiliza as notas do ENEM. De acordo com portal do MEC, o PROUNI “concede bolsas de estudo integrais e parciais de 50% em instituições privadas de educação superior, em cursos de graduação e sequenciais de formação específica, a estudantes brasileiros sem diploma de nível superior” [24]. No que diz respeito à evasão na rede privada para os ingressos via PROUNI, o MEC afirma, usando como referência os anos de 2009 e 2010, que a evasão via ingressantes por PROUNI foi inferior ao percentual observado no sistema universitário privado brasileiro como um todo:

Tendo como referência os anos de 2009 e 2010, observa-se que foi de 15,6% o percentual de alunos que deixaram o ensino superior antes da conclusão de seus cursos em instituições privadas. No mesmo período, entre os alunos contemplados com bolsas do programa do governo federal, esse percentual foi de apenas 4% [12].

⁴De Oliveira, Vitor Wagner Neto; Carvalho, Camila. Evasão na licenciatura: estudo de caso. Revista Trilhas da História, v. 3, n. 6, p. 97-112, 2014.

⁵Portilho, Lúcio Antônio et al. A adoção do SiSU e a evasão na Universidade Federal de Uberlândia. In: XVIII SemeAd – Seminários em Administração. São Paulo: FEA-USP, nov. 2015.

Sendo assim, a forma de ingresso via SISU pode exercer um efeito sobre evasão diferente nas instituições privadas em relação ao que ocorre nas instituições públicas, frente à outras formas de ingresso como Vestibular e Programa de Avaliação Seriada (PAS).

No que diz respeito ao ingresso por sistema de cotas ou o acesso universal (sem cotas), a evasão também pode diferir. De acordo com o portal do MEC, a distribuição das cotas em universidades é feita da seguinte forma no Brasil:

As vagas reservadas às cotas (50% do total de vagas da instituição) são subdivididas — metade para estudantes de escolas públicas com renda familiar bruta igual ou inferior a um salário mínimo e meio per capita e metade para estudantes de escolas públicas com renda familiar superior a um salário mínimo e meio. Em ambos os casos, também será levado em conta percentual mínimo correspondente ao da soma de pretos, pardos e indígenas no estado, de acordo com o último censo demográfico do Instituto Brasileiro de Geografia e Estatística (IBGE). [23]

Em um estudo feito sobre os efeitos da política de cotas na UnB no período de 2004 a 2006 verificou-se que os cotistas apresentaram taxas de evasão inferiores aos não cotistas [16]. Um estudo sobre evasão feito na Universidade Estadual do Rio de Janeiro (UERJ), considerando-se a taxa de evasão no primeiro ano para os ingressantes no ano de 2005, revelou que “entre os alunos não cotistas, 15,6% abandonaram a UERJ no primeiro ano. Este número é percentualmente cerca de três vezes maior do que o apresentado pelos cotistas (5,8%) [53]”. Ainda de acordo com o estudo feito na UERJ “a utilidade da faculdade para os grupos cotistas parece ser maior e tal valoração se tem traduzido em uma maior persistência e taxas de graduação” [53].

Em relação ao efeito da concorrência (relação candidatos/vagas) apresentada pelo curso e a sua respectiva taxa de evasão, também há alguns trabalhos que correlacionam de forma inversa a evasão e a relação candidatos/vagas, ou seja, propoem que a medida que a concorrência aumenta, há uma tendência de que a evasão caia.

Em um estudo de evasão na Universidade Federal de Goiás (UFG), verificou-se que vários alunos ingressam em cursos menos concorridos visando serem transferidos para outros cursos [80]. A comissão especial de estudos sobre evasão nas universidades públicas brasileiras afirma que “baixas remunerações no mercado de trabalho diminuem a procura pelos cursos, afetam a qualidade da formação de seus alunos e aumentam a propensão a evasão”[25]. Em um estudo feito pelo instituto Lobo [88] correlacionando concorrência e evasão nas instituições superiores públicas e privadas, verificou-se que a mais alta taxa de evasão em 2005, que foi da área de Serviços (28%), correspondeu a menor relação candidatos/vagas (1,3). A área de Agricultura e Veterinária teve a menor taxa de evasão (13%) e a maior relação candidatos/vagas (4,5). Cursos como os de Medicina, conforme

mencionado por Spinosa⁶ (2008 apud Gomes et al., 2010) [48], que costumam ter elevados níveis de concorrência, costumam também ter índices baixos de evasão, o que se justifica pela satisfação e motivação do discente ao ser aprovado em um curso bastante concorrido e pelos anseios de realização profissional e financeira.

O que foi possível perceber é que as causas que envolvem a evasão são multifacetadas, sendo um problema amplo, complexo, sobre o qual exercem influência diversas variáveis. Conforme visto até aqui, além de fatores sócio-econômicos, institucionais, vocacionais e pessoais, políticas governamentais também merecem ser estudadas, pois alunos ingressantes por formas de ingresso diferentes, alunos cotistas e não cotistas parecem apresentar uma sobrevivência diferente quanto ao evento evasão. É ainda necessário analisar separadamente as instituições privadas e públicas, pois se assumirmos que as causas de evasão em instituições públicas são exatamente as mesmas para as instituições privadas, estaremos enviesando a pesquisa. Conforme visto, é plausível que fatores sócio-econômicos exerçam maior peso sobre a decisão de evasão em escolas particulares em relação às públicas. Sendo assim, uma análise pluralista é conduzida no Capítulo 4, levando-se em conta a complexidade do problema de evasão.

2.2.2 Causas da Evasão nos Cursos de Computação

Existem múltiplas causas associadas à evasão nos cursos superiores, conforme apresentado anteriormente. Em se tratando da área de Computação a situação não é diferente. Essa seção elenca diversos fatores relacionados à evasão nos cursos de graduação da área de Computação, de acordo com a produção bibliográfica referente ao assunto.

Os cursos de graduação em Computação possuem em suas matrizes curriculares diversas disciplinas voltadas para cálculos matemáticos e para uso de algoritmos, que é a base para que o aluno aprenda a desenvolver programas de computador. Devido à relevância das disciplinas de algoritmos para a formação superior de um profissional da área de Computação, cabe também analisar como o insucesso nessas disciplinas pode contribuir para o fenômeno da evasão.

Em relação às disciplinas voltadas para o ensino de algoritmos, Hinterholz Jr. [50] afirma que “essas disciplinas costumam ter altos índices de evasão e reprovação, sendo um dos gargalos existentes nos cursos de graduação, particularmente de computação, dificultando ou até mesmo impedindo a continuidade dos alunos no curso”. Rodrigues⁷ (2002 apud Hinterholz Jr, 2009) [50] destaca Diversos problemas que podem ser citados no processo de ensino e aprendizagem de algoritmos e programação:

⁶Gomes, Maria José et al. Evasão acadêmica no ensino superior: estudo na área da saúde. Revista Brasileira de Pesquisa em Saúde/Brazilian Journal of Health Research, 2010.

⁷Rodrigues, M. C. Como Ensinar Programação?. Informática-Boletim Informativo Ano I n, v. 1, 2002.

[...]dificuldade de adaptação dos alunos desenvolverem raciocínio lógico quando estão acostumados a decorar o conteúdo; falta de motivação do aluno gerada pelo despreparo e o desânimo quando ele acredita que a disciplina constitui um obstáculo extremamente difícil de ser superado.

Valaski e Paraiso [91] fazem uma conexão das disciplinas de programação com os altos índices de reprovação:

Tradicionalmente, as disciplinas ligadas a programação de computadores tem um alto índice de reprovação. Vários alunos ingressantes em cursos de Computação tem seu primeiro contato com programação nesta disciplina. A necessidade de desenvolver um raciocínio lógico, descrevendo este através de uma linguagem com sintaxe bastante específica, é determinante para este alto índice de reprovações.

Como agravante, podemos ter problemas ligados à didática no ensino de algoritmos em sala de aula, o que dificulta a aprendizagem, conforme pode ser verificado em Borges⁸ (2000 apud Hinterholz Jr, 2009) [50]: “o modo tradicional não consegue facilmente motivar os alunos a se interessar pela disciplina. Entre outras razões, isso ocorre, pois não é clara para os mesmos a importância de certos conteúdos para sua formação”. É importante que o ensino de programação seja prazeroso e englobe situações reais e dinâmicas para envolver o discente.

Há também os problemas inerentes à dificuldade de abstração algorítmica por parte do aluno, conforme cita Pereira e Rapkiewicz⁹ (2004 apud Hinterholz Jr, 2009) [50]:

[...]refere-se às dificuldades apresentadas por grande parte dos alunos em assimilar as abstrações envolvidas no processo de ensino e aprendizagem de programação. Além disso, muitos dos alunos não conseguem desenvolver o raciocínio lógico necessário para o posterior desenvolvimento de programas.

De acordo com Branco Neto e Schuvartz¹⁰ (2007 apud Prietch e Pazeto, 2010) [77]:

Os cursos da área de computação enfrentam um grande problema com as disciplinas de introdução à programação de computadores, as quais visam ensinar como utilizar o computador para solucionar problemas. Acadêmicos iniciantes, ao se depararem com a disciplina, sentem-se incapazes de programar, devido ao conjunto de habilidades que a programação exige como capacidade para solucionar problemas, raciocínio lógico, habilidade matemática, capacidade de abstração, entre outras

⁸Borges, Marcos AF. Avaliação de uma metodologia alternativa para a aprendizagem de programação. In: VIII Workshop de Educação em Computação–WEI. 2000.

⁹Júnior, J. C. R. P.; Rapkiewicz, Clevi Elena. O processo de ensino-aprendizagem de fundamentos de Programação: uma visão crítica da pesquisa no Brasil. In: Anais do XII Workshop sobre Educação em Computação (SBC). 2004.

¹⁰Neto, Wilson Castello Branco; Schuvartz, Aguinaldo Antonio. Ferramenta Computacional de Apoio ao Processo de Ensino-Aprendizagem dos Fundamentos de Programação de Computadores. In: Anais do Simpósio Brasileiro de Informática na Educação. 2007. p. 520-528.

Ressalta-se que essas disciplinas de algoritmos, essenciais para que o aluno aprenda a programar e, portanto, indispensáveis em sua formação profissional, são vistas nos períodos iniciais do curso, o que reforça a necessidade, por parte da instituição de ensino, de adotar estratégias coibidoras da evasão logo no início do curso.

Segundo Piva Jr. e Freitas [75] “a dificuldade apresentada por estudantes na representação algorítmica deve-se à dificuldade dos mesmos em extrair os detalhes que compõem uma situação problema”. De acordo com Paula et al. [74], em se tratando de discentes de Computação, “é difícil para os alunos conseguirem construir representações mentais que de fato abstraíam por completo um problema, sendo necessário buscar meios que estimulem os alunos a desenvolver tal capacidade”. Um estudo realizado na Universidade AL-AQSA – Palestina por Abu-Oda e El-Halees [1] envolvendo 1290 registros de alunos no período de 2005 a 2011, usando técnicas de mineração de dados, constatou que o êxito em disciplinas iniciais do curso relacionadas a algoritmos diminui a probabilidade de abandono do aluno.

Diante das dificuldades dos discentes no processo de abstração algorítmica, o professor deve buscar estratégias no sentido de tornar as disciplinas de algoritmos mais inteligíveis, conforme colocado por Rodrigues¹¹ (2004 apud Piva Jr. e Freitas, 2011) [74]:

O professor Methanias coloca muito bem a preocupação que impera nos processos de ensino de algoritmos. Ele diz “Ao ensinar programação, o professor deve sempre ter em mente que os alunos não são especialistas da área de computação e todos os conceitos apresentados são novos. Portanto, o professor deve ser capaz de fazer o aluno compreender a abstração envolvida com toda simbologia utilizada. Para a grande maioria, 'A', é apenas uma letra do alfabeto. Como fazer o aluno entender que, 'A', agora é uma variável ?”

Em um estudo feito sobre evasão no curso de Licenciatura Plena em Informática na Universidade Federal de Mato Grosso (UFMT), constatou-se que:

No que tange a cursos de computação, é comum os candidatos terem uma visão distorcida das habilidades, objetivos e perfil do egresso. Geralmente pensam que o foco destes cursos é aprender a trabalhar com softwares básicos, tais como editores de texto, navegadores de Internet, dentre outros. Neste sentido, no primeiro ano ocorre muita desistência, pois os acadêmicos percebem o intuito do profissional que se almeja formar e as habilidades que serão adquiridas ao longo dos anos na instituição [77].

Sendo assim, muitos alunos evadem devido à frustração por desconhecerem o foco real de um curso superior na área de Computação. Os alunos podem ter afinidade com um determinado software, ou com uma área bem específica da Computação, como Computação

¹¹Rodrigues, M. C. Como Ensinar Programação?. Informática-Boletim Informativo Ano I n, v. 1, 2002.

Gráfica por exemplo, e ao perceberem que o curso oferece uma formação científica mais generalista, acabam abandonando.

Um outro aspecto que pode influenciar a evasão são as perspectivas salariais em relação à formação que o curso confere. A área de Computação possui diferentes cursos e a remuneração pode diferir entre eles:

[...] o principal motivo é por tratar-se de um curso de licenciatura, onde as perspectivas salariais e a inserção no mercado de trabalho são mais restritas que outras áreas da informática. Neste aspecto, os professores do curso têm debatido a respeito de como atender as necessidades regionais. Visando atender este anseio, as alternativas encontradas são através da adequação da matriz, inserindo disciplinas específicas para sanar tais dificuldades, bem como ofertar alguns cursos de extensão visando à melhor qualificação dos futuros profissionais [77].

No caso descrito anteriormente, destaca-se a questão da atratividade de mercado, onde cursos com melhores perspectivas remuneratórias tendem a apresentar menores taxas de evasão. A média salarial para algumas profissões na área de Tecnologia da Informação no Brasil é apresentada na Tabela 2.5, de acordo com o Site Nacional de Empregos (SINE) [89]. Esse site (www.sine.com.br) não possui relações com instituições de governo e não deve ser confundido com o Sistema Nacional de Emprego (SINE) do Ministério do Trabalho e Previdência Social.

Tabela 2.5: Média salarial nos cursos de Computação

Profissão	Média Salarial
Professor de Informática	2263.22
Engenheiro de Computação	4948.14
Analista de Sistemas	4742.94
Engenheiro de Software	5294.66

Fonte: obtido a partir do Portal SINE [89]

A Tabela 2.5 apresenta o salário médio mensal, em Reais, em 2016, mas não contempla todas as profissões na área de Computação, onde se pode encontrar muitos cargos, vários deles bem específicos e focados em alguma tecnologia. No entanto, a Tabela 2.5 ilustra a discrepância que existe entre os salários de professor de informática e demais formações. A relação entre a evasão em Licenciatura em Computação e os salários oferecidos necessita de uma análise mais profunda que foge ao escopo desse trabalho. A princípio pode-se conjecturar que essa baixa remuneração pode ser um dos motivos para a evasão nesse curso.

O conhecimento matemático estabelece uma correlação com disciplinas algorítmicas. Barcelos e Silveira [9] discutiram sobre isso:

A falta do domínio adequado de conhecimentos matemáticos pelos alunos é um possível fator explicativo para a falta de interesse e evasão em cursos da área. Uma revisão da literatura permite apontar que diversos pesquisadores (CAMPBELL; MCCABE, 1984; WILSON; SHROCK, 2001; BEAUBOUEF, 2002; SETTI, 2009) indicam possíveis correlações entre o conhecimento matemático prévio dos alunos e seu desempenho em cursos introdutórios de Computação, bem como a relevância de tópicos do conhecimento matemático para uma melhor compreensão e modelagem dos processos computacionais. Coincidentemente ou não, constata-se um baixo rendimento na aprendizagem de Matemática pelos alunos no ensino básico.

Os índices de reprovações em disciplinas Matemáticas nos cursos superiores de Computação costumam ser elevados, conforme expresso na Tabela 2.6, retirada do estudo de Pazeto e Prietch [77] sobre a evasão no curso de Licenciatura Plena em Informática na UFMT mostrando a porcentagem média, de 2001 à 2008, de alunos reprovados por disciplina do curso, no 1º Ano:

Tabela 2.6: Reprovações por disciplina - Licenciatura em Informática (UFMT)

Primeiro Ano		
	Disciplina	% Média de Reprovações
1	Programação I	48,44
2	Lógica Matemática-Elem.de Lóg.Digital	47,03
3	Álgebra para Computação - Vetores e Geometria Analítica	49,60
4	Historia da Educação Brasileira (Optativa)	21,62
5	Inglês Instrumental	25,20
6	Português - Leitura e Produção de Texto	22,56
7	Física Geral e Experimental I	24,02
8	Cálculo I	56,99

Fonte: Pazeto e Prietch [77]

Conforme verificado na Tabela 2.6, a disciplina de Cálculo I apresentou maior percentual médio de reprovação no período 2001 a 2008. Com percentuais elevados, encontram-se também as disciplinas relacionadas à algoritmos e programação. Conforme expresso na tabela, essas disciplinas são estudadas logo no primeiro ano do curso e são disciplinas essenciais para que o aluno adquira as bases necessárias para outras disciplinas focadas em programação ou disciplinas matemáticas como Cálculo II.

Em relação ao uso da Matemática em cursos superiores de Computação, Gruner [49] afirma que os estudantes de Ciência da Computação na África do Sul que pretendem tornar-se pesquisadores ou cientistas é atualmente baixa e a popularidade da Matemática abordada no currículo dos cursos está em queda.

Embora existam vários trabalhos que relacionam as dificuldades matemáticas e as dificuldades em algoritmos com evasão e desmotivação dos alunos, obviamente, por ser um problema multifacetado, há outros fatores que podem levar à evasão nos cursos de Computação.

Em relação às causas de evasão em cursos superiores da área de Computação, um estudo realizado por Xenos et al. [92], elencou as causas de evasão em cursos superiores de Ciência da Computação na Hellenic Open University - Patras, Grécia: Profissional (62,1%), acadêmico (46,2%), familiar (17,8%), pessoais (8,9%), relacionado a saúde (9,5%). Ressalta-se que o curso de Ciência da Computação em estudo na Hellenic Open University é um curso à distância. Isso contribuiu para que as questões profissionais fossem mais citadas, visto que vários alunos alegaram dificuldades de conciliar o serviço e os estudos.

Percebe-se que são múltiplas as causas que podem estar relacionadas à evasão em cursos de Computação. Fazendo um breve apanhado das causas mencionadas até o momento nesse estudo e fazendo uso da classificação usada por Barroso e Falcão [10] usando 3 grandes grupos de causas, temos na Tabela 2.7 um pequeno resumo:

Tabela 2.7: Causas de evasão em cursos de Computação

Classificação de Causas de evasão Segundo Barroso e Falcão	Motivo de evasão
Questões sócio-econômicas	Dificuldades profissionais
	Mercado de trabalho pouco atraente (caso dos cursos de licenciatura)
Questão vocacional	Desconhecimento do foco do curso - verificação de falta de afinidade.
Questão institucional	Dificuldades em disciplinas de algoritmos/ Programação
	Dificuldades em disciplinas Matemáticas
	Dificuldades de abstração
	Dificuldades dos professores em repassar de maneira clara os conteúdos

Muitas das causas de evasão citadas podem ser enquadradas como questões sócio-econômicas, vocacionais ou institucionais. É claro que, embora essas 3 classificações sejam bastante amplas, elas não cobrem todas as causas de evasão. O aluno pode ser motivado a evadir, por exemplo, por questões pessoais, afetivas, familiares, depressão etc. Ou seja, as questões mencionadas, embora sejam recorrentes, não cobrem todas as causas de evasão, deixando evidente a complexidade do problema. Portanto, ao classificar os

problemas de evasão para os cursos de Computação, não se pode ficar limitado à apenas esses 3 grupos de questões, embora eles tenham se mostrado úteis para classificar diversas causas de evasão.

2.3 Técnicas Estatísticas

O *Box-plot* é um recurso gráfico que pode facilitar a comparação entre conjuntos de valores em relação a uma comparação feita a partir de dados tabelados. Seu uso facilita a compreensão dos dados de evasão apresentados no Capítulo 4. A técnica estatística de Análise de Sobrevivência possibilitará um estudo da evasão comparando-se diversos fatores como sexo, formas de ingresso, cotistas e não cotistas etc, levando-se em conta o fator tempo. Essas técnicas estatísticas são detalhadas nas subseções a seguir.

2.3.1 Box-plot

O *Box-plot*, também chamado de Diagrama de Caixa, é um recurso gráfico para descrição de valores que facilita a comparação entre diferentes conjuntos de dados. Ele consiste em “um gráfico estatístico que possibilita representar a distribuição de um conjunto de dados com base em alguns parâmetros descritivos” [15]. De acordo com Maroco¹² (2007 apud Capela e Capela, 2011) [15] de maneira geral, as estatísticas apresentadas em um gráfico do tipo *Box-plot* incluem a mediana, o primeiro e o terceiro quartis, os valores mínimos e máximos e os chamados outliers, que são aqueles valores que se mostram discrepantes em relação ao conjunto de dados plotado no gráfico. “A existência de outliers, valores extremamente altos ou extremamente baixos, pode indicar tanto dados incorretos como dados válidos que carecem de uma atenção especial.” [15].

Para melhor visualização da utilidade do *Box-plot* na comparação de dois ou mais conjuntos de dados, é exibido um exemplo. A Tabela 2.8 trás valores fictícios de evasão em 5 cursos de graduação de uma universidade X, em um período de 10 anos:

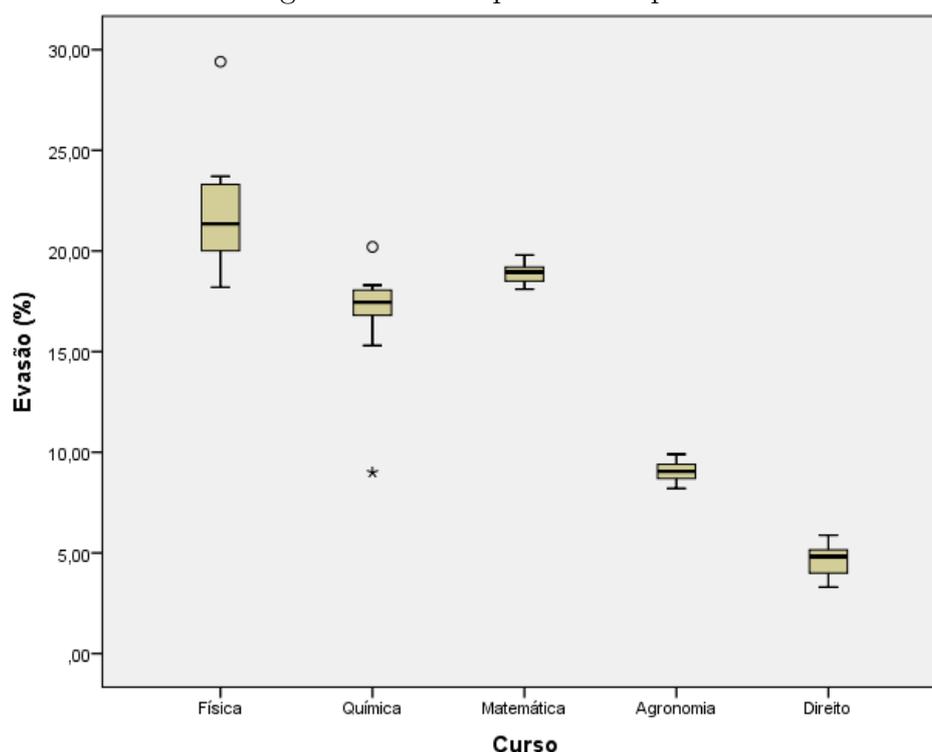
¹²Maroco, João. Análise Estatística com a utilização do SPSS. 3.ed. Edições Silabo, 842p., 2007.

Tabela 2.8: Evasão anual para cursos da universidade X - dados fictícios

Taxa de Evasão Anual (%)										
Curso	Ano									
	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009
Física	22,30	29,40	20,01	18,20	21,80	23,70	20,20	20,90	19,99	23,30
Química	18,30	17,40	9,00	15,30	20,20	17,40	17,50	17,80	16,80	18,05
Matemática	19,00	18,10	18,50	18,90	19,10	19,20	19,25	19,80	18,70	18,40
Agronomia	9,00	8,50	8,70	8,20	9,30	9,50	9,40	9,90	9,10	8,80
Direito	5,00	3,30	4,00	5,50	5,88	3,90	4,50	4,85	4,80	5,15

Pode ser complicado comparar a evasão no período 2000-2009 entre os cursos da Tabela 2.8, especialmente quando se pretende identificar informações como dados discrepantes dentro de um conjunto de dados. O diagrama *Box-plot* facilita essa tarefa. A Figura 2.1 exemplifica o uso de diagrama *Box-plot* para a Tabela 2.8:

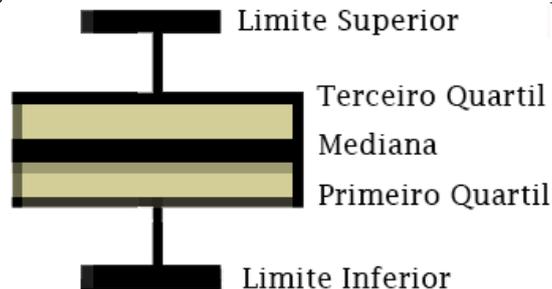
Figura 2.1: Exemplo de Box-plot



A Figura 2.1 foi gerada a partir dos dados da Tabela 2.8 por meio do software IBM SPSS Statistics versão 22. No eixo X temos os cursos e no eixo Y temos os percentuais de evasão. É perceptível que no período de 10 anos em avaliação o curso de Física apresentou maiores taxas de evasão, seguido do curso de Matemática e logo em seguida do de Química.

Isso pode ser deduzido pelo posicionamento das caixas no diagrama: aquelas posicionadas mais acima correspondem às maiores taxas de evasão. Percebe-se também que algumas caixas são maiores que outras. Por exemplo, a caixa correspondente ao curso de Física é maior que as demais. Isso acontece por que o curso de Física tem maiores variações nos valores de 10 anos em relação aos outros cursos, ou seja, a variabilidade dos dados é maior no curso de Física e é menor em cursos como Matemática e Agronomia. Percebe-se um traço preto cortando cada caixa. Este traço corresponde ao valor mediano do conjunto de valores de cada caixa. Cada caixa dispõe de uma haste inferior e de uma haste superior, que correspondem, respectivamente, ao limite inferior e ao limite superior. Percebe-se ainda que para os cursos de Física e de Química existem alguns valores que não fazem parte da caixa. Eles são representados por um círculo e por um asterisco. Esses valores são chamados Outliers, que são valores discrepantes em relação ao conjunto de dados observado. Essa é outra vantagem do *Box-plot*, a rápida identificação de valores incomuns. O outlier do curso de Física, por exemplo, corresponde à evasão do ano de 2001, que foi de 29,40%. A Figura 2.2 resume a estrutura de uma caixa em *Box-plot*:

Figura 2.2: Estrutura da caixa de um Box-plot



Percebe-se, a partir da Figura 2.2 que a caixa é constituída pelo terceiro e primeiro quartis. Os limites inferior e superior são obtidos, respectivamente, a partir das Fórmulas 2.1 e 2.2 [15]:

$$limite_inferior = 1^\circ_quartil - 1,5(3^\circ_quartil - 1^\circ_quartil) \quad (2.1)$$

$$limite_superior = 3^\circ_quartil + 1,5(3^\circ_quartil - 1^\circ_quartil) \quad (2.2)$$

Valores inferiores ao limite inferior são considerados outliers e valores superiores ao limite superior também são considerados outliers.

A interpretação de dados torna-se muitas vezes mais fácil a partir de um *Box-plot* do que via tabela, sendo um recurso bastante poderoso para se comparar grupos distintos. Sendo assim, no Capítulo 4, o *Box-plot* é usado para representação das taxas de evasão anuais nos cursos.

2.3.2 Análise de Sobrevivência

De acordo com Soares e Colosimo¹³ (1995 apud Lima Junior et al., 2012) [61]:

A análise de sobrevivência é um conjunto de métodos estatísticos que remontam a meados do século XX, mas foram desenvolvidos e se popularizaram majoritariamente em torno da década de 80”.

Segundo Maller and Zhou¹⁴ (1996 apud Lima Junior et al., 2012):

Apesar de estarem historicamente relacionados à pesquisa em epidemiologia e em clínica médica, os métodos da análise de sobrevivência são adequados à abordagem em diversas questões importantes em engenharia, sociologia, psicologia e educação.

A análise de sobrevivência tem atualmente, portanto, aplicações em diversas áreas que não se restringem à área médica, sendo seu emprego muito interessante quando o fator tempo é importante no estudo. O foco da análise de sobrevivência é justamente a verificação do tempo até a ocorrência de um evento, como a morte de um paciente, ou o tempo até o retorno de uma enfermidade, o tempo até que um automóvel apresente seu primeiro defeito após a venda, o tempo até que uma pessoa em liberdade com ficha criminal cometa novo crime, o tempo que um discente leva para se formar ou para evadir etc [61]. Fica claro então como a análise de sobrevivência pode ser útil em diferentes campos do conhecimento quando o fator tempo exerce relevância no estudo.

O emprego da análise de sobrevivência é, portanto, bastante adequado ao estudo da evasão escolar para se mensurar quanto tempo o discente leva até sair do curso, sem que tenha obtido o diploma de graduação.

A análise de sobrevivência tem sido utilizada em comparação com técnicas de mineração de dados clássicas e apresentado bons resultados. Segundo Kishleitner (2008) [56] regressão logística, análise discriminante, análise de sobrevivência, árvores de decisão, inferência bayesiana e redes neurais são algumas das técnicas que tem sido muito utilizadas para auxiliar na concessão, no acompanhamento, na cobrança e na retenção de clientes de cartão de crédito. Figini (2009) [44] menciona o uso da análise de sobrevivência para previsão de taxas de cancelamento de serviços de clientes ao longo do tempo. No estudo são usados dados de uma empresa de serviços de mídia. A empresa em questão usa uma árvore de classificação para prever a evasão de clientes e essa técnica é comparada com análise de sobrevivência, sendo que a abordagem proposta, com base na modelagem de análise de sobrevivência, levou a conclusões mais robustas.

Na análise de sobrevivência existe o conceito de dados censurados, que são os dados correspondentes aos indivíduos para os quais não se observou o evento de interesse:

¹³Soares, José Francisco; Colosimo, Enrico Antônio. Métodos estatísticos na pesquisa clínica. Anais da XL Reunião Anual da Região Brasileira da Sociedade Internacional de Biometria, 1995.

¹⁴Maller, Ross A.; Zhou, Xian. Survival analysis with long-term survivors. New York: Wiley, 1996.

Nos estudos com análise de sobrevivência, os dados destes participantes são aproveitados na análise final, mesmo que não desenvolvam o evento em estudo. Tal pode ocorrer porque abandonaram o estudo, foram perdidos no seguimento ou o estudo chegou ao seu término. Quando tal acontece, designa-se o indivíduo por censurado. Assim, censurado significa que o indivíduo não desenvolveu o evento até ao fim de observação no estudo (independentemente do motivo pelo qual tal lhe aconteceu ou do que lhe aconteceu posteriormente) [11]

No contexto deste estudo, por exemplo, em que se analisa a evasão escolar, os dados dos alunos falecidos são exemplos de dados censurados, já que esses alunos foram removidos do estudo sem se observar o evento de interesse, que nesse caso é a evasão. Discentes removidos do estudo devido à conclusão do curso também constituem casos censurados. Um outro exemplo de dados censurados seriam os dados daqueles discentes que, após decorrido o tempo do estudo, não se tenha observado a evasão.

As censuras podem receber diferentes classificações. Segundo Colosimo e Giolo¹⁵ (2006 apud Oliveira e Lins, 2011) [27], podemos ter os seguintes tipos de censura:

- Censura à direita do tipo I: aquela em que o estudo será terminado após um período pré-estabelecido de tempo independente das ocorrências do evento de interesse.
- Censura à direita do tipo II: quando as observações são acompanhadas até se obter um número pré-determinado de ocorrências do evento de interesse.
- Censura aleatória: acontece quando um participante sai ou é retirado do acompanhamento sem que tenha sido observada a ocorrência do evento de interesse.

Segundo Colosimo e Giolo (2006 apud Oliveira e Lins, 2011) [27] o tempo de vida na análise de sobrevivência é formado pelo instante inicial, a escala de medida e o evento de interesse. A escala de medida pode ser dada como tempo real (como horas, dias, meses etc) ou em intervalos regulares de tempo. Nem sempre é possível usar uma escala de medida do tempo com precisão em horas ou dias por exemplo. Os dados do INEP, conforme é apresentado no Capítulo 4 (estudos realizados), não permitem essa precisão pois são informações anuais sem a identificação de quantos dias o aluno levou até evadir. Já em se tratando dos dados da UnB é possível medir o tempo em períodos cursados (períodos de 6 meses).

De acordo com Colosimo e Giolo (2006 apud Lima Júnior et al., 2012) [61] as técnicas de análise de sobrevivência são enquadradas em 3 categorias: não-paramétricas, semi-paramétricas e paramétricas. A Tabela 2.9 descreve algumas características de cada um desses grupos:

¹⁵Colosimo, Enrico Antônio; Giolo, Suely Ruiz. Análise de sobrevivência aplicada. In: ABE-Projeto Fisher. Edgard Blücher, 2006.

Tabela 2.9: Categorias das técnicas de Análise de Sobrevivência

Técnica	Características
Não-paramétrica	Não fazem quase nenhuma restrição sobre a distribuição dos tempos de vida na população que gerou a amostra
	Mais flexíveis
	Não permitem testar o efeito de muitas variáveis ao mesmo tempo
	Principal técnica: estimador de Kaplan-Meier
Semi-paramétricas	Chamadas Regressão de Cox
	Vantagem sobre o modelo não-paramétrico: permite testar a significância estatística de diversas co-variáveis sobre a distribuição de tempos de vida
Paramétricas	Chamadas modelo de tempo de vida acelerado
	Definição rígida da forma funcional das distribuições de tempo de vida na população
	Técnicas pouco utilizadas

Fonte: Colosimo e Giolo (2006 apud Lima Júnior et al., 2012)[61]

Cabe discutir um pouco mais a respeito das técnicas de análise de sobrevivência não-paramétricas, devido a sua maior flexibilidade. Botelho et al. (2009) [11] diferenciam as duas principais técnicas não paramétricas de análise de sobrevivência empregadas em estudos:

O método Actuarial consiste em dividir o tempo de seguimento em intervalos iguais e estimar a probabilidade que os participantes que chegaram ao início de cada intervalo tinham de desenvolver o evento até final desse intervalo. Pelo contrário, o método de Kaplan-Meier consiste em dividir o tempo de seguimento em intervalos, cujos limites correspondem ao tempo de seguimento em que houve eventos. Este é um método mais utilizado actualmente, pois utilizamos a data exacta do evento, sendo os resultados mais precisos.

De acordo com Bustamante-Teixeira et al. (2002) [14] “A distinção essencial entre o método atuarial e o método de Kaplan-Meier é que este último elimina a necessidade de assumir que as censuras das observações ocorram uniformemente durante este intervalo”. O método de Kaplan-Meier considera, portanto, o momento exato em que o evento ocorreu, em vez de considerar número fixo de intervalos como o método atuarial. Portanto, o método de Kaplan-Meier é mais preciso. Os métodos não-paramétricos são bastante usados e se mostram bastante flexíveis. Dentre eles, se destaca o método de Kaplan-meier. Devido a essa maior precisão, este método é adotado nos estudos do Capítulo 4.

Segundo Lima Junior (2012)[61], “Chama-se função de sobrevivência $S(t)$ a probabilidade de que um indivíduo possua tempo de vida maior que t . Ou seja, a probabilidade de que, decorrido um tempo t , esse indivíduo tenha sobrevivido ao evento terminal”. A função de sobrevivência pode ser descrita pela Fórmula 2.3 [61]:

$$S(t) = Pr(T > t) \quad (2.3)$$

Ainda de acordo com Lima Junior (2012)[61], “em amostras numerosas, a função de sobrevivência $S(t)$ pode ser pensada como a fração de sobreviventes em função do tempo.”

De acordo com Colosimo e Giolo¹⁶ (2006 apud Lima Junior et al.,2012) [61], o Estimador de Kaplan-Meier (EKM) pode ser definido formalmente pela Fórmula 2.4:

$$EKM(t) = \prod_{t_i=0}^{t_i} \left(1 - \frac{d_i}{n_i} \right) \quad (2.4)$$

Segue a explicação para a fórmula do EKM apresentada:

[...]suponha que uma amostra compreenda N medidas de tempo de vida e que, dentre essas medidas, haja somente k observações distintas e não censuradas. Considere que essas medidas distintas de tempo estejam dispostas em ordem crescente $t_1 < t_2 < t_3 < \dots < t_k$. Sejam d_i o número de eventos terminais ocorridos em t_i e n_i o número de indivíduos que poderiam ser atingidos pelo evento terminal no tempo t_i [61].

Existem vários testes que podem ser usados para verificar a significância estatística em estudos de análise de sobrevivência. O software SPSS Statistics, em suas versões mais recentes, disponibiliza os seguintes testes estatísticos para análise de sobrevivência: Logrank, Breslow e Tarone-Ware. Em relação a estes testes, pode-se dizer o seguinte:

De maneira geral, esses testes consistem em calcular uma estatística ponderada que, sob a hipótese nula (hipótese de que os grupos de medidas são amostras originadas da mesma população), tem distribuição conhecida. Entretanto os testes possuem regras de atribuição de peso para melhor discriminar as curvas em determinadas etapas de sua evolução temporal. No teste Logrank, os pesos são atribuídos de maneira a enfatizar diferenças ao final das distribuições de tempo de vida. O teste de Breslow enfatiza diferenças no início da distribuição do tempo de vida. O teste de Tarone-Ware, por sua vez, foi desenvolvido com o objetivo de discriminar distribuições em fases intermediárias do tempo de vida [61].

No Capítulo 4, são apresentadas as curvas de sobrevivência usando o método de Kaplan Meier. Para construção dessas curvas foi usado o software SPSS Statistics e os 3 testes estatísticos mencionados anteriormente: Logrank, Breslow e Tarone-Ware, visto que cada

¹⁶Colosimo, Enrico Antônio; Giolo, Suely Ruiz. Análise de sobrevivência aplicada. In: ABE-Projeto Fisher. Edgard Blücher, 2006.

um enfatiza diferenças em momentos distintos da distribuição do tempo de vida, e assim, o teste de significância estatística poderá compreender diferentes etapas.

Embora o método de Kaplan Meier seja bastante flexível e ótimo para se traçar as curvas de sobrevivência ele possui como limitação não ser possível analisar o efeito de duas ou mais variáveis ao mesmo tempo. Nesses casos, é necessário fazer uso da regressão de Cox, que é usado para se verificar os riscos proporcionais de cada variável em relação ao evento terminal.

De acordo com Oliveira e Lins (Oliveira e Lins, 2011) [27] “os modelos de Cox são ditos de risco proporcional porque as razões entre as taxas de risco para indivíduos com diferentes valores ou em diferentes categorias das variáveis explicativas são assumidas como independentes do tempo”. A regressão de Cox é indicada nas circunstâncias em que:

[...]se deseja estudar sobrevivência sob o prisma de causalidade ou da predição, pois fornece as estimativas das razões de risco dos fatores estudados, podendo-se avaliar o impacto que alguns fatores de risco ou fatores prognósticos têm no tempo até a ocorrência do evento de interesse [14].

A função de risco do modelo de Cox é modelada pela Fórmula 2.5 [8]:

$$\lambda = \lambda_0(t)exp(\beta'x) \quad (2.5)$$

Em que “ $\lambda_0(t)$ é uma função do tempo não-especificada, denominada de função de risco de base, e β é um vetor de parâmetros desconhecidos associados aos fatores de risco x ” [8]. A estimação dos parâmetros β da equação “é realizada por meio do método da máxima verossimilhança parcial, sendo o exponencial de cada parâmetro interpretado como uma razão de riscos” [8].

No Capítulo 4 desse trabalho são empregados então o método de Kaplan Meier para traçar as curvas de sobrevivência comparando as categorias para diversas variáveis. Por exemplo, considerando-se uma variável INGRESSO, que diz respeito a forma de ingresso do aluno no curso, pode-se comparar as categorias de forma de ingresso por SISU e por Vestibular. Para as situações multivariadas, analisando-se o efeito de diversas variáveis simultaneamente e o risco proporcional inerente a cada variável no que diz respeito ao evento evasão, é empregada a regressão de Cox. Conforme dito anteriormente, a abordagem paramétrica de análise de sobrevivência, devido à rigidez do modelo e às limitações, não é utilizada nesse estudo.

2.4 Mineração de Dados

A mineração de dados tem uma importância fundamental no processo de descoberta de conhecimento em bases de dados. Nesta seção são explorados o conceito e as tarefas de mineração de dados, as metodologias de mineração de dados e a mineração de regras de associação com o algoritmo Apriori.

2.4.1 Conceito e Tarefas de Mineração de Dados

Quando nos deparamos com uma grande base de dados podem existir ali muitas informações interessantes e relevantes que ainda não são conhecidas, ocultas em meio a uma massa de dados.

De acordo com Silva e Silva “a mineração de dados é uma área de pesquisa multidisciplinar, envolvendo basicamente Banco de Dados, Estatística e Aprendizado de Máquina” [84]. Conforme Fayyad et al.¹⁷ (1996 apud Silva e Silva, 2014) [84] “a mineração de dados é parte principal de um processo que tem como entrada uma Base de Dados e como saída um Conhecimento”. “As técnicas de Data Mining visam atender às aplicações de níveis administrativos mais elevados, tais como: marketing (mala direta direcionada), planejamento de estoque, abertura de novas filiais e outras decisões estratégicas” [79]. Um exemplo de conhecimento obtido a partir de uma técnica de mineração de dados seria “SE (situacao_disciplina_algoritmo = 'reprovado') E (situacao_disciplina_calculo1 = 'reprovado') ENTÃO (evadido='sim')” [79].

A mineração de dados é vista como a principal etapa de um processo mais amplo que é o Knowledge Discovery in Databases (KDD) que significa descoberta de conhecimento em bases de dados [18]. O KDD além da mineração de dados, inclui duas outras fases: “pré-processamento de dados (preparação de dados, abrangendo mecanismos para captura, organização e tratamento dos dados) e pós-processamento dos resultados obtidos na mineração de dados” [18]. Existem diferentes tarefas realizadas por técnicas de mineração de dados [39]: Classificação: Construção de um modelo que possa ser aplicado a dados não classificados para categorizá-los em classes; Estimativa (ou regressão): Define um valor para alguma variável contínua desconhecida; Associação: Usado para verificar as associações entre variáveis em um banco de dados; Segmentação (ou Clustering): Consiste em particionar uma população heterogênea em vários subgrupos mais homogêneos; Sumarização: Consiste em encontrar uma descrição compacta para um subconjunto de dados.

¹⁷Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic. The KDD process for extracting useful knowledge from volumes of data. Communications of the ACM, v. 39, n. 11, p. 27-34, 1996.

2.4.2 Metodologias de Mineração de Dados

Existem três principais metodologias de Data Mining: Cross Industry Standard Process for Data Mining (CRISP-DM), Sample, Explore, Modify, Model, Assessment (SEMMA) [19] e KDD.

O CRISP-DM foi concebido no final de 1996 por veteranos do mercado de mineração de dados: DaimlerChrysler, SPSS e a NCR e sua construção não se deu de forma teórica ou acadêmica, usando princípios técnicos, visto que essas abordagens foram tentadas no passada sem muito êxito, mas foi baseado na experiência prática e real de como as pessoas realizam projetos de mineração de dados [17].

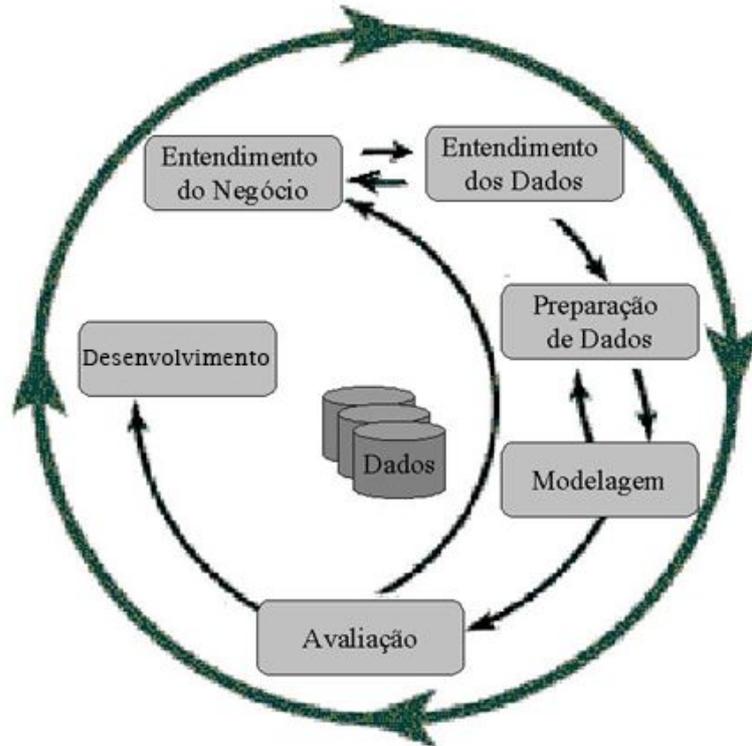
A metodologia CRISP-DM tem uma maior aceitação mundial e é composta por 6 fases [19]:

- **Compreensão do negócio:** é necessário compreender os objetivos de acordo com as regras de negócio, definir o problema a ser tratado com mineração de dados e criar um plano preliminar.
- **Compreensão dos dados:** é necessário conhecer aspectos dos dados como a qualidade dos mesmos, os subconjuntos de dados e as variáveis que interessam e formular hipóteses.
- **Preparação dos dados:** após a compreensão dos dados, procede-se com ações como limpeza dos dados, transformação de variáveis a fim de preparar os dados para aplicação das técnicas de mineração de dados.
- **Modelação:** nessa etapa as técnicas de mineração, devidamente ajustadas, são aplicadas para a obtenção de bons resultados e descoberta de novo conhecimento.
- **Avaliação:** Os resultados obtidos na fase de modelação são analisados. Os passos anteriores também são revisados para se verificar se cumprem os objetivos do negócio. Assim, pode-se tomar a melhor decisão quanto ao uso dos resultados.
- **Desenvolvimento:** Compreende as ações a serem tomadas diante dos resultados obtidos com a mineração de dados já avaliados. Envolve desde a produção de relatórios até a repetição do processo de mineração de dados se assim for necessário. Os conceitos precisam ser apresentados de forma inteligível para o cliente.

A Figura 2.3, retirada do guia passo a passo do CRISP-DM [17] ilustra as fases descritas anteriormente e enfatiza que elas não são rígidas quanto à sequência de execução, pois pode-se avançar ou retroceder nas fases sempre que necessário. As setas na Figura 2.3 indicam as dependências mais importantes e freqüentes entre fases. O círculo exterior simboliza a natureza cíclica da mineração de dados. A mineração de dados não termina

uma vez que tenha sido implementada. As lições aprendidas durante o processo e a solução implementada pode desencadear novas questões:

Figura 2.3: Fases do modelo de Referência CRISP-DM



Devido ao fato de ser um modelo com boa aceitação no mercado, concebido no universo empresarial por empresas diretamente ligadas a tecnologia e com um foco prático, a metodologia CRISP-DM foi adotada nesse trabalho. Maiores detalhes sobre metodologias são discutidas no Capítulo 3 (metodologia).

2.4.3 Mineração de Regras de Associação - Algoritmo Apriori

Para descobrir associações importantes entre os dados, uma opção interessante é o uso de técnicas de mineração de regras de associação. O emprego de mineração de regras de associação nesse trabalho, onde se estuda evasão escolar, é bem apropriado pois se pretende investigar as causas que estão associadas a evasão nos cursos superiores de Computação. Segundo Baker et al. [7] trabalhos envolvendo mineração de dados educacionais é algo relativamente recente. Minaei-Bidgolil et al.¹⁸ (2006 apud Manhães et al., 2011) [64] realizaram um estudo usando regras de associação para extração de padrões em bases de dados a partir de sistemas educacionais. Os autores mostram que um conjunto de regras

¹⁸Minaei-Bidgolil, B. et al. Association analysis for a web-based educational system. Data Mining in E-Learning(Advances in Management Information), v. 4, p. 139-156, 2006.

revelam os atributos associados ao desempenho dos estudantes em uma disciplina de física oferecida on-line. Um outro trabalho utilizando regras de associação em um contexto educacional foi o realizado por Hämäläinen et al.¹⁹ (2004 apud Manhães et al.,2011) [64] que analisaram disciplinas de programação de computadores usando regras de associação e modelos probabilísticos, o que levou os autores a identificar alguns fatores relevantes na predição dos resultados finais nas disciplinas. Santos et al. [81] usaram mineração de regras de associação via algoritmo Apriori para detectar padrões de comportamento dos alunos no estado de ânimo desanimado, usando para isso dados de alunos da disciplina de Análise de Demonstrações Contábeis, ofertada pelo Departamento de Ciências Contábeis e Atuariais da Universidade Federal do Rio Grande do Sul (UFRGS), no semestre de 2014/2, verificando-se que os alunos que se sentiram desanimados, enviaram menos de 50% das atividades e tiveram dificuldades com o conteúdo, o que refletiu no conceito final.

De acordo com Romão et al. [79] “Uma das técnicas mais atraentes é a Mineração de Regras de Associação, que tem como destaque o algoritmo Apriori. Ele pode trabalhar com um número grande de atributos, gerando várias alternativas combinatórias entre eles.” O algoritmo Apriori é portanto muito utilizado em se tratando de mineração de regras de associação em grandes massas de dados, encontrando todos os conjuntos de itens frequentes, chamados itemsets frequentes (Lk) [78]. A mineração de regras de associação tem uma aplicação muito interessante, por exemplo, em se tratando de supermercados, quando se tem um banco de dados com a relação de itens comercializados. Nesse sentido, o algoritmo pode nos dizer se em uma determinada transação de venda de um produto, um outro produto também é vendido pelo supermercado. “O objetivo, então, é encontrar todas as regras de associação relevantes entre os itens, do tipo $X(\textit{antecedente}) \Rightarrow Y(\textit{consequente})$ ” [79]. A mineração de regras de associação tem uma utilidade bem evidente nesse universo de transações comerciais, e essa utilidade pode ser estendida para outros campos, como para análise de dados do ensino. No que tange ao problema em estudo nessa dissertação, que é a evasão nos cursos superiores da área de Computação, podemos verificar, por exemplo, se há associação entre a evasão e o desempenho dos alunos em determinadas disciplinas, se há relação com a forma de ingresso, com o sexo, com o período de saída etc. Um exemplo de regra de associação seria:

Regra 1:

$$\{Evasão = sim\} \Rightarrow \{Reprovado_algoritmo = sim, Reprovado_calculo_1 = sim\}$$

A Regra 1 apresentada é apenas para fins de exemplificação, não compondo os estudos

¹⁹Hämäläinen, Wilhelmiina et al. Data mining in personalizing distance education courses. In world conference on open learning and distance education, Hong Kong, pp. 1–11, 2004

realizados. Ela indica que o aluno que evade também reprova em Algoritmos e também reprova em Cálculo 1. No entanto, para que esteja completa, essa regra ainda necessita de algumas informações, como por exemplo: Qual o percentual de alunos que evade que também reprova em Algoritmos e em Cálculo 1? Essa é uma informação essencial para elucidarmos a força da regra, visto que não faria sentido considerar como uma regra forte se apenas 10% dos alunos evadidos também reprovarem em Cálculo 1 e Algoritmos. Nesse sentido há dois conceitos fundamentais para a mineração de regras de associação: Confiança e Suporte.

O grau de confiança seria a probabilidade de que uma transação que tenha um item, também contenha o outro item. “A toda regra de associação $A \Rightarrow B$ associamos um grau de confiança, denotado por $conf(A \Rightarrow B)$ ” [30]. Portanto, a confiança mede a $P(B|A)$, ou seja, a frequência dentre as transações contendo A que também contêm B. Para a Regra 1 exibida como exemplo, se dissermos que confiança é de 60%, isso indica que 60% daqueles que evadiram também reprovaram em Algoritmos e em Cálculo 1. Sendo assim, a confiança evidencia a força da regra. A Fórmula 2.6 formaliza a definição de confiança [30]:

$$conf(A \Rightarrow B) = \frac{\text{número de transações que contêm } A \text{ e } B}{\text{número de transações que contêm } A} \quad (2.6)$$

O grau de suporte corresponde a proporção de transações que contém os itens. “Em termos estatísticos, o suporte de uma regra $X \Rightarrow Y$ é a probabilidade $P(X \cup Y)$, onde $X \cup Y$ indica que a transação contém os dois conjuntos de itens X e Y” [30]. Sendo assim, para a Regra 1 exibida como exemplo, se dissermos que o suporte é de 30%, indica que em todo o banco de dados a ocorrência de evadidos = sim, de reprovados em Algoritmos = sim e de reprovados em Cálculo 1 = sim corresponde a 30% do total de transações verificados nesse banco de dados.

Ao se usar o algoritmo Apriori para minerar regras de associação, é importante que se defina uma confiança mínima para não se obter regras sem peso para o estudo. Essa confiança, conforme descrito anteriormente, não deve ser muito pequena, sob pena de obter várias regras inexpressivas, além de obter como saída do algoritmo uma enxurrada de regras que posteriormente teriam que ser filtradas para separar as mais relevantes. É interessante que se defina também um suporte mínimo. Contudo, pode ser interessante definir suportes baixos, inferiores ao valor da confiança, visto que algumas regras específicas, interessantes e menos perceptíveis podem ter um valor de suporte pequeno.

Um outro aspecto importante que se deve ressaltar é que, ainda que uma regra obtida apresente uma confiança elevada, isso não garante que essa regra seja interessante para o estudo. Vamos tomar como exemplo a Regra 2 apresentada:

Regra 2:

$\{Evadido = sim, Cursou_1_4_periodos = sim\} \Rightarrow \{Cursou_11_mais_periodos = não\}$

Confiança: 100%

Suporte: 40%

A Regra 2 apresentada é apenas para fins de exemplificação, não compondo os estudos realizados. A regra poderia ser lida da seguinte forma: 100% dos alunos que evadem e que cursam apenas entre 1 e 4 períodos, não cursam 11 ou mais períodos. Embora essa regra tenha uma confiança máxima, de 100%, ela não agrega nenhuma informação relevante ao estudo conduzido, visto que se um aluno cursou apenas de 1 a 4 períodos, é evidente que ele não cursou 11 ou mais períodos. Regras como essa podem ser obtidas como saída do algoritmo Apriori e devem ser omitidas ou descartadas do estudo, pois não representam um novo conhecimento.

Considerando-se um banco de dados contendo informações sobre os alunos evadidos nos cursos de graduação de uma determinada universidade em um período de 10 anos e tendo-se como objetivo obter associações entre os dados desse banco de dados que tragam informações relevantes sobre o problema da evasão, vamos considerar os itens descritos a seguir [79] :

- $I = \{i_1, i_2, \dots, i_m\}$ é um conjunto de literais, denominados itens. São as variáveis associadas aos alunos. Exemplo: forma de ingresso, cotista, situação na disciplina x, sexo etc;
- T é um determinado conjunto de itens de um aluno, tal que $T \subseteq I$;
- D é uma tabela representando todas as características e atributos de todos alunos; e
- X, Y são conjuntos de itens específicos dos alunos, tal que $X \subseteq T$ e $Y \subseteq T$

Considerando-se os itens descritos pode-se dizer que: “Uma regra de associação é uma implicação da forma $X \Rightarrow Y$, onde $X \subset I$, $Y \subset I$ e $X \cdot Y = \emptyset$. A regra $X \Rightarrow Y$ pertence a D com confiança c se $c\%$ dos registros em D que contém X também contém Y . A regra $X \Rightarrow Y$ tem suporte s em D se $s\%$ dos registros em D contém $X \cup Y$ ” [79].

O algoritmo Apriori usa duas funções, chamadas de Apriori_gen e de Genrules, sendo que a primeira gera os conjuntos de itens candidatos eliminando os conjuntos de itens não frequentes e a segunda extrai as regras de associação [78]. “O algoritmo Apriori é executado de forma iterativa: os itemsets frequentes de tamanho k são calculados a partir dos itemsets frequentes de tamanho $k-1$ que já foram calculados no passo anterior (a partir dos itemsets frequentes de tamanho $k-2$, etc) [30]”. Para descrever resumidamente

o funcionamento do algoritmo Apriori, vamos considerar que estejamos no passo k e que já tenhamos obtido no passo anterior o conjunto L_{k-1} dos itemsets frequentes de tamanho $k-1$:

Tabela 2.10: Etapas de execução do algoritmo Apriori

Passo	Nome da fase	Descrição
k-1	Definição dos itemsets frequentes de tamanho unitário	Determina os conjuntos de itens frequentes de tamanho unitário (1-itemsets frequentes).
k	Geração dos candidatos	Os itemsets frequentes do passo $k-1$ (L_{k-1}) são usados para gerar os conjuntos de itens potencialmente frequentes, os itemsets candidatos (C_k). A geração dos itemsets candidatos usa como argumento os itemsets frequentes L_{k-1} , usando a função Apriorigen, que retorna um superconjunto de todos os k -itemsets frequentes. Une-se então L_{k-1} com L_{k-1} .
	Poda dos candidatos	São eliminados os itemsets c_k que pertencem a C_k , desde que um dado $(k-1)$ -subset de c_k não pertença a L_{k-1} , onde C_k é o conjunto de k -itemsets candidatos e c_K corresponde a cada membro desse grupo.
	Cálculo do suporte	É feita a descoberta das regras de associação via algoritmo Genrules. A geração de regras, para qualquer itemset frequente, significa encontrar todos os subsets não vazios de L . Assim, para todo e qualquer subset A , produz-se uma regra em que se A então $(L - A)$ somente se a razão (suporte (L) /suporte (A)) é ao menos igual a confiança mínima estabelecida pelo usuário.

Fonte: De Amo [30] e Romão et al. [79]

O algoritmo Apriori é usado no Capítulo 4 com o intuito de descoberta de conhecimento e para validar as verificações feitas via análise de sobrevivência. Quando à confiança e ao suporte mínimo a serem utilizados, isso é abordado com maiores detalhes no Capítulo 3 (metodologia).

2.5 Classificação dos Cursos Superiores Adotada pelo INEP

O INEP classifica os cursos superiores no censo da educação superior em 8 grandes áreas do conhecimento, usando a Classificação Internacional Padronizada da Educação (ISCED - *International Standard Classification of Education*) proposto pela OCDE [33], sendo adotada essa classificação desde o ano 2000. A adoção de um modelo de classificação internacional permite comparar internacionalmente as estatísticas brasileiras da educação superior, além da classificação OCDE ser flexível e permitir a adaptação às características e variedades dos cursos brasileiros [33]. As oito áreas gerais do conhecimento (ou oito grandes áreas) estão descritas a seguir:

- Educação
- Humanidades e Artes
- Ciências Sociais, Negócios e Direito
- Ciências, Matemática e Computação
- Engenharia, Produção e Construção
- Agricultura e Veterinária
- Saúde e Bem Estar Social
- Serviços

Cada uma dessas grandes áreas é dividida em áreas detalhadas e cada uma dessas áreas detalhadas, por sua vez, é dividida em áreas específicas. Como o foco desse estudo são os cursos de Computação, convém detalhar um pouco mais a grande área de Ciências, Matemática e Computação (que é onde estão classificados os cursos de Computação). A Tabela 2.11 exhibe as áreas Específicas e Detalhadas da área geral de Ciências, Matemática e Computação [33]:

Tabela 2.11: Grande área de Ciências, Matemática e Computação

Área Geral	Áreas Específicas	Áreas Detalhadas
Ciências, Matemática e Computação	Ciências da Vida	Ciências da vida (Curso gerais)
		Biologia e Bioquímica
		Ciências Ambientais
	Ciências Físicas	Ciências Físicas (Curso gerais)
		Física
		Química
		Ciência da Terra
	Matemática e Estatística	Matemática
		Estatística
	Computação	Ciência da Computação
		Uso do Computador
		Processamento da Informação

Dentre as áreas relacionadas na Tabela 2.11, algumas são voltadas mais para estudos biológicos e químicos. Outras tem um enfoque mais matemático, como Matemática, Física, Estatística e a própria Computação. A evasão para essas áreas é detalhada no Capítulo 4 (estudos realizados).

Capítulo 3

Metodologia

Neste capítulo é descrita a metodologia de estudo empregada, discorrendo como os dados foram tratados e analisados. Os estudos apresentados no Capítulo 4 foram feitos a partir de dados fornecidos pelo INEP e de dados fornecidos pela UnB. A metodologia de estudo aplicada ao longo deste trabalho segue as etapas previstas no CRISP-DM.

Em relação aos testes estatísticos, “os níveis de significância usualmente adotados são 5%, 1% e 0,1%” [73]. Considera-se, nos estudos apresentados, o p-valor (nível de significância) $< 0,05$ aceitável para avaliar a significância dos resultados obtidos. Sendo assim, níveis de significância inferiores a esse valor foram considerados significantes. Ressalta-se que a significância de 0,05 adotada apenas dá uma evidência de uma hipótese ser ou não rejeitada, dado o nível de significância escolhido. Sendo assim, o p-valor não é usado nesse trabalho de forma taxativa para afirmar algo com convicção.

3.1 Tratamento e Análise dos Dados do INEP

Foram analisados os microdados do censo da educação superior fornecidos pelo INEP, disponibilizados para download em <http://portal.inep.gov.br/basica-levantamentos-acessar>. Existem no portal do INEP dados do censo da educação superior disponíveis desde 1995. O mais recente disponibilizado para download são os dados referentes ao ano de 2014.

Nessa dissertação o período estudado usando os dados do INEP é de 2010 a 2014. Os dados do censo da educação superior fornecidos pelo INEP passaram a ser individualizados a partir de 2009, conforme o resumo técnico do Censo da Educação Superior de 2009 [34]. Como nesse estudo pretende-se acompanhar o status de cada aluno ao longo do curso, para realizar por exemplo a Análise de Sobrevivência, é necessário dispor de dados individuais, não fazendo sentido, portanto, trabalhar com dados anteriores a 2009, pois estes eram agregados. Os dados de 2009 não foram usados visto que houve algumas

mudanças metodológicas no censo da educação superior em 2010 em relação a 2009 que poderiam causar problemas na interpretação dos dados.

De acordo com o resumo técnico do censo da educação superior de 2010 [35], em 2009, o atributo grau acadêmico dos cursos de graduação, no censo, permitia que um mesmo curso fosse declarado simultaneamente como sendo bacharelado e licenciatura e, diante disso, a partir de 2010, os cursos que anteriormente era classificados como bacharelado e licenciatura passaram a serem classificados em uma das seguintes opções: como dois cursos, sendo um bacharelado e outro licenciatura; como dois cursos, sendo um de licenciatura, outro de bacharelado e uma Área Básica de Curso; como apenas um curso, sendo bacharelado ou licenciatura. Diante desse ajuste metodológico e do fato de que as taxas de evasão são apresentadas também por cursos, os dados de 2009 não foram considerados.

Os microdados do censo da educação superior que o INEP disponibiliza em seu portal para download, possuem informações, para cada ano, sobre os cursos ofertados no país, sobre os alunos, sobre os docentes, sobre as instituições de ensino e sobre os locais de oferta dos cursos, disponibilizando um arquivo para cada uma dessas categorias de informações. foram utilizados os arquivos referentes aos cursos, aos alunos e às instituições de ensino. Os arquivos referentes aos anos de 2010, 2011 e 2012 estão disponíveis em formato TXT, juntamente com um arquivo em formato SPSS e outro em formato SAS contendo instruções para leitura das variáveis e um script para leitura das variáveis. Para os anos de 2013 e 2014 os arquivos foram disponibilizados em formato CSV.

Para permitir a compreensão dos dados, o INEP disponibiliza em cada censo da educação superior um arquivo “LEIA-ME” contendo o significado das variáveis e seus possíveis valores. Como nos censos do período 2010-2014 cada aluno possui um código único gerado pelo próprio INEP que permite o seu acompanhamento ao longo do curso, e existe uma variável identificando o status do aluno no curso (se cursando, formado, transferido, desvinculado etc) é possível verificar a taxa de evasão anual em diferentes cursos e também acompanhar o tempo de permanência do aluno no curso via Análise de Sobrevivência. Infelizmente, o INEP não dispõe de informações sobre o desempenho do aluno no curso (notas em disciplinas cursadas). Portanto, a associação entre evasão e desempenho do aluno no curso não é possível via dados do INEP.

Para prover a preparação dos dados, o software Excel foi utilizado para limpeza de variáveis não relevantes ao estudo. Os arquivos de aluno, em cada ano, possuem pouco mais de 90 variáveis, os arquivos de curso pouco mais de 70 e os arquivos das instituições de ensino superior pouco mais de 40. A tabela de alunos, em 2010, possui 8337219 registros, em 2011 possui 8961724 registros, em 2012 são 9565483 registros, em 2013 9929289 registros e em 2014 10793935. Os dados foram transpostos para o banco de dados MYSQL. As variáveis que foram usadas no estudo são descritas na Tabela 3.1:

Tabela 3.1: Variáveis usadas nos estudos com dados do INEP

Variável	Descrição
co_aluno	Código de identificação gerado pelo INEP para o aluno
co_aluno_situacao	Situação do aluno no curso
in_sexo_aluno	Sexo do aluno
co_grau_academico	Grau acadêmico do aluno
ano_ingresso	Ano de ingresso do aluno
co_curso	Código único de identificação do curso
no_curso	Nome do curso
co_ocde	Código de identificação do curso segundo OCDE
no_ocde	Nome do curso segundo OCDE
no_ocde_area_geral	Nome da área geral segundo OCDE
co_ocde_area_geral	Código da área geral segundo OCDE
no_area_especifica	Nome da área específica segundo OCDE
co_ocde_area_especifica	Código da área específica segundo OCDE
no_area_detalhada	Nome da área detalhada segundo OCDE
co_ocde_area_detalhada	Código da área detalhada segundo OCDE
tp_atributo_ingresso	Identifica se o aluno é ingresso pela Área Básica de Ingresso
co_nivel_academico	Código do nível acadêmico do curso
in_reserva_vagas	Informa se o aluno ingressou por meio de reserva de vagas
in_ing_vestibular	Informa se o aluno ingressou por meio de vestibular
in_ing_enem	Informa se o aluno ingressou por meio de ENEM
inscritos	Informa a quantidade de inscritos no curso
vagas	Informa a quantidade de vagas no curso
tempo	Informa o tempo de permanência do aluno no curso
co_ies	Código da instituição de ensino superior
no_ies	Nome da instituição de ensino superior

Algumas variáveis apresentadas na Tabela 3.1 foram criadas a partir de variáveis existentes nas tabelas. A variável tempo foi criada contando-se o tempo de permanência do aluno no curso e foi usada nos estudos de Análise de Sobrevivência. A variável inscritos e vagas foram obtidas a partir da soma de todos os inscritos e de todas as vagas em cada curso, pois o INEP tem várias variáveis de contagem de inscritos e de vagas de acordo com o turno em que o curso é ofertado (matutino, vespertino ou noturno) e de acordo com a modalidade de ensino (presencial ou a distância).

Alguns filtros foram utilizados no MYSQL antes de se calcular as taxas anuais de evasão e aplicar as técnicas de Análise de Sobrevivência. De acordo com os arquivos

“LEIA-ME” que acompanham cada censo da educação superior, é necessário o uso do filtro `tp_ atributo_ ingresso=0` ao se classificar estatísticas por área de conhecimento. O valor 0 no filtro, nesse caso, indica que o curso não se trata de uma área básica de ingresso. As áreas básicas de ingresso não correspondem a cursos de bacharelado, licenciatura ou tecnólogo. A área básica de ingresso consiste em cursar um conjunto básico de disciplinas (denominado de “ciclo básico” por algumas instituições de educação superior) e após isso, escolher de uma entre duas ou mais formações acadêmicas [36]. Portanto, as áreas básicas de ingresso não foram consideradas nesse estudo. Um outro filtro usado nesse estudo foi `co_ nivel_ academico=1`, que corresponde aos cursos superiores de nível acadêmico graduação. O censo da educação superior traz também dados dos cursos de nível acadêmico sequencial de formação específica. Os cursos sequenciais de formação específica também são cursos superiores assim como os de graduação. Eles são destinados “a atualização de qualificações técnicas, profissionais, acadêmicas ou de desenvolvimento intelectual”. Entretanto, diferente dos cursos de graduação, eles não conferem diplomas de bacharel, licenciado ou tecnólogo e portanto, não permitem que o aluno ingresse em uma pós-graduação. Os cursos sequenciais de formação específica fogem ao escopo desse trabalho. Só foram considerados nesse estudo os cursos de graduação.

A variável `co_ aluno_ situação` foi usada para verificar se o aluno evadiu ou não. Esta variável pode assumir os seguintes valores: cursando, matrícula trancada, formando, transferido para outro curso da mesma instituição, desvinculado e falecido. A quantidade de alunos evadidos em cada ano foi calculada como a quantidade de matrículas desvinculadas acrescida da quantidade de matrículas de alunos transferidos no ano, em cada curso/área em avaliação. Os alunos com status desvinculado são aqueles que abandonaram o curso antes da conclusão, quer seja por iniciativa própria, quer seja por políticas institucionais como reprovações sucessivas ou jubramento. As matrículas totais foram obtidas somando-se a quantidade de matrículas de todos os alunos para cada ano e para cada curso/área. As taxas de evasão anuais foram obtidas a partir de scripts MYSQL, sendo a relação entre o número de evadidos e o número de matrículas totais. Neste estudo, transferências são consideradas evasões mesmo que não ocorra evasão da instituição, pois trabalha-se com o conceito de evasão de curso. A Fórmula 3.1 foi utilizada para o cálculo do percentual de evasão anual:

$$E = (M_{desvinculados} + M_{transferidos}) / M_{totais} \cdot 100 \quad (3.1)$$

Onde: E corresponde ao percentual de evasão, M_{totais} corresponde ao somatório de todas as matrículas de alunos.

Além das variáveis descritas na Tabela 3.1 obtidas a partir dos dados do INEP, foi também incluída uma variável usando dados da Coordenação de Aperfeiçoamento de Pes-

soal de Nível Superior (CAPES), para verificar se a instituição oferta ou não curso de pós-graduação na área de Computação. Essa variável foi utilizada na Análise de Sobrevivência para verificar se as instituições que oferecem verticalização do ensino com cursos de pós-graduação na área de Computação apresentam sobrevivência nos cursos de graduação em Computação diferente daquelas que não oferecem essa verticalização. Os dados sobre oferta de pós-graduação nos cursos de Computação foram obtidos no endereço <https://sucupira.capes.gov.br/sucupira/public/consultas/coleta/programa/quantitativos/quantitativoIes.jsf?areaAvaliacao=2&areaConhecimento=10300007>.

As taxas de evasões anuais são exibidas via *Box-plot* e separadas por mantenedoras institucionais públicas ou privadas. São consideradas as instituições privadas com fins lucrativos e as instituições privadas sem fins lucrativos. São consideradas instituições públicas as federais, estaduais e municipais. O software utilizado para geração dos gráficos foi o SPSS Statistics versão 22.

A relação entre a concorrência e a evasão é exibida via gráfico de dispersão e é aplicado o teste estatístico não paramétrico Rô de Spearman para testar a correlação entre a concorrência e a evasão. O software utilizado para obtenção dos resultados é o SPSS Statistics. O teste não paramétrico foi adotado em relação ao paramétrico pois algumas variáveis usadas não mostraram distribuição normal. A concorrência é obtida a partir da relação candidatos/vagas nos cursos.

A Análise de Sobrevivência foi feita utilizando-se o método de Kaplan-Meier e a regressão de Cox, sendo aplicadas para alunos dos cursos de Computação ingressantes no ano de 2010, acompanhando-se esses alunos até o ano de 2014 (período de 5 anos). Foram avaliados um total de 23696 registros de alunos de escolas públicas e 80027 registros de alunos de escolas privadas. O software utilizado para obtenção dos resultados foi o SPSS Statistics versão 22. Os resultados são separados por mantenedora institucional (pública ou privada). Foram aplicados 3 testes estatísticos na Análise de Sobrevivência: Logrank, Breslow e Tarone-Ware.

3.2 Tratamento e Análise dos Dados da UnB

Foram analisados os dados fornecidos pela Secretaria de Assuntos Acadêmicos (SAA) da UnB no período 2005-2015, referentes aos seguintes cursos superiores de Computação: Bacharelado em Ciência da Computação, Licenciatura em Computação, Engenharia de Computação e Engenharia de Software. A base de dados foi fornecida contendo registros de 916 alunos do curso de Licenciatura em Computação, 903 alunos de Ciência da Computação, 531 alunos de Engenharia de Computação e 525 alunos de Engenharia de Software totalizando 2875 registros. Há uma quantidade menor de registros nos cursos

de Engenharia de Computação e em Engenharia de Software pois esses cursos são mais recentes: iniciaram respectivamente nos anos de 2009 e de 2008.

A base de dados foi fornecida pela UnB em formato XLSX com total de 75 variáveis. A preparação dos dados, incluindo a limpeza, criação de variáveis e discretização, foi feita com o software Excel. Após esse tratamento, foram usadas as variáveis descritas na Tabela 3.2:

Tabela 3.2: Variáveis usadas nos estudos com dados da UnB

Variável	Descrição
numero_matricula	Número de matricula do aluno no curso
nome	Nome do aluno
evadido	Situação do aluno se evadido ou não
periodos_cursados	Número de períodos cursados
forma_ingresso	Forma de ingresso do aluno
sexo	Sexo do aluno
cotista	Informa se o aluno ingressou por cotas ou sistema universal
reprovado_calculo_1	Informa se o aluno reprovou na disciplina Cálculo 1
reprovado_algoritmo	Informa se o aluno reprovou em disciplinas de algoritmos
curso	Nome do curso do aluno
telefone	Número de telefone do aluno
email	Endereço de mail do aluno

Além de ser feita uma limpeza na base, algumas variáveis foram discretizadas para facilitar os estudos pois possuíam uma série de valores:

- evadido: foi discretizada com dois valores possíveis: 0, correspondendo a não evadido e 1 a evadido;
- forma_ingresso: 1 para ingresso via vestibular, 2 para ingresso via SISU, 3 para ingresso via PAS e 4 para outras formas de ingresso;
- cotista: 1 para cotista e 0 para não cotista;
- reprovado_calculo_1: 1 para aqueles que sofreram uma ou mais reprovações em Cálculo 1 e 0 para os que nunca reprovaram em Cálculo 1.
- reprovado_algoritmo: 1 para aqueles que sofreram uma ou mais reprovações em disciplinas de algoritmos e 0 para os que nunca reprovaram em Algoritmos.

A Análise de Sobrevivência foi feita utilizando-se o método de Kaplan-Meier e a regressão de Cox, sendo aplicadas para alunos dos cursos de Computação que ingressaram

no período 2005-2015 (período de 11 anos). O tempo nas curvas de sobrevivência é exibido em períodos cursados (períodos de 6 meses). O software utilizado para obtenção dos resultados foi o SPSS Statistics. Foram aplicados 3 testes estatísticos na Análise de Sobrevivência: Logrank, Breslow e Tarone-Ware.

Além da Análise de Sobrevivência, foi aplicado o algoritmo Apriori para verificação de associação entre as variáveis na base de dados fornecida pela UnB. A confiança mínima para execução do algoritmo foi fixada em 70%. Quanto ao suporte, fixou-se inicialmente o suporte mínimo de 1% e foi-se aumentando gradativamente conforme as regras obtidas não apresentassem relevância. Suportes pequenos podem revelar regras específicas e confianças maiores revelam regras fortes. Para execução da mineração de regras de associação foi usado o software R versão 3.1.2.

Para levantamento das causas de evasão foi conduzido um estudo de caso com os alunos evadidos da UnB. Para contatar os alunos evadidos, foi verificado, junto a Secretaria de Registros Acadêmicos da UnB, os nomes e endereços de e-mail de todos os alunos evadidos nos cursos de Computação avaliados, no período 2005 – 2015. Nos cursos em que não se obteve sucesso no contato via e-mail, tentou-se o contato telefônico. Foram considerados evadidos aqueles discentes que foram desligados do curso antes de concluí-lo, por qualquer motivo exceto por motivos de falecimento. Para cada um desses alunos foi enviado um e-mail explicando os motivos da pesquisa e um link para preenchimento do formulário que foi disponibilizado online via Formulários Google. Importante ressaltar que o formulário dessa pesquisa não requereu identificação do aluno.

Para verificação dos motivos da evasão, foram disponibilizados 21 itens em um formulário de escala Likert, onde os alunos deveriam responder, em uma escala de 0 a 5 (6 níveis), como cada item influenciou a sua saída do curso, sendo 0 correspondente a nenhuma influência e 5 correspondente a grande influência. Os participantes também puderam apontar outros motivos que os levaram a evadir, caso julgassem como relevante alguma causa não apontada no formulário de escala. Também puderam responder quais as disciplinas que consideraram mais difíceis ao longo do curso e o porquê de terem escolhido cursar um curso superior na área de Computação.

Para análise dos dados do formulário Likert, foi usado o teste de Kruskal Wallis, com o teste post hoc de Dunn para agrupar os grupos homogêneos, com intervalo de confiança a 95%. O teste não paramétrico foi adotado pois as variáveis não atenderam a normalidade.

Outros fatores verificados foram sexo, formas de ingresso, período de saída do curso, recebimento ou não de algum auxílio estudantil, se o curso foi o primeiro curso de formação superior do aluno.

Capítulo 4

Estudos Realizados

Este capítulo trata dos estudos realizados sobre a evasão nos cursos superiores de Computação e é dividido em duas Seções: 4.1 Estudo dos Microdados do INEP - trata dos estudos realizados usando os censos da educação superior disponibilizados pelo INEP em seu portal; 4.2 Estudo dos dados da UnB - Trata dos estudos realizados com os dados fornecidos pela UnB.

4.1 Estudo dos Microdados do INEP

Nesta seção a evasão e a concorrência são apresentadas para o período 2010-2014. São exibidas as evasões anuais nas grandes áreas, na área geral de Ciências, Matemática e Computação e na área de Computação além da concorrência. São exibidas também as curvas de sobrevivência para os cursos de Computação levando-se em conta os seguintes fatores: sexo, formas de ingresso, ingresso por cota ou sistema universal e oferta ou não de pós-graduação em Computação na instituição.

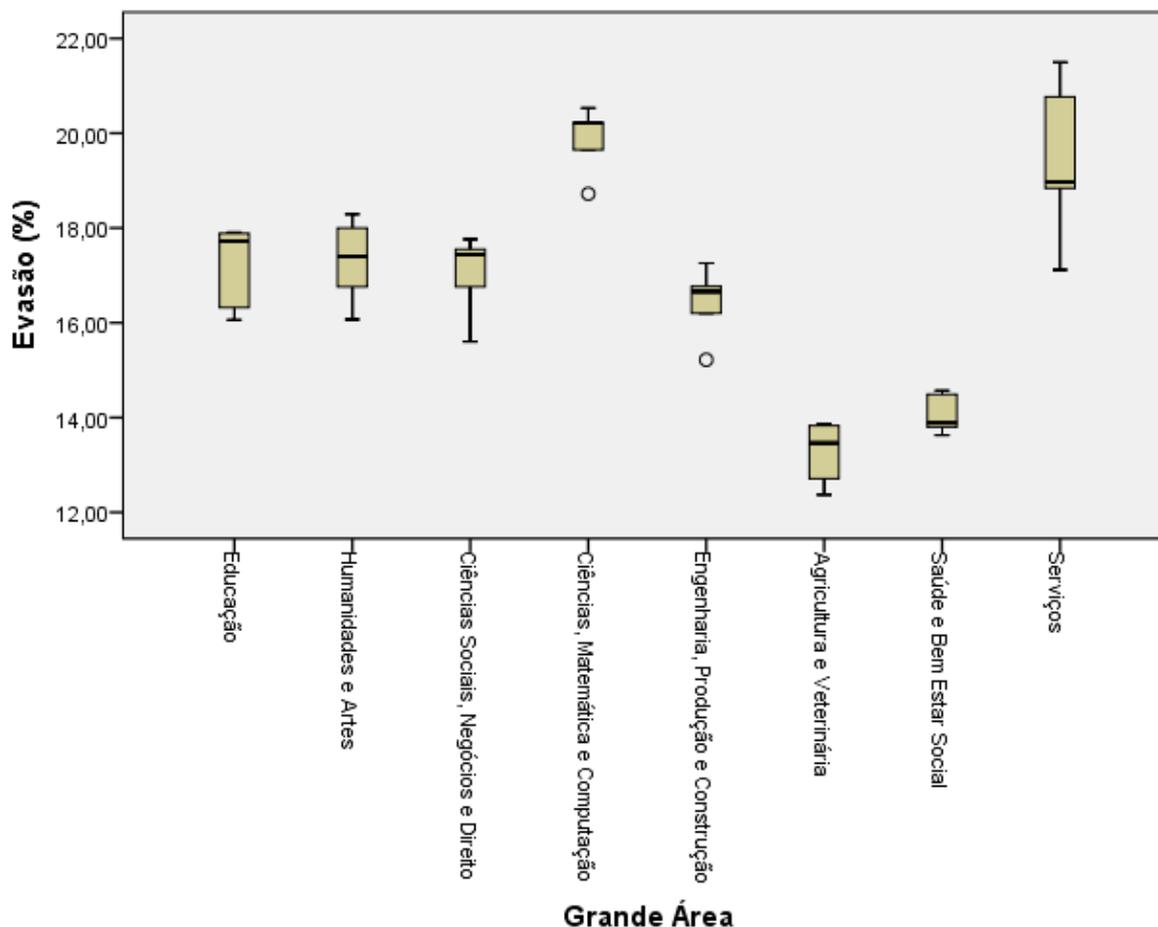
4.1.1 Evasão e Concorrência

Nesta subseção são apresentadas as taxas anuais de evasões e a concorrência (relação candidatos/vagas) para as grandes áreas do conhecimento. Após a comparação da evasão nessas grandes áreas, é feito um detalhamento para a grande área de Ciências, Matemática e Computação e posteriormente para a área de Computação.

Os Diagramas apresentados a seguir foram gerados a partir das taxas de evasão anuais e relações candidatos/vagas que estão disponíveis no apêndice A dessa dissertação. O Eixo X do gráfico traz a relação das grandes áreas ou cursos avaliados e o eixo Y traz os percentuais de evasão. Foi considerada a evasão anual nos anos de 2010 a 2014.

A Figura 4.1 ilustra a taxa de evasão nas áreas gerais do conhecimento, nas instituições privadas, de acordo com a classificação OCDE [33] adotada pelo INEP e foi gerada a partir dos valores de evasão da Tabela A.1 presente no apêndice:

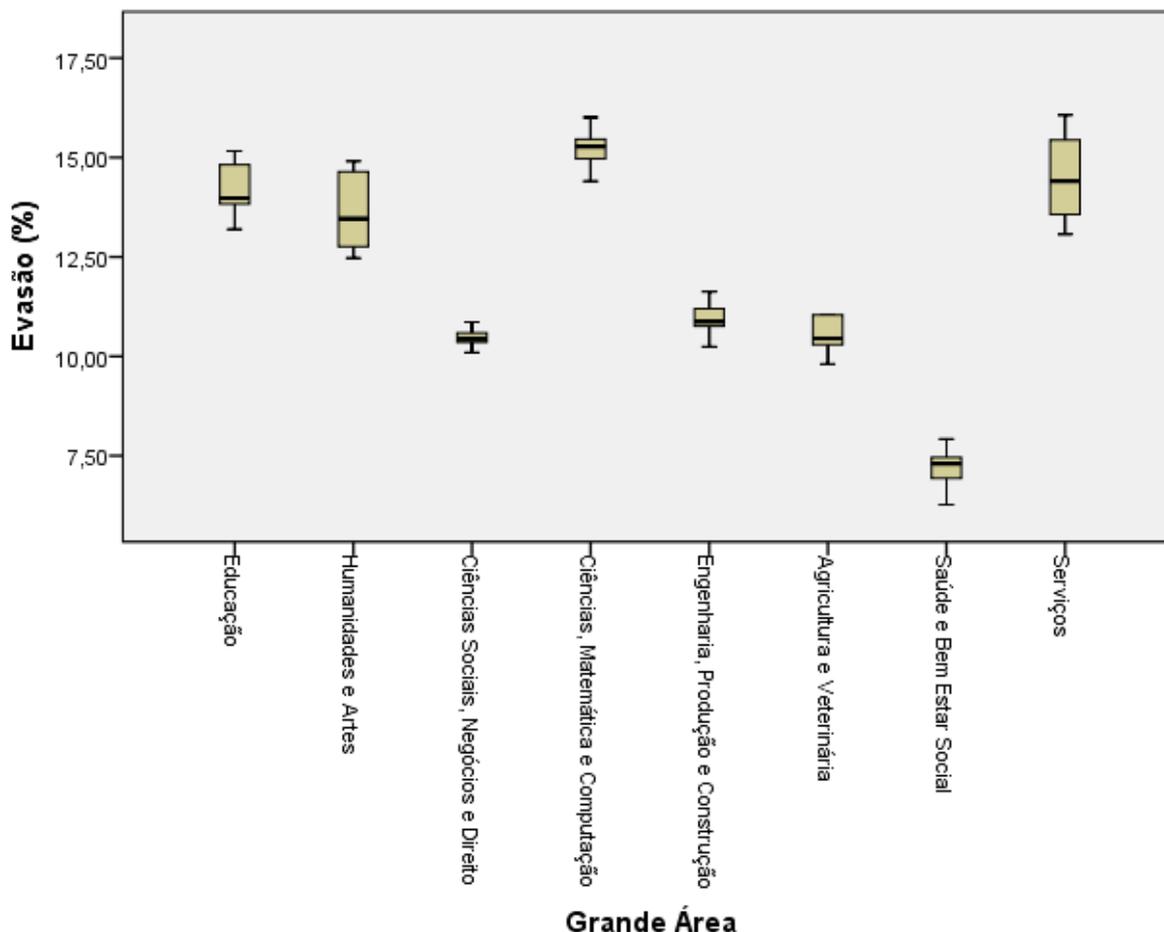
Figura 4.1: Evasão nas grandes áreas do conhecimento (instituições privadas)



Olhando para a Figura 4.1 é fácil perceber que a maior evasão no período 2010-2014 corresponde à áreas de Ciências, Matemática e Computação e à área de Serviços. A evasão nestas duas áreas não diferiu muito, visto que as suas caixas encontram-se aproximadamente na mesma altura no gráfico. A área de Ciências Matemática e Computação apresenta um valor atípico de evasão inferior aos demais sendo a taxa de 18,72% referente ao ano de 2010. Como se pode perceber, a variabilidade dos valores de evasão é maior para a grande área de Serviços (apresenta caixa maior) do que para a grande área de Ciências, Matemática e Computação.

A Figura 4.2 ilustra a taxa de evasão nas áreas gerais do conhecimento, nas instituições públicas e foi gerada a partir dos valores de evasão da Tabela A.1 presente no apêndice:

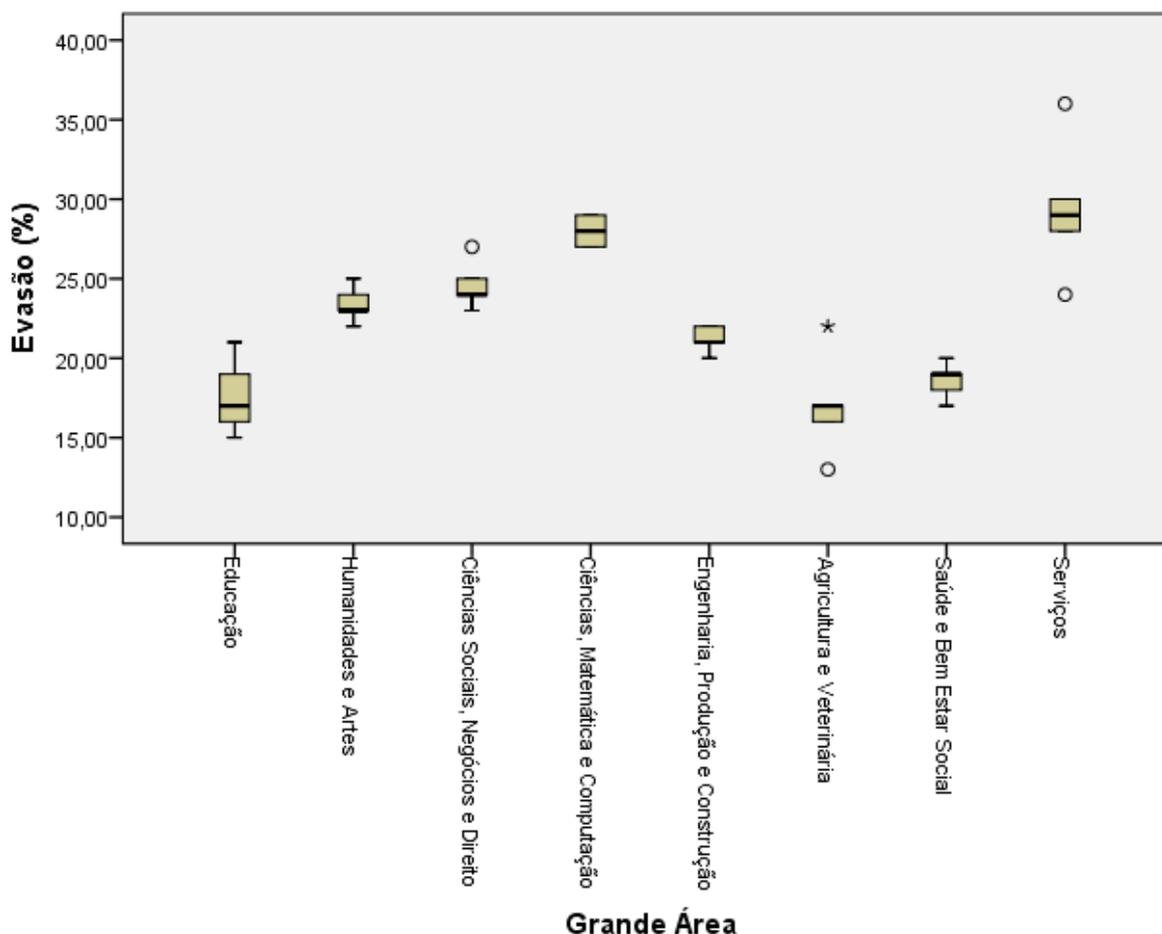
Figura 4.2: Evasão nas grandes áreas do conhecimento (instituições públicas)



Tal como foi verificado nas instituições privadas, nas instituições públicas a área de Ciências, Matemática e Computação e a área de Serviços apresentaram as maiores taxas de evasão no período 2010-2014, pois suas caixas estão posicionadas mais acima no gráfico em relação às demais. A variabilidade da evasão é maior na grande área de Serviços em relação à grande área de Ciências, Matemática e Computação.

A Figura 4.3 mostra o digrama *Box-plot* construído a partir de informações apresentadas pelo instituto Lobo[88] exibindo as taxas de evasão anuais no período 2001-2005 em todas as instituições de ensino superior (considerando as públicas e as privadas). O gráfico foi construído a partir da Tabela A.2:

Figura 4.3: Evasão nas grandes áreas do conhecimento (instituto Lobo)

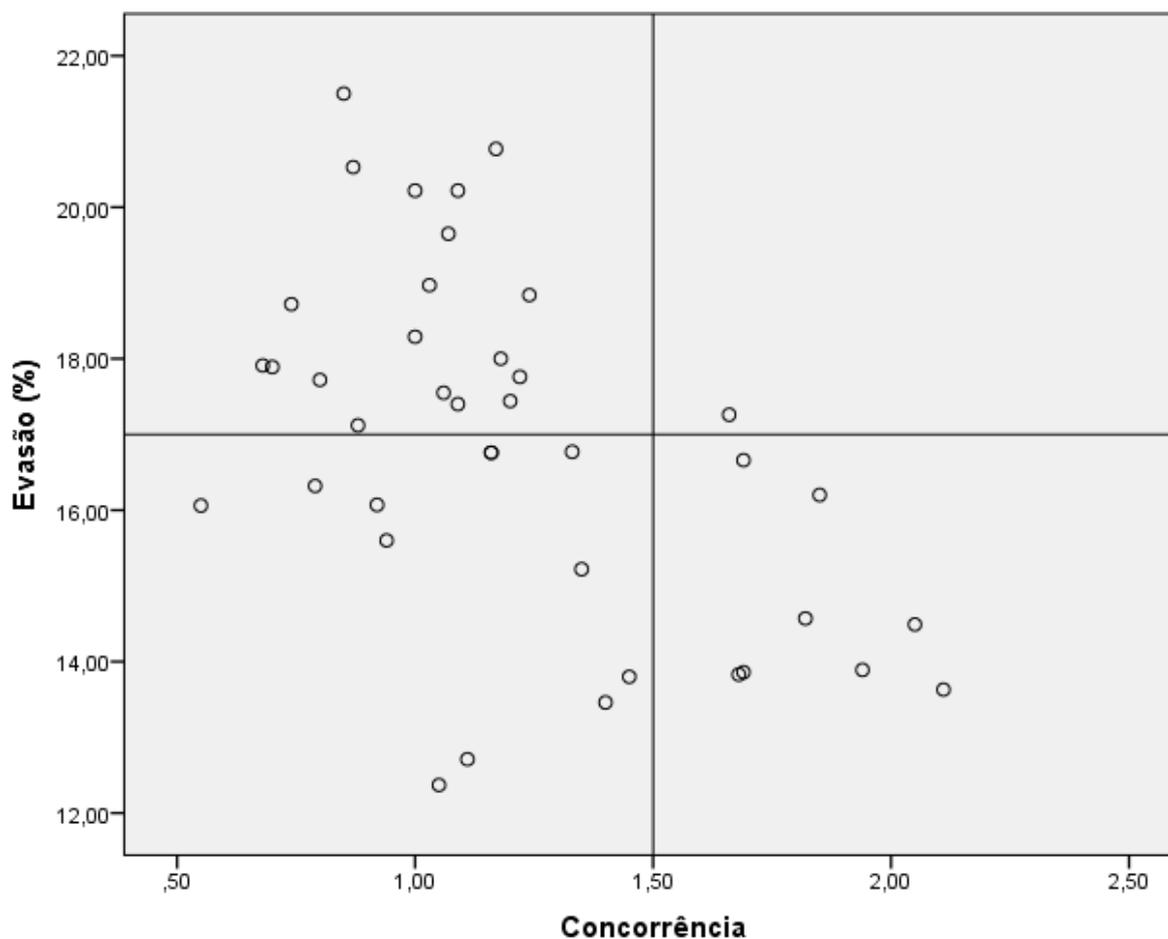


Tal como foi verificado para as instituições públicas e privadas no período 2010 a 2014, os estudos realizados pelo instituto Lobo também revelam que as maiores taxas de evasão no período 2001 a 2005 foi da grande área de Serviços e da grande área de Ciências, Matemática e Computação.

Os gráficos exibidos até aqui mostram uma evidência de que a evasão é maior na grande áreas de Ciências Matemática e Computação e na área de Serviços em relação às outras grandes áreas. Pelo posicionamento das caixas nos gráficos para essas duas grandes áreas, não se pode dizer que a evasão entre elas seja diferente, visto que a altura delas nos gráficos é muito semelhante.

A Figura 4.4, que foi criada a partir da Tabela A.1, mostra, a partir de um gráfico de dispersão, a correlação entre evasão e concorrência (relação candidatos/vagas) no período 2010 a 2014 nas instituições privadas:

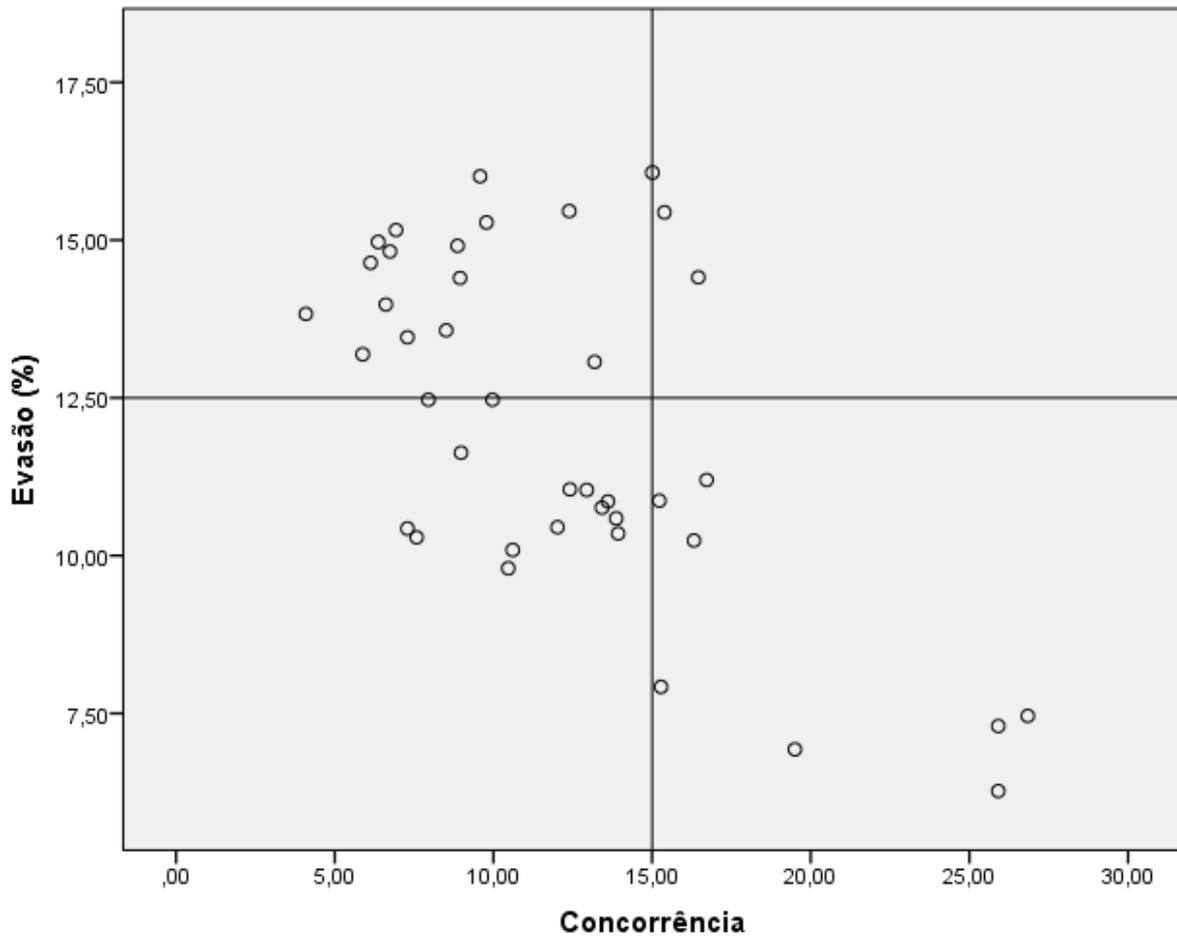
Figura 4.4: Evasão vs Concorrência (instituições privadas)



A Figura 4.4 foi dividida em 4 quadrantes para facilitar a interpretação. Percebe-se que a maioria dos casos de maior concorrência (acima de 1,5 candidato/vaga) apresentam menores taxas de evasão (abaixo de 17%) e a maioria dos casos onde se verifica menor concorrência (abaixo de 1,50 candidato/vaga) a evasão é maior (acima de 17%). O gráfico mostra então uma evidência de uma certa correlação inversa entre concorrência e evasão.

A Figura 4.5 mostra, também a partir de um gráfico de dispersão, a correlação entre evasão e concorrência no período 2010-2014 para as instituições públicas:

Figura 4.5: Evasão vs Concorrência (instituições públicas)



Tal como foi verificado nas instituições privadas, nas públicas a evasão e a concorrência mostram uma certa correlação inversa, conforme se pode observar a maioria dos casos de maior concorrência (acima de 15 candidatos por vaga) associa-se aos casos de menor evasão (abaixo de 12,50%) enquanto a maioria dos casos de menor concorrência (abaixo de 15 candidatos por vaga) associa-se aos casos de maior evasão (acima de 12,50%).

Uma outra evidência de que a concorrência e a evasão se associam inversamente pode ser verificada nas Tabelas 4.1 e 4.2, que mostram o teste de correlação não paramétrico Rô de Spearman obtido a partir do SPSS Statistics, aplicado ao caso das instituições privadas e públicas:

Tabela 4.1: Teste de correlação - evasão vs concorrência (instituições privadas)

			CONCORRÊNCIA	EVASÃO
Rô de Spearman	CONCORRÊNCIA	Coefficiente de Correlação	1,000	-,485**
		Sig. (2 extremidades)		,002
		N	40	40
	EVASÃO	Coefficiente de Correlação	-,485**	1,000
		Sig. (2 extremidades)	,002	
		N	40	40
**. A correlação é significativa no nível 0,01 (2 extremidades).				

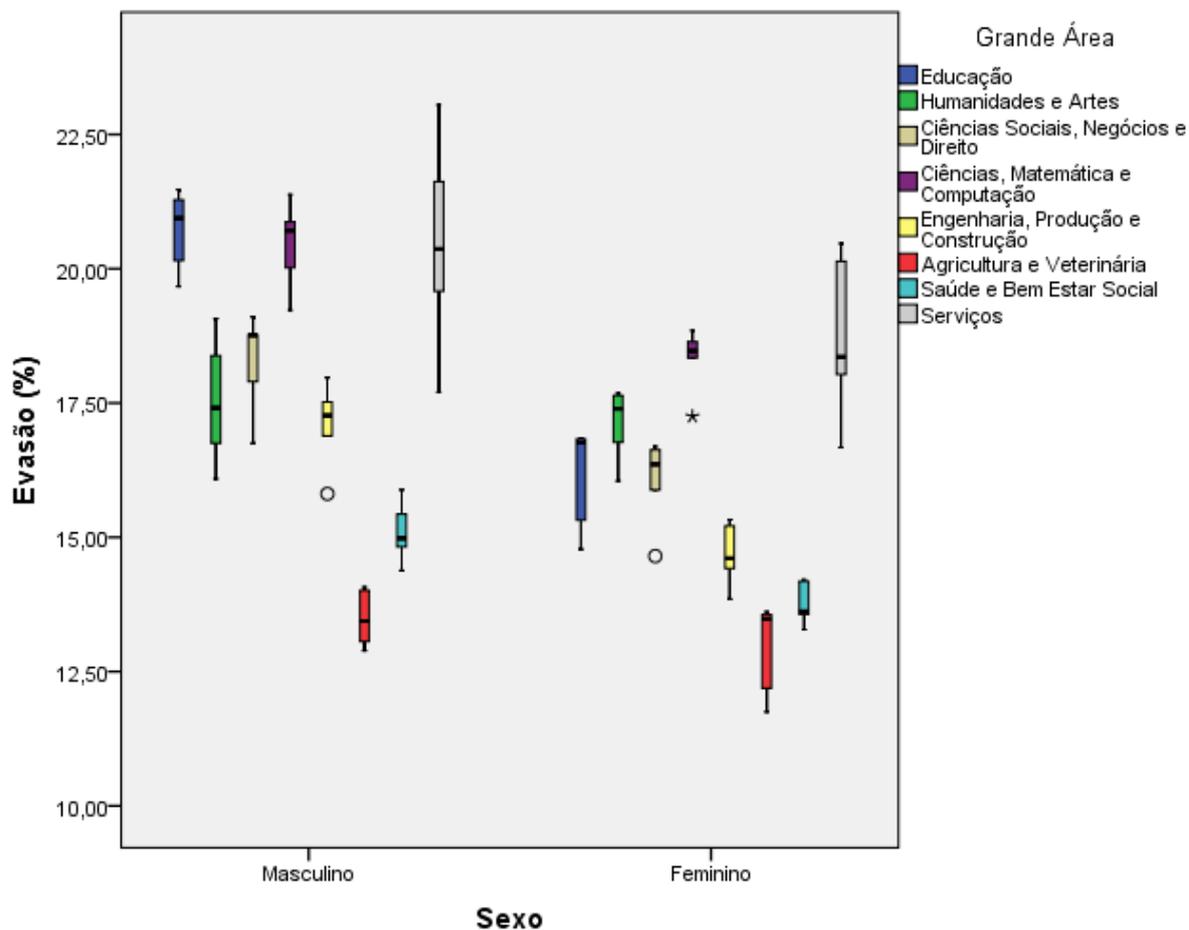
Tabela 4.2: Teste de correlação - evasão vs concorrência (instituições públicas)

			CONCORRÊNCIA	EVASÃO
Rô de Spearman	CONCORRÊNCIA	Coefficiente de Correlação	1,000	-,487**
		Sig. (2 extremidades)		,001
		N	40	40
	EVASÃO	Coefficiente de Correlação	-,487**	1,000
		Sig. (2 extremidades)	,001	
		N	40	40
**. A correlação é significativa no nível 0,01 (2 extremidades).				

As Tabelas 4.1 e 4.2 apresentam o Coeficiente de Correlação de Spearman usado para verificar se há correlação estatística entre concorrência e evasão, sendo -0,485 para as instituições privadas e de -0,487 para as instituições públicas. Foram testados 40 valores ($N = 40$) para cada grupo. A classificação do coeficiente de correlação pode ser interpretado assim: 0,10 até 0,30 (fraco); 0,40 até 0,60 (moderado); 0,70 até 1 (forte) [28]. Um valor negativo para o coeficiente indica que há uma correlação inversa, sendo que os valores do coeficiente variam de -1 a 1. Temos uma evidência então de que há uma correlação moderada inversa entre concorrência e evasão, tanto nas instituições privadas quanto nas públicas. O “N” apresentado nas tabelas corresponde ao número de valores testados: ele corresponde a 40 pois foram testados 5 valores de evasão e 5 valores de concorrência (correspondentes aos anos de 2010 a 2014) em 8 grandes áreas ($8 \times 5 = 40$). Os valores testados encontram-se na Tabela A.1. O teste de significância estatística (p-valor) retornou um valor inferior a 0,05, indicando que a correlação é significativa. No Capítulo 1 foi levantada a hipótese de que a relação candidatos/vagas é inversamente proporcional à evasão, ou seja, há uma tendência de que cursos superiores mais concorridos tenham taxas de desistência inferiores aos cursos superiores com menor concorrência. Conforme visto até aqui, há evidências de que a evasão seja moderadamente inversa em relação à concorrência.

A Figura 4.6 foi construída a partir dos valores da Tabela A.3 e compara a evasão entre os alunos do sexo masculino e feminino nas instituições privadas no período 2010-2014, nas grandes áreas do conhecimento:

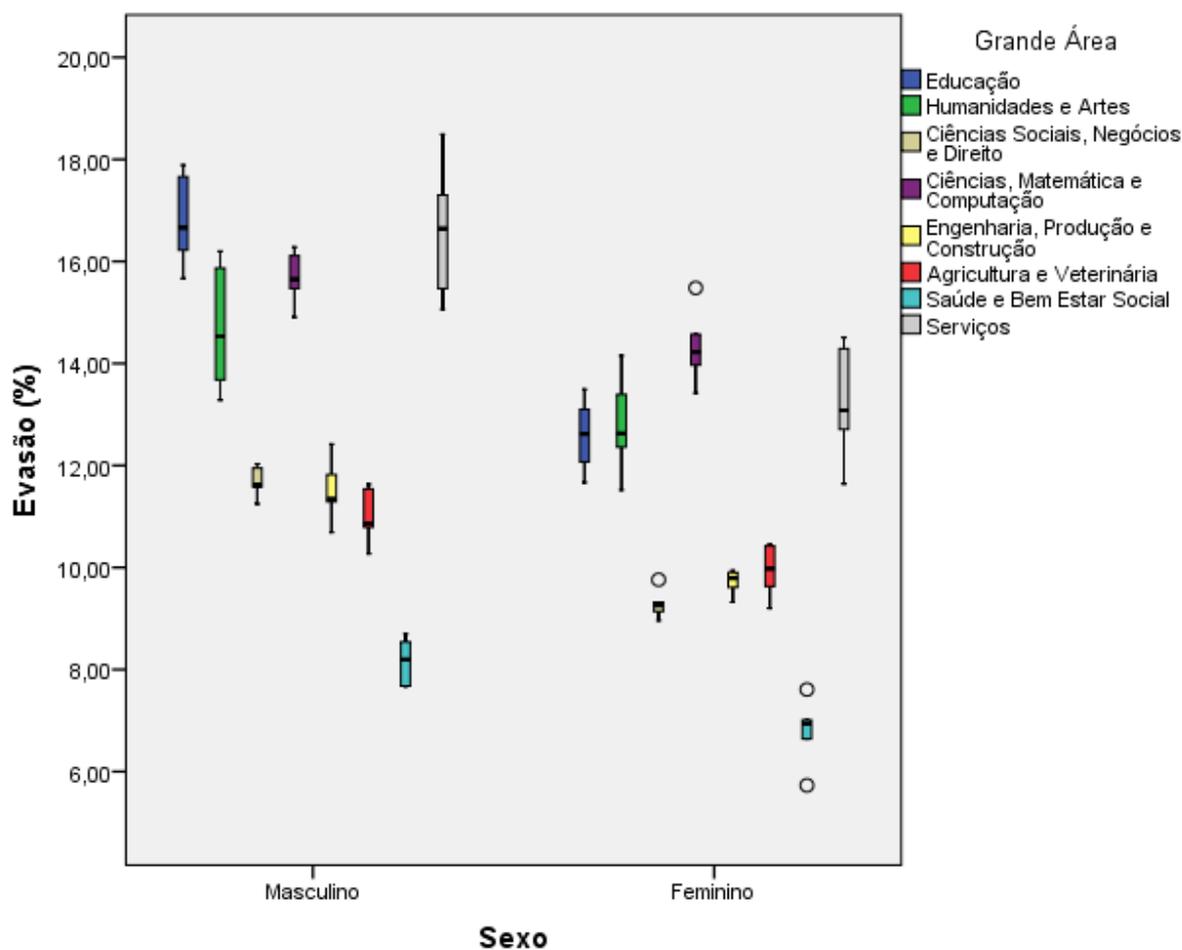
Figura 4.6: Evasão por sexo nas grandes áreas do conhecimento (instituições privadas)



Na Figura 4.6 cada grande área é representada por uma caixa de cor diferente. O eixo X é separado entre a evasão no sexo masculino e a evasão no sexo feminino. É possível verificar que as caixas correspondentes ao sexo masculino encontram-se posicionadas acima das caixas de mesma cor no sexo feminino. Isso indica que a evasão nas grandes áreas, no sexo masculino, foi maior que a evasão no sexo feminino no período 2010-2014. Temos então uma evidência de que a evasão no sexo masculino, de maneira geral, é maior que a evasão no sexo feminino, nos cursos superiores.

A Figura 4.7 foi construída a partir dos valores da Tabela A.4 e compara a evasão entre os alunos do sexo masculino e feminino nas instituições públicas no período 2010-2014, nas grandes áreas do conhecimento:

Figura 4.7: Evasão por sexo nas grandes áreas do conhecimento (instituições públicas)



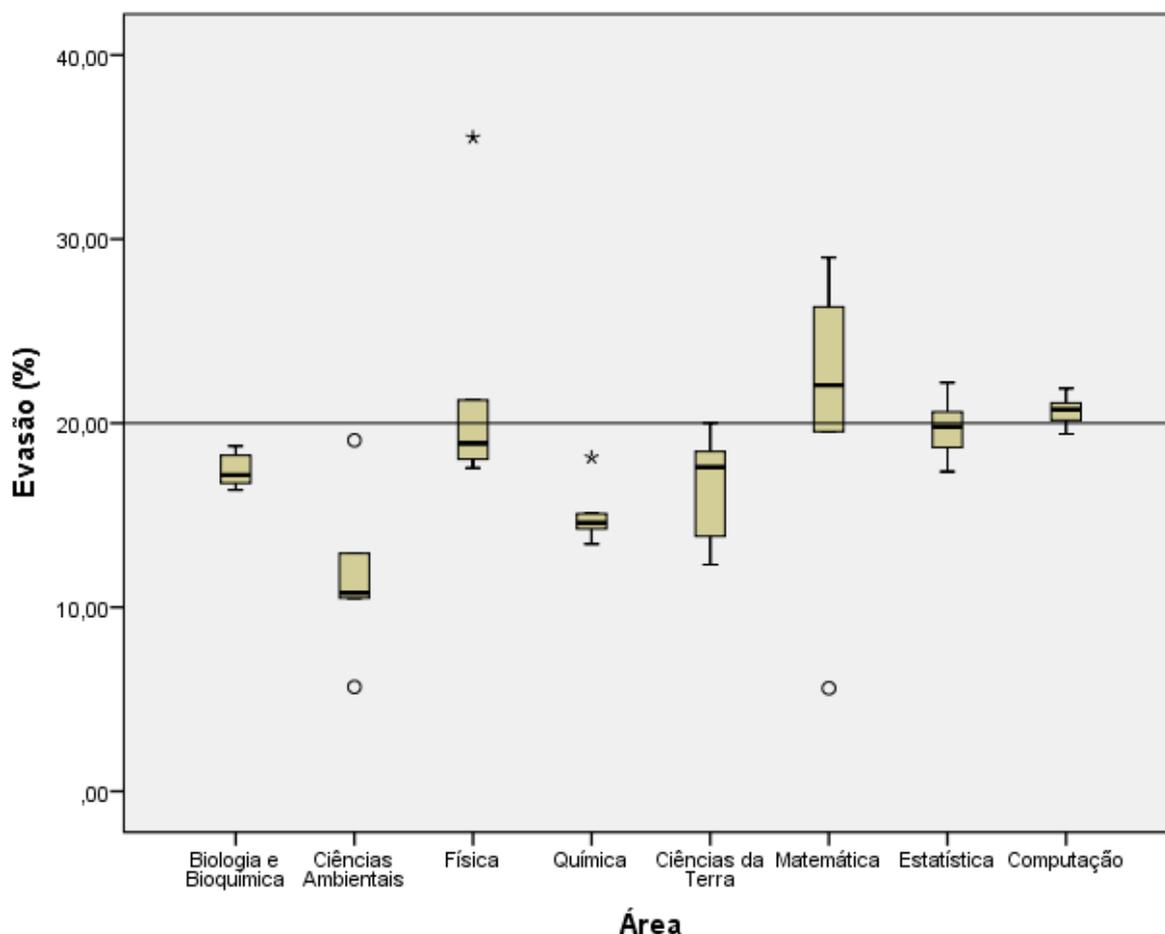
Tal como foi verificado nas instituições privadas, nas instituições públicas a evasão se mostra superior no público masculino em cada uma das grandes áreas do conhecimento, visto que, a partir da observação do *Box-plot* percebe-se que as caixas do sexo masculino encontram-se posicionadas acima das caixas correspondentes de mesma cor do sexo feminino. A Figura 4.7 mostra uma evidência de que a evasão nas áreas gerais do conhecimento é maior no sexo masculino em relação ao sexo feminino.

A seguir são apresentados os *Box-plots* detalhando-se a grande área de Ciências, Matemática e Computação para comparar a evasão nos cursos de Computação com os demais cursos desta grande área. A Figura 4.7 mostra a evasão nos cursos classificados pelo INEP como sendo de Computação (envolvendo as áreas detalhadas de Ciência da Computação, Uso do Computador e Processamento da Informação) em comparação com os cursos de outras áreas que compõem a grande área de Ciências, Matemática e Computação.

Na Figura 4.8 foi omitida a área detalhada de Ciências Físicas (Cursos Gerais) pois essa área só apresentou matrículas no ano de 2010 para as instituições privadas. Foi

omitida também a área detalhada de Ciências da vida (Curso gerais) pois não se verificou matrículas para essa área em nenhum dos anos avaliados. O *Box-plot* foi gerado a partir dos valores da Tabela A.5:

Figura 4.8: Evasão na grande área de Ciências, Matemática e Computação (instituições privadas)

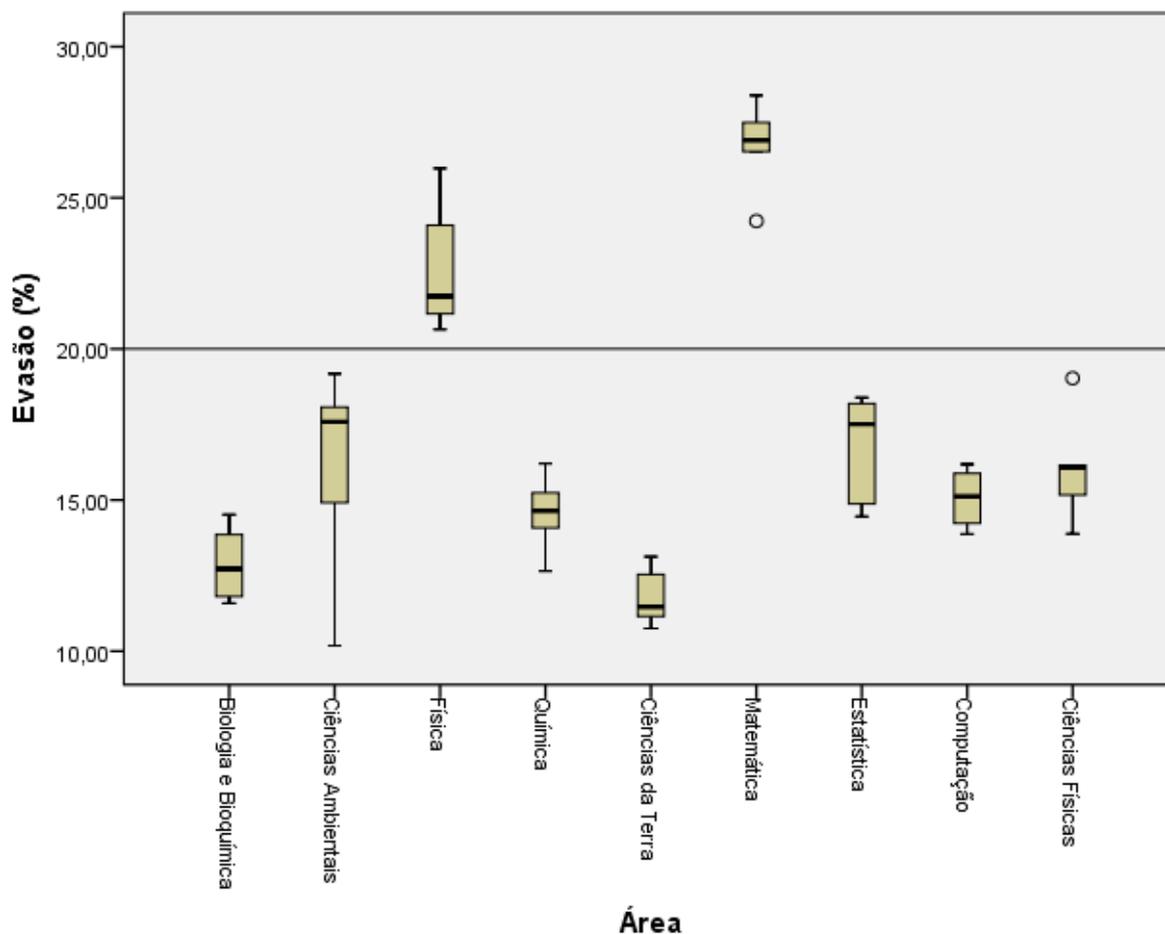


No Capítulo 1 foi levantada a hipótese de que os cursos da grande área de Ciências, Matemática e Computação que requerem maior uso de conhecimentos matemáticos e de abstração algorítmica possuem maiores taxas de evasão. As caixas posicionadas mais acima no *Box-plot* da Figura 4.8 são as de Matemática, Estatística, Computação e Física. Se assumirmos que essas áreas requerem maiores conhecimentos matemáticos e abstração algorítmica em relação às áreas de Biologia e Bioquímica, Ciências Ambientais, Química e Ciências da Terra, teremos uma evidência de confirmação dessa hipótese.

A Figura 4.9 ilustra a evasão nos cursos de Computação (segundo o INEP) nas instituições públicas em comparação com outras áreas da grande área de Ciências, Matemática e Computação. No *Box-plot* da Figura 4.9 foi omitida a área de Ciências da vida (Curso

gerais) pois não se verificou matrículas para essa área nos anos de 2010 e 2014 e o número de matrículas nos anos de 2011, 2012 e 2013 foi bem pequeno (inferior a 200 matrículas), conforme pode ser verificado na Tabela A.5.

Figura 4.9: Evasão na grande área de Ciências, Matemática e Computação (instituições públicas)



Pela análise da Figura 4.9 percebe-se que os cursos que se destacam quanto à evasão são os de Física e de Matemática. Os cursos com menores taxas de evasão são os da área de Biologia e Bioquímica e de Ciências Ambientais. Aqui podemos ver que os cursos com maiores taxas de evasão são também cursos com grande exigência matemática. A evasão nos cursos de Computação mostrou-se aproximadamente no mesmo patamar em relação às áreas de Ciências Físicas, Estatística, Química e Ciências Ambientais.

Nas Tabelas 4.3 e 4.4 temos a comparação entre a evasão nos cursos classificados pelo INEP como sendo de Computação com a média nacional considerando-se as 8 grandes áreas do conhecimento:

Tabela 4.3: Evasão na Computação em relação à média nacional (instituições privadas)

Ano	Média Nacional(%)	Computação (%)
2010	15,86	19,42
2011	17,46	20,74
2012	16,82	20,16
2013	16,81	21,10
2014	17,16	21,89

Tabela 4.4: Evasão na Computação em relação à média nacional (instituições públicas)

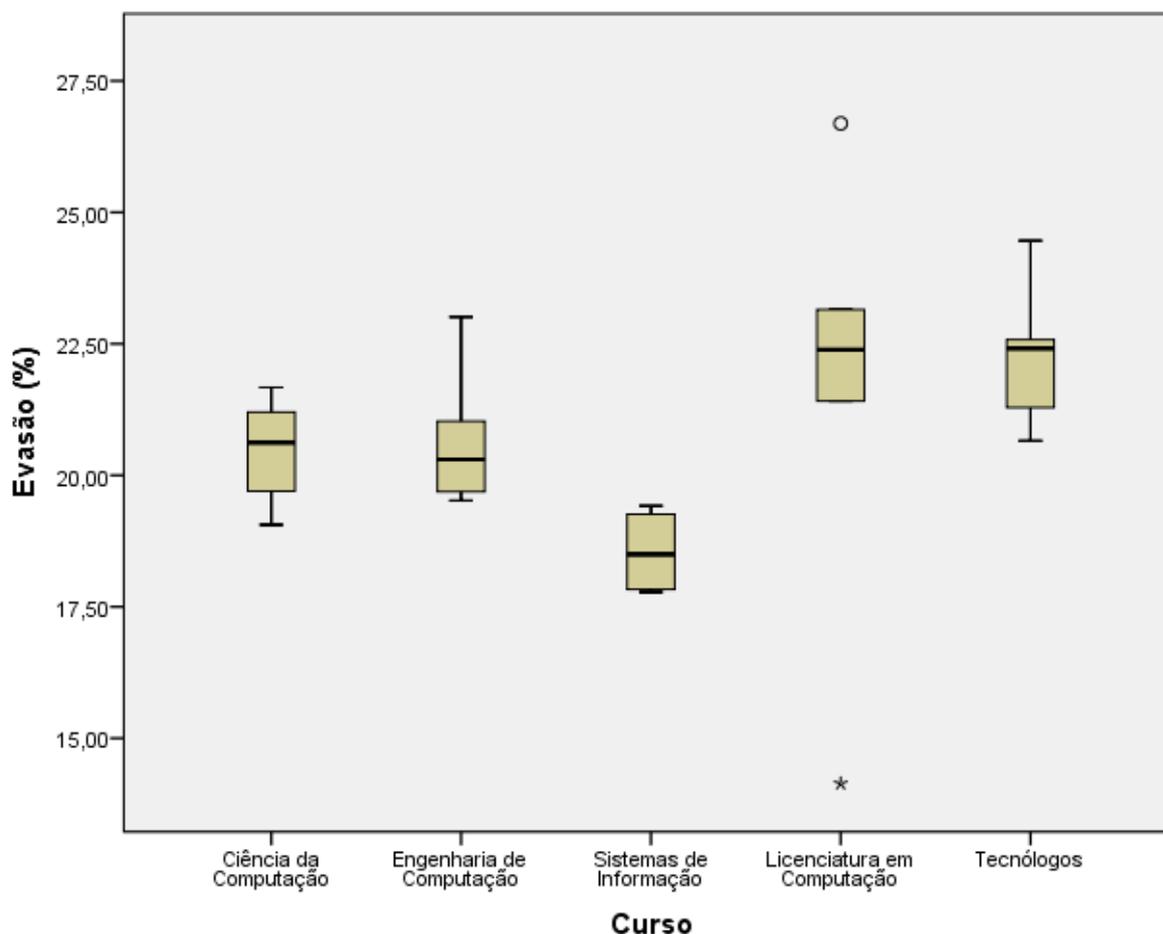
Ano	Média Nacional (%)	Computação (%)
2010	12,16	14,23
2011	11,46	13,87
2012	12,00	16,18
2013	12,45	15,12
2014	12,35	15,88

Nota-se que a evasão nos cursos de Computação, tanto nas instituições públicas quanto nas privadas, ficou acima da média nacional no período 2010-2014.

Os gráficos apresentados a seguir exploram a evasão em diferentes cursos de Computação. É apresentada a evasão para os cursos de Ciência da Computação, Engenharia de Computação, Engenharia de Software, Sistemas de Informação, Licenciatura em Computação e cursos de Tecnólogos. É importante observar que, a abordagem de escolha dos cursos de Computação usados nos gráficos é a mesma adotada pela Sociedade Brasileira de Computação (SBC) em relatórios estatísticos sobre a educação superior em Computação [72]. O curso de Engenharia de Computação não é classificado pelo INEP na área geral de Ciências, Matemática e Computação mas na área geral de Engenharia, Produção e Construção e os cursos de Licenciatura em Computação são classificados na área geral de Educação. Sendo assim, nem todos os cursos apresentados são classificados como cursos de Computação pelo INEP.

A Figura 4.10, que foi construída a partir dos valores da Tabela A.6, ilustra a evasão nos cursos de Computação nas instituições privadas no período 2010-2014. O curso de Engenharia de Software, que é um curso relativamente recente, foi omitido nesse *Box-plot*, visto que só foram registradas matrículas para esse curso nas instituições privadas nos anos de 2013 e 2014, sendo de 122 e 296 alunos respectivamente:

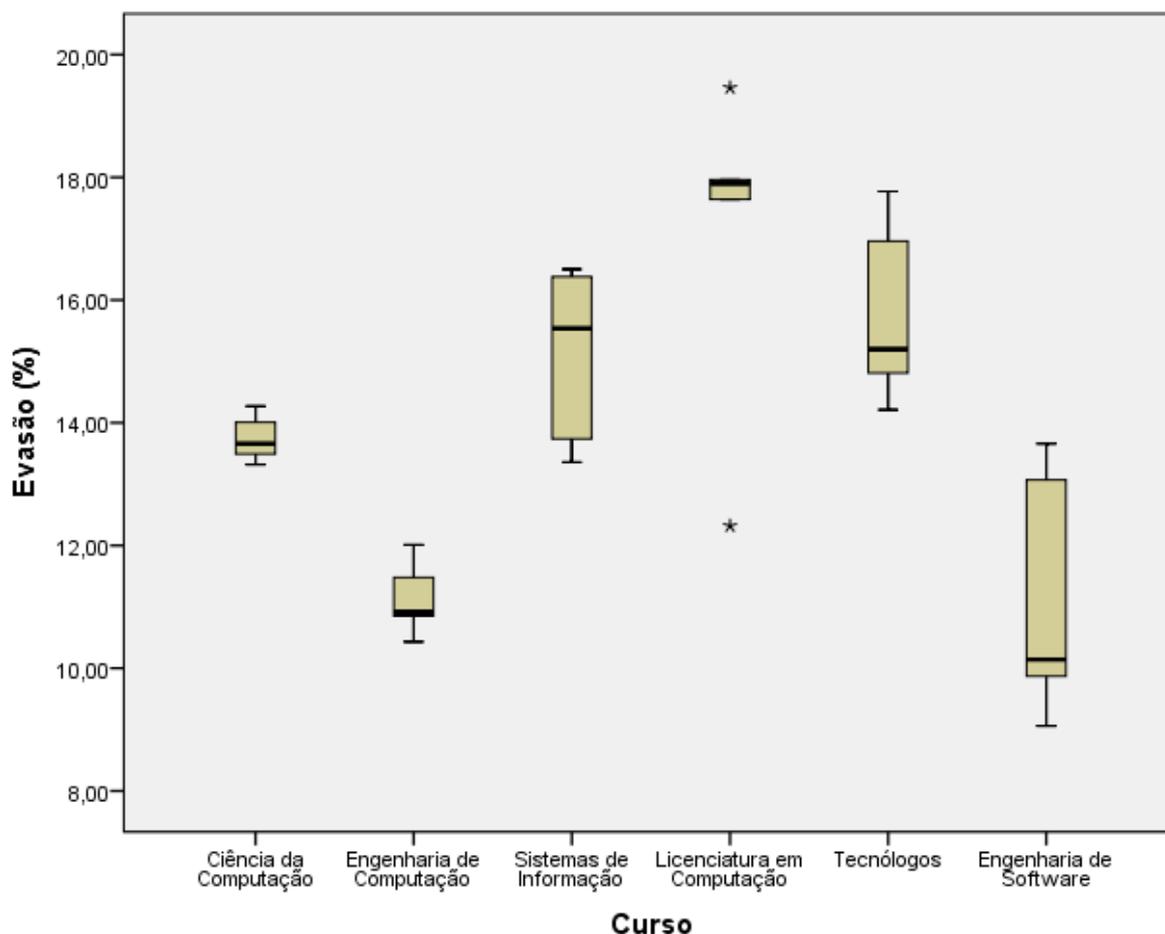
Figura 4.10: Evasão nos cursos de Computação (instituições privadas)



A análise da Figura 4.10 nos permite constatar que a evasão foi superior nos cursos de Licenciatura em Computação e nos cursos de Tecnólogos. O curso de Sistemas de Informação apresentou evasão menor e os cursos de Ciência da Computação e Engenharia de Computação evasão aproximadamente no mesmo patamar. O curso de Licenciatura em Computação apresentou um valor discrepante para a evasão no ano de 2012 de 14,14%, pequeno em relação aos demais valores e apresentou também um valor discrepante em 2014, de 26,69%, superior às demais taxas de evasão. Conforme mencionado na Subseção 2.2.2, cursos com melhores perspectivas remuneratórias tendem a apresentar menores taxas de evasão. A área profissional de professor de informática, oferece salários em média mais baixos que outras áreas. Contudo, essa relação entre a evasão na Licenciatura em Computação e a atratividade de mercado requer maiores análises e um maior aprofundamento.

A Figura 4.11, que foi construída a partir dos valores da Tabela A.6, ilustra a evasão nos cursos de Computação nas instituições públicas:

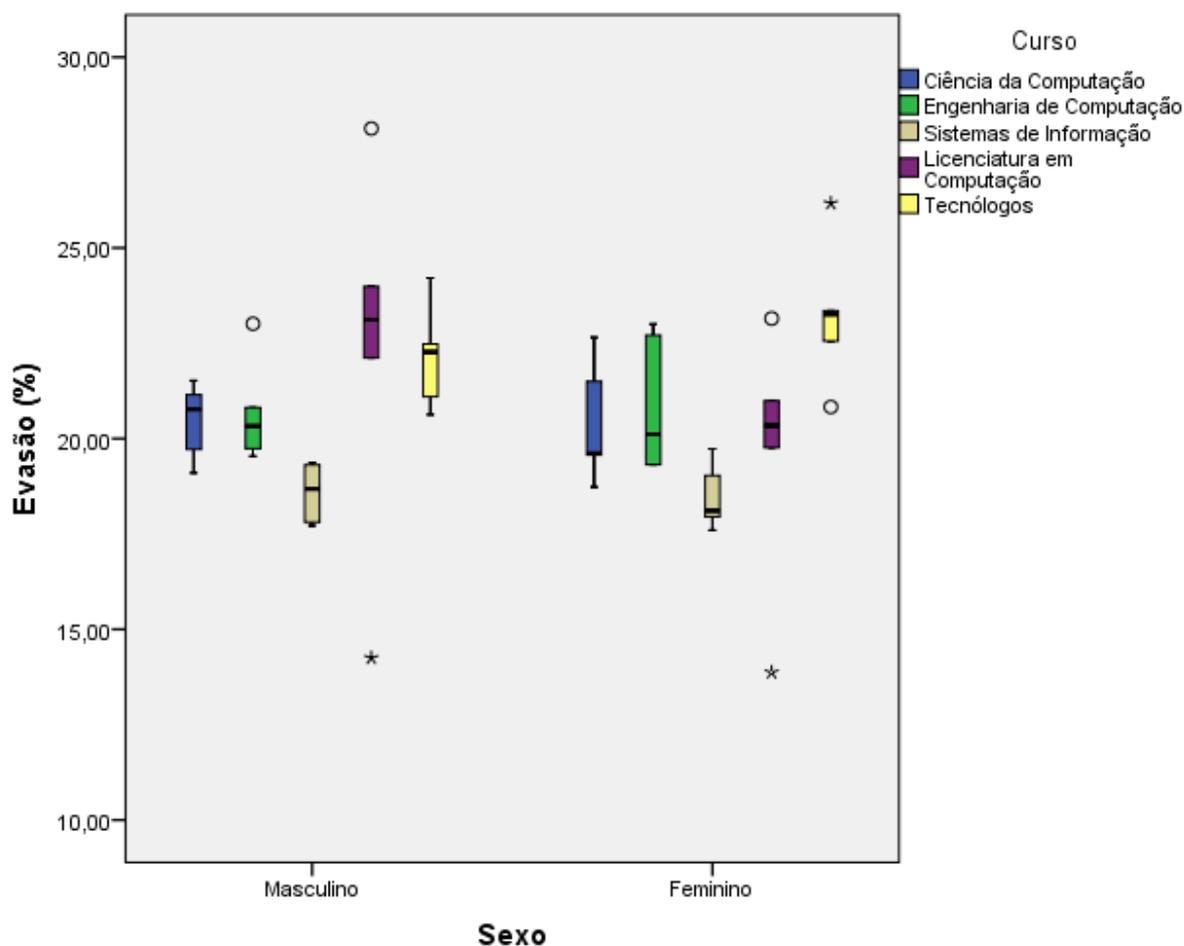
Figura 4.11: Evasão nos cursos de Computação (instituições públicas)



De maneira similar ao que foi verificado nas instituições privadas, nas instituições públicas os cursos com maior evasão foram os de tecnólogos e os de Licenciatura em Computação. O curso de Sistemas de Informação também apresentou uma evasão próxima dos cursos de tecnólogos. O curso de Licenciatura em Computação apresentou um valor discrepante para evasão, pequeno em relação aos demais conjuntos de dados, no ano de 2010 (12,32%) e apresentou um valor discrepante elevado para evasão no ano de 2013 (19,46%).

A Figura 4.12, que foi construída a partir dos dados da Tabela A.7, compara a evasão em relação ao sexo nos cursos de Computação nas instituições privadas:

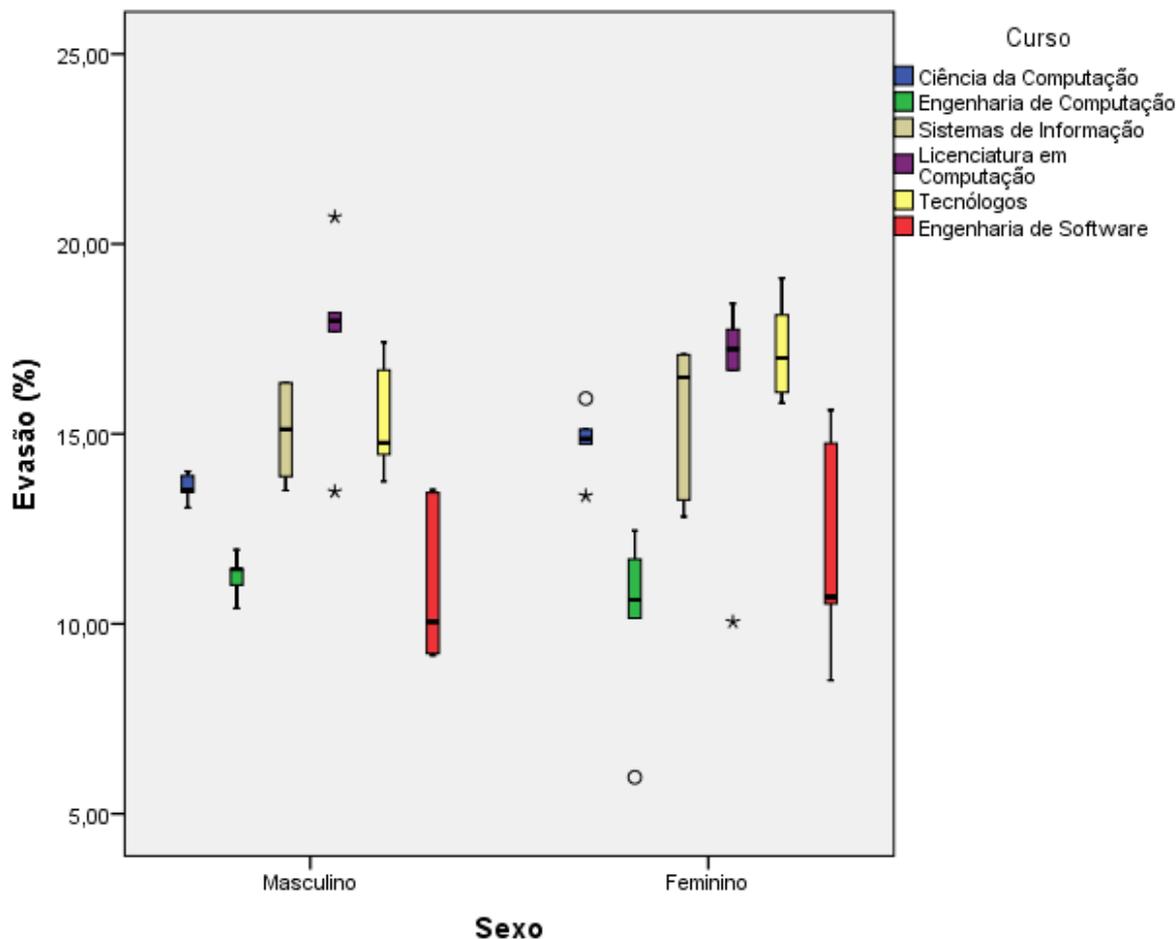
Figura 4.12: Evasão por sexo nos cursos de Computação (instituições privadas)



A partir da observação da Figura 4.12 é possível verificar que a evasão não difere muito entre o público masculino e o feminino, nas instituições privadas, nos cursos de Computação, uma vez que as caixas de cores iguais estão posicionadas aproximadamente na mesma altura, com algumas variações como no curso de Licenciatura em Computação, em que a caixa do sexo masculino está posicionada mais acima, indicando uma evasão superior.

A Figura 4.13, que foi construída a partir dos dados da Tabela A.8, compara a evasão em relação ao sexo nos cursos de Computação nas instituições públicas:

Figura 4.13: Evasão por sexo nos cursos de Computação (instituições públicas)



Tal como foi verificado para as instituições privadas, nas instituições públicas, a evasão nos cursos de Computação não diferiu muito entre os sexos masculino e feminino.

4.1.2 Análise de Sobrevivência

As curvas de sobrevivência foram traçadas para os alunos ingressantes em 2010 nos cursos de Ciência da Computação, Engenharia de Computação, Licenciatura em Computação, Engenharia de Software, Sistemas de Informação e para os cursos de Tecnólogos. Inicialmente, são apresentadas as curvas de sobrevivência usando o método de Kaplan-Meier. Em seguida, a análise multivariada usando regressão de Cox é exibida.

A Figura 4.14 exhibe as curvas de sobrevivência para as instituições privadas, levando-se em consideração o sexo do aluno:

Figura 4.14: Curvas de sobrevivência (sexo) - instituições privadas

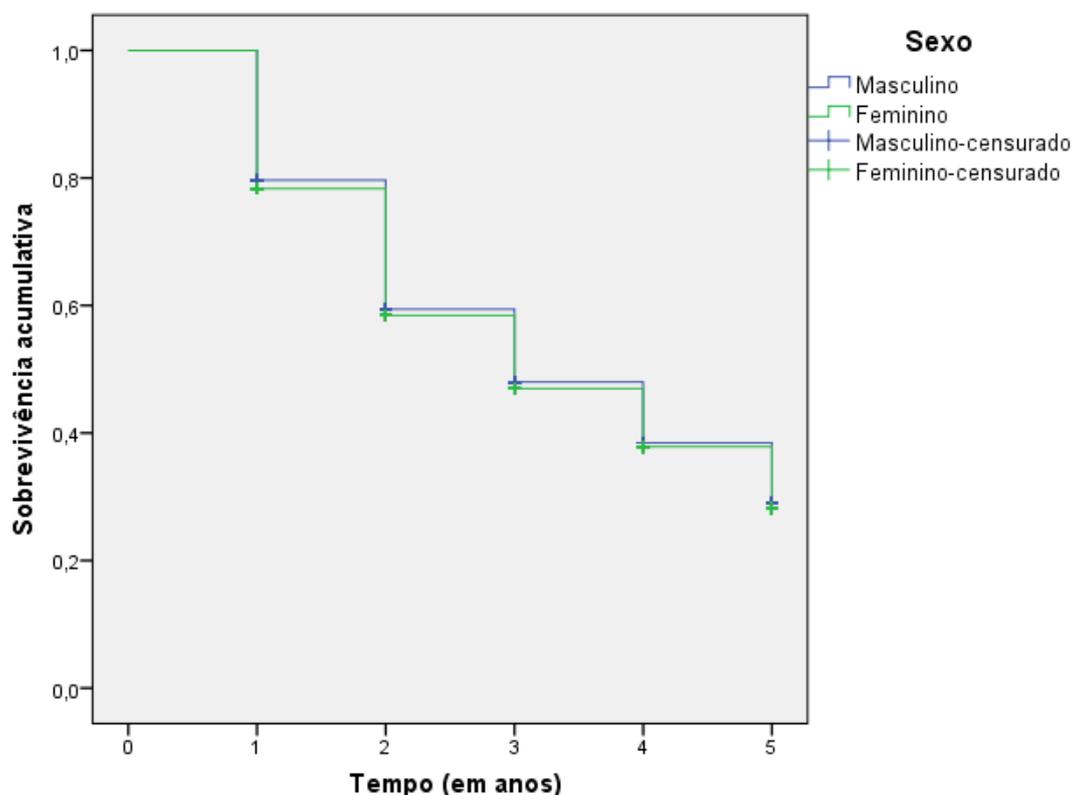


Tabela 4.5: Estatísticas - curvas de sobrevivência (sexo) - instituições privadas

	Qui-quadrado	Graus de liberdade	Sig. (p-valor)
Log Rank (Mantel-Cox)	4,786	1	0,029
Breslow (Generalized Wilcoxon)	6,970	1	0,008
Tarone-Ware	5,999	1	0,014

O gráfico de sobrevivência da Figura 4.14 foi construído a partir de 68404 registros do sexo masculino e 11623 registros do sexo feminino. A curva posicionada mais acima no gráfico indica uma maior sobrevivência em relação ao evento evasão. Nota-se, portanto, que o sexo masculino apresentou uma sobrevivência superior ao sexo feminino pois a sua curva se encontra um pouco acima do sexo feminino. Os 3 testes estatísticos empregados revelam que a sobrevivência do sexo masculino para os ingressos em 2010 é significativamente superior à sobrevivência do sexo feminino ($p\text{-valor} < 0,05$).

A Figura 4.15 exibe as curvas de sobrevivência para as instituições públicas, levando-se em consideração o sexo do aluno:

Figura 4.15: Curvas de sobrevivência (sexo) - instituições públicas

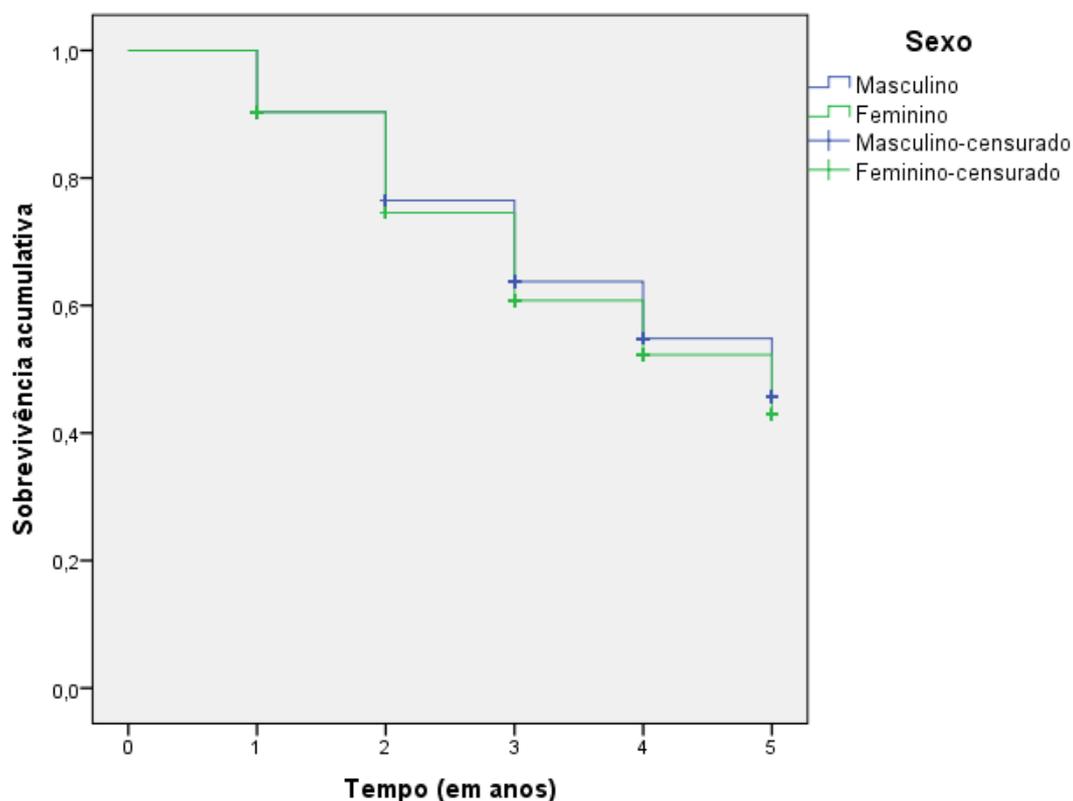


Tabela 4.6: Estatísticas - curvas de sobrevivência (sexo) - instituições públicas

	Qui-quadrado	Graus de liberdade	Sig. (p-valor)
Log Rank (Mantel-Cox)	12,962	1	0,000
Breslow (Generalized Wilcoxon)	12,653	1	0,000
Tarone-Ware	13,015	1	0,000

O gráfico de sobrevivência da Figura 4.15 foi construído a partir de 18648 registros do sexo masculino e 5048 registros do sexo feminino. Nota-se que o sexo masculino apresentou uma sobrevivência superior ao sexo feminino pois a sua curva se encontra um pouco acima da curva do sexo feminino, tal como ocorreu nas instituições privadas. Os 3 testes estatísticos empregados revelam diferenças estatisticamente significativas entre as curvas. Cabe observar que no primeiro ano a sobrevivência apresentada pelos diferentes sexos foi praticamente a mesma, visto que as curvas estão sobrepostas no ano 1. A diferença mais perceptível entre elas inicia-se a partir do ano 2.

A Figura 4.16 exhibe as curvas de sobrevivência para as instituições privadas, levando-se em consideração se o aluno é cotista ou não:

Figura 4.16: Curvas de sobrevivência (cotista) - instituições privadas

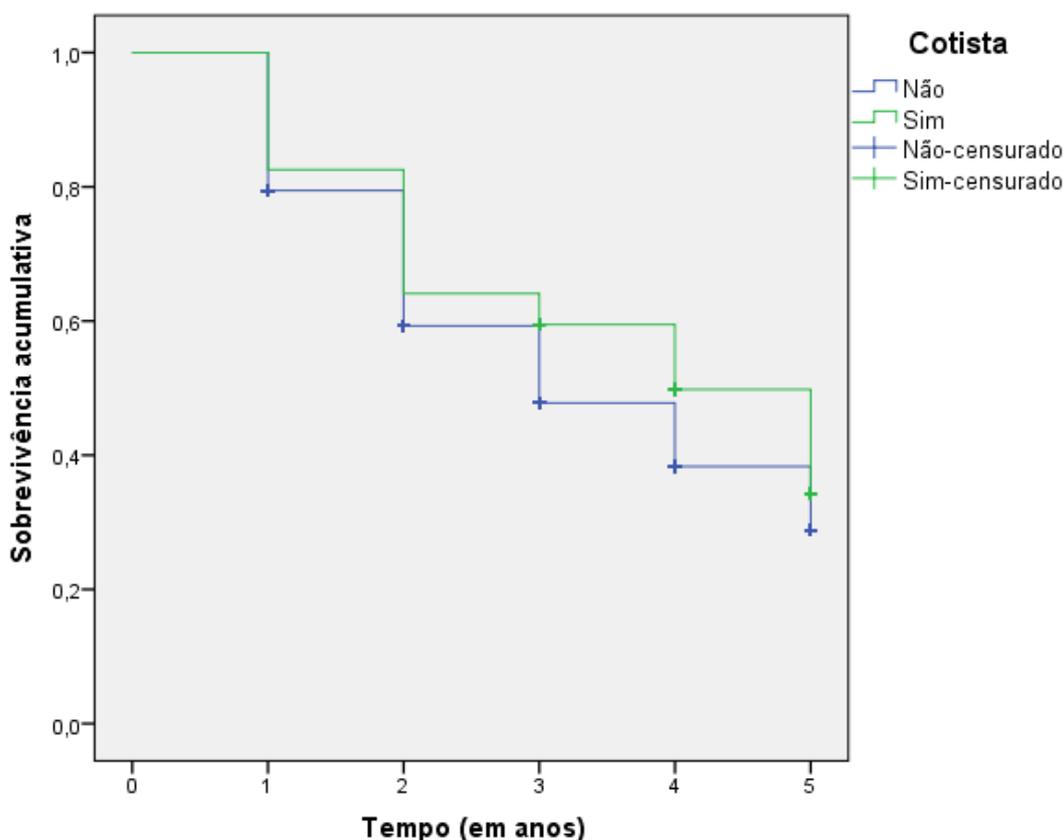


Tabela 4.7: Estatísticas - curvas de sobrevivência (cotista) - instituições privadas

	Qui-quadrado	Graus de liberdade	Sig. (p-valor)
Log Rank (Mantel-Cox)	6,014	1	0,014
Breslow (Generalized Wilcoxon)	5,950	1	0,015
Tarone-Ware	6,439	1	0,011

O gráfico de sobrevivência da Figura 4.16 foi construído a partir de 79832 registros de não cotistas e 195 registros de cotistas. Em se tratando de instituições privadas é natural que o número de cotistas seja pequeno. Nota-se que os cotistas apresentaram sobrevivência superior aos não cotistas. Os 3 testes estatísticos empregados revelam diferenças estatisticamente significativas entre as curvas.

A Figura 4.17 exhibe as curvas de sobrevivência para as instituições públicas, levando-se em consideração se o aluno é cotista ou não:

Figura 4.17: Curvas de sobrevivência (cotista) – instituições públicas

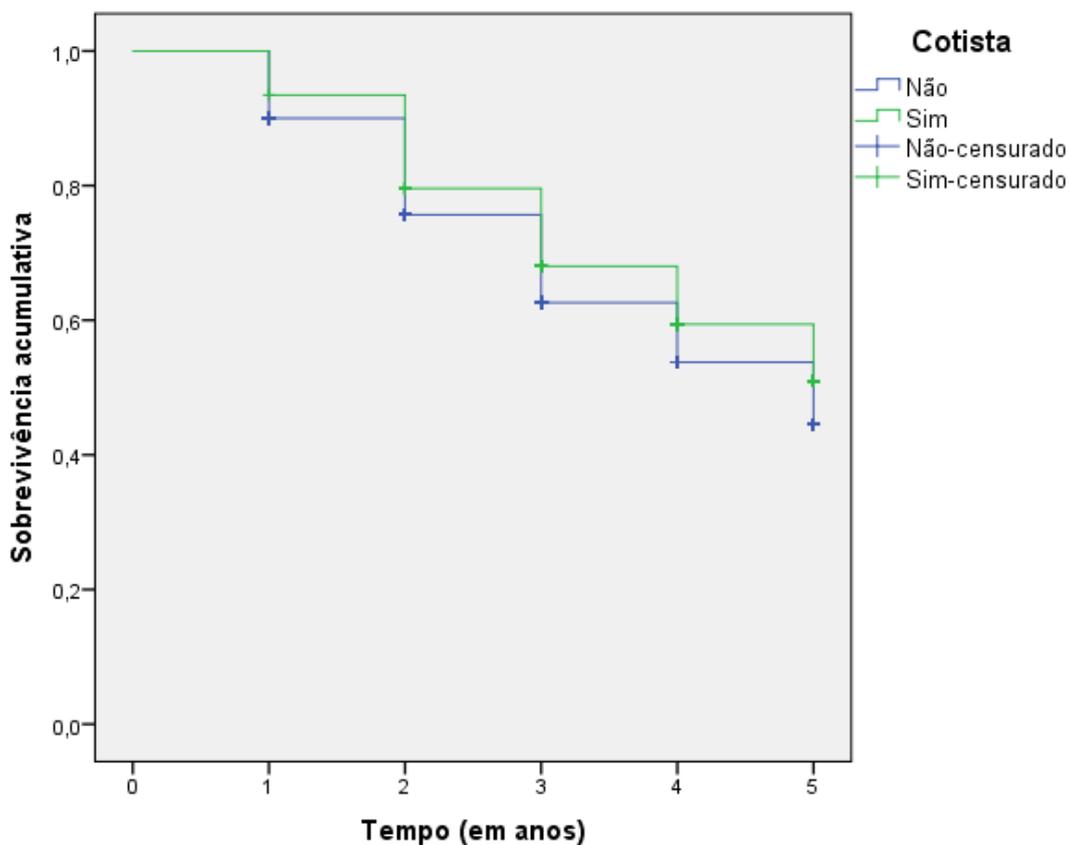


Tabela 4.8: Estatísticas - curvas de sobrevivência (cotista) - instituições públicas

	Qui-quadrado	Graus de liberdade	Sig. (p-valor)
Log Rank (Mantel-Cox)	34,747	1	0,000
Breslow (Generalized Wilcoxon)	35,241	1	0,000
Tarone-Ware	35,332	1	0,000

O gráfico de sobrevivência da Figura 4.17 foi construído a partir de 21523 registros de não cotistas e 2173 registros de cotistas. Tal como verificado nas instituições privadas nota-se que os cotistas apresentaram sobrevivência superior aos não cotistas. Os 3 testes estatísticos empregados revelam diferenças estatisticamente significativas entre as curvas.

A Figura 4.18 exibe as curvas de sobrevivência para as instituições privadas, levando-se em consideração a forma de ingresso do aluno:

Figura 4.18: Curvas de sobrevivência (ingresso) - instituições privadas

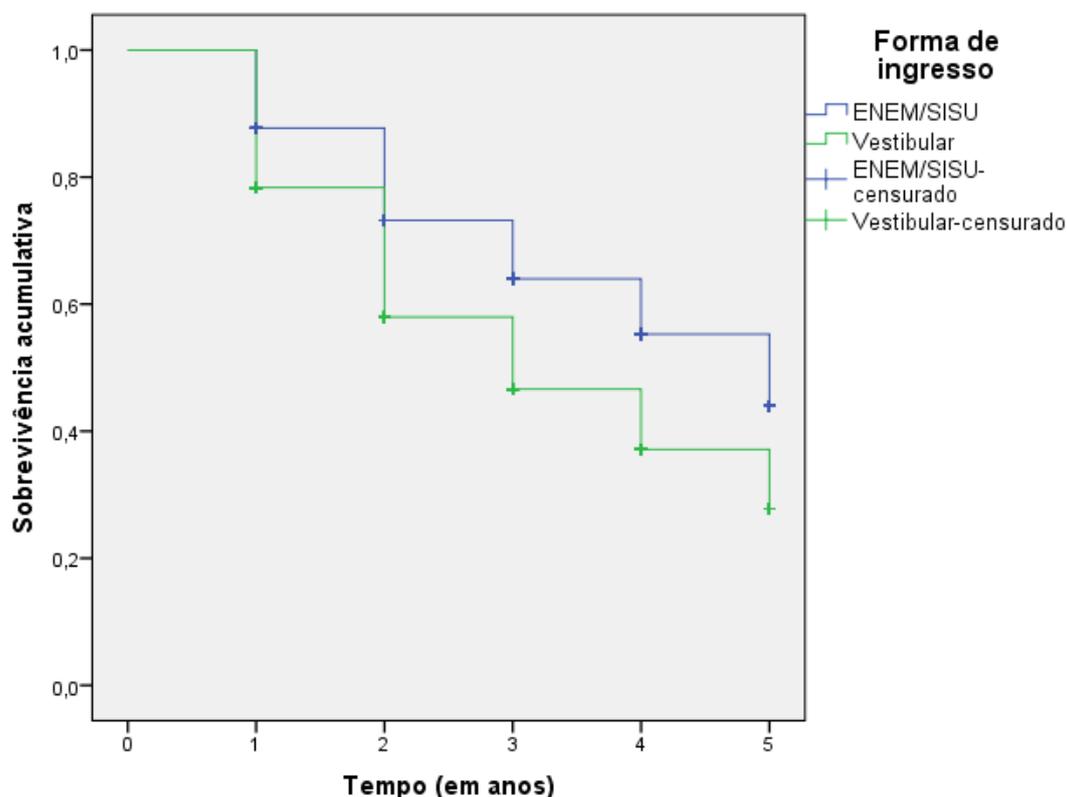


Tabela 4.9: Estatísticas - curvas de sobrevivência (ingresso) - instituições privadas

	Qui-quadrado	Graus de liberdade	Sig. (p-valor)
Log Rank (Mantel-Cox)	728,803	1	0,000
Breslow (Generalized Wilcoxon)	725,816	1	0,000
Tarone-Ware	750,920	1	0,000

O gráfico de sobrevivência da Figura 4.18 foi construído a partir de 6228 registros de ingressantes via ENEM/SISU e 58650 registros de ingressantes via Vestibular. A curva de sobrevivência para outras formas de ingresso foi omitida desse gráfico visto que nos microdados do censo da educação superior de 2010 não é descrito qual é o outro tipo de forma de seleção além de ENEM e além do Vestibular. Os ingressos via SISU mostraram sobrevivência superior nas instituições privadas em relação aos ingressos via Vestibular. Conforme mencionado na Subseção 2.2.2, O ENEM é usado pelo PROUNI para seleção de candidatos bolsistas nas universidades privadas. É natural que um aluno bolsista

ofereça maior resistência à evasão. Os 3 testes estatísticos empregados revelam diferenças estatisticamente significativas entre as curvas.

A Figura 4.19 exibe as curvas de sobrevivência para as instituições públicas, levando-se em consideração as formas de ingresso por Vestibular e via ENEM/SISU:

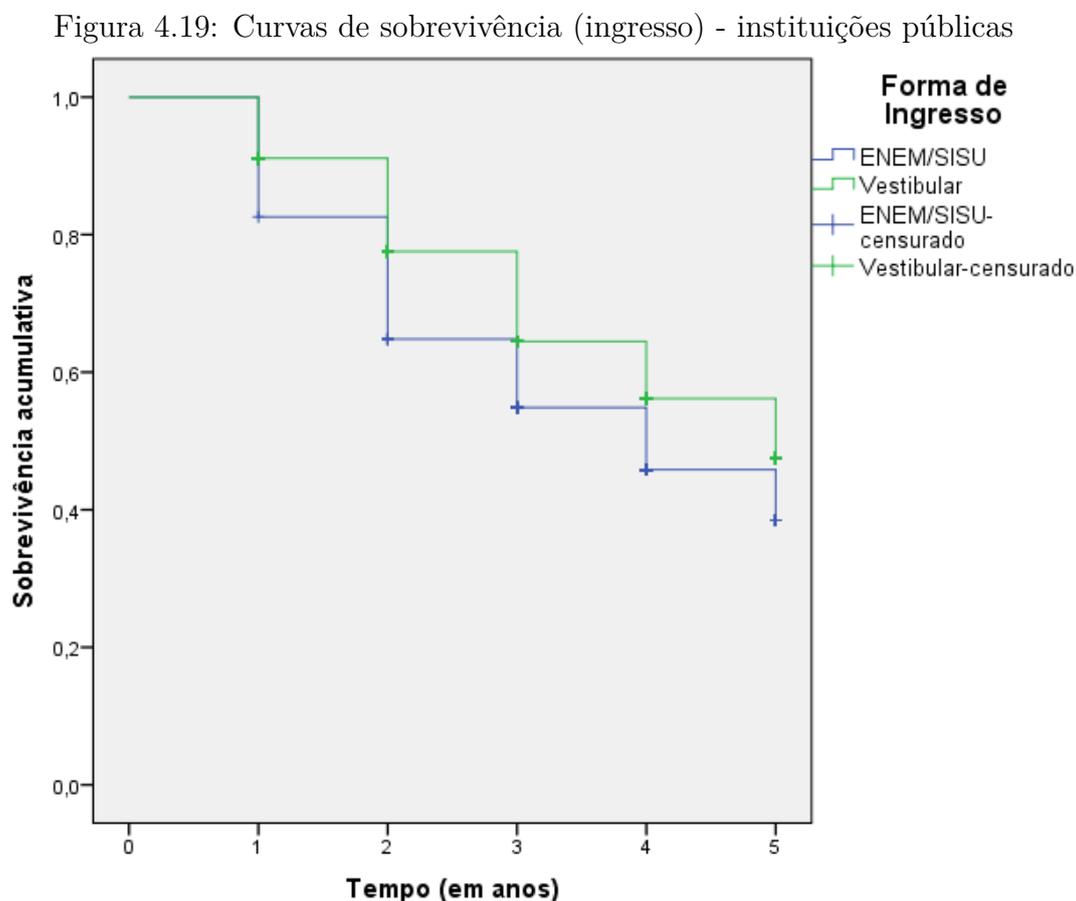


Tabela 4.10: Estatísticas - curvas de sobrevivência (ingresso) - instituições públicas

	Qui-quadrado	Graus de liberdade	Sig. (p-valor)
Log Rank (Mantel-Cox)	108,744	1	0,000
Breslow (Generalized Wilcoxon)	141,453	1	0,000
Tarone-Ware	126,519	1	0,000

Diferente do que foi verificado para as instituições privadas, nas instituições públicas os alunos que ingressaram via Vestibular apresentaram sobrevivência superior aos ingressos via ENEM/SISU. O gráfico de sobrevivência da Figura 4.19 foi construído a partir de 2129 registros de ingressantes via ENEM/SISU e 16368 registros de ingressantes via Vestibular.

Isso também foi elencado na Subseção 2.2.2, em que se relacionou o ingresso via SISU a uma mobilidade maior dos alunos.

A Figura 4.20 exibe as curvas de sobrevivência para as instituições privadas comparando aqueles alunos que estudaram em instituições ofertantes de pós-graduação na área de Computação, com aqueles alunos que estudaram em instituições não ofertantes de pós-graduação em Computação:

Figura 4.20: Curvas de sobrevivência (pós-graduação em Computação) – instituições privadas

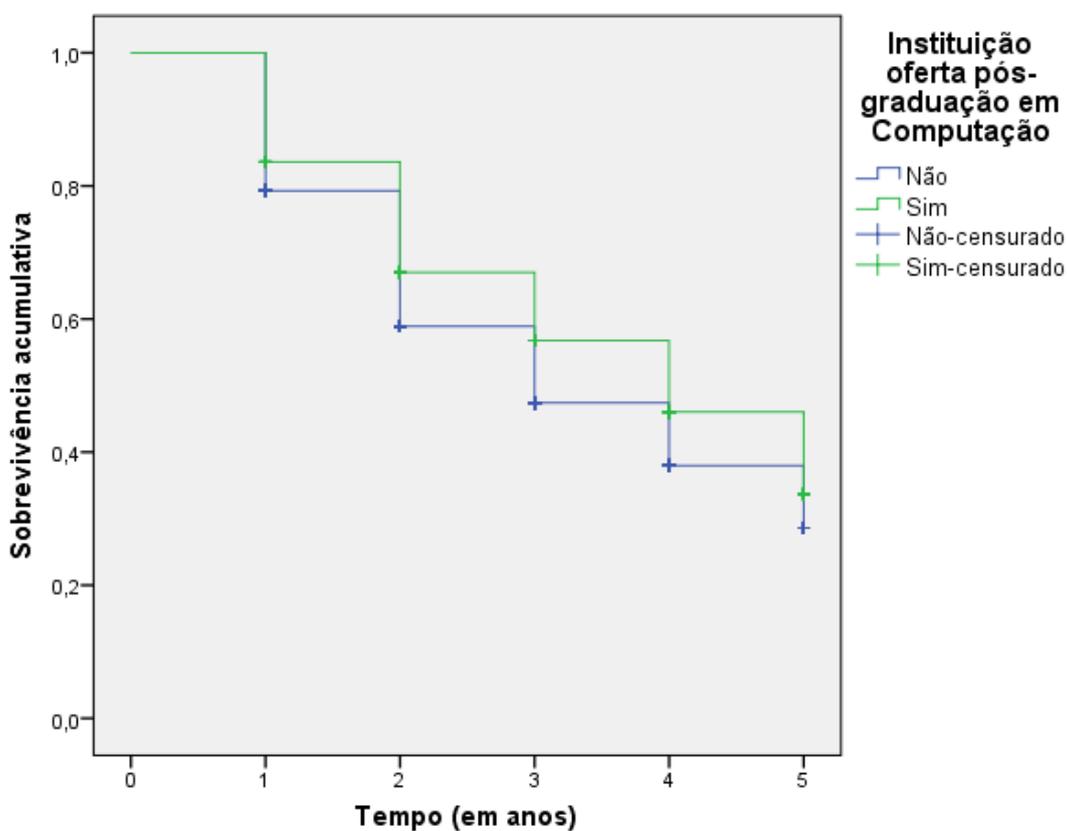


Tabela 4.11: Estatísticas - curvas de sobrevivência (pós-graduação em Computação) - instituições privadas

	Qui-quadrado	Graus de liberdade	Sig. (p-valor)
Log Rank (Mantel-Cox)	83,501	1	0,000
Breslow (Generalized Wilcoxon)	107,453	1	0,000
Tarone-Ware	101,899	1	0,000

A Figura 4.20 mostra que aqueles alunos que fizeram seus cursos superiores de Computação em instituições ofertantes de pós-graduação em Computação, apresentaram sobrevivência superior em relação aos alunos que estudaram em instituições que não oferecem essa verticalização do ensino. As curvas de sobrevivência foram traçadas a partir de 3656 registros de alunos que estudaram em instituições ofertantes de pós-graduação em Computação e de 76371 alunos que não estudaram em instituições ofertantes de pós-graduação em Computação.

A Figura 4.21 exibe as curvas de sobrevivência para as instituições públicas comparando aqueles alunos que estudaram em instituições ofertantes de pós-graduação na área de Computação, com aqueles alunos que estudaram em instituições não ofertantes de pós-graduação em Computação:

Figura 4.21: Curvas de sobrevivência (pós-graduação em Computação) – instituições públicas

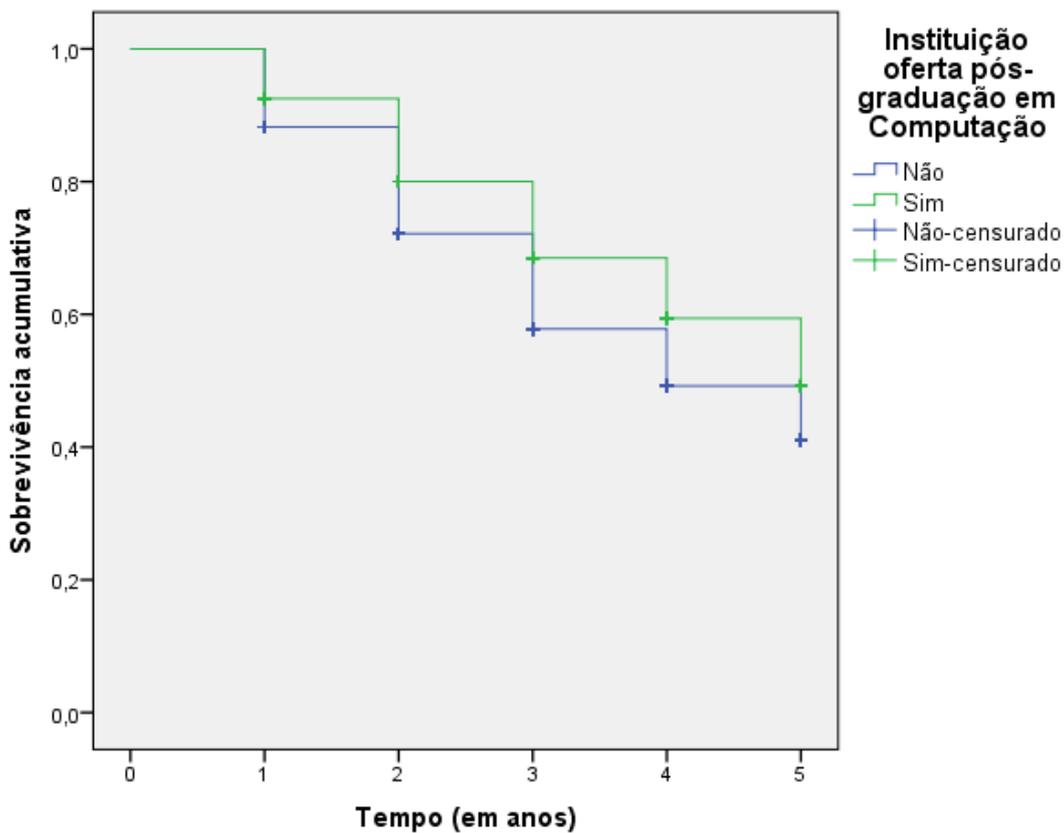


Tabela 4.12: Estatísticas - curvas de sobrevivência (pós-graduação em Computação) - instituições públicas

	Qui-quadrado	Graus de liberdade	Sig. (p-valor)
Log Rank (Mantel-Cox)	235,994	1	0,000
Breslow (Generalized Wilcoxon)	284,898	1	0,000
Tarone-Ware	265,402	1	0,000

O resultado obtido foi o mesmo em relação às instituições privadas: alunos estudantes de cursos superiores de Computação em instituições ofertantes de pós-graduação em Computação apresentaram sobrevivência superior aos que estudaram em instituições que não ofertam essa verticalização do ensino. Foram usados registros de 11711 alunos que estudaram em instituições ofertantes de pós-graduação em Computação e 11985 registros de alunos que estudaram em instituições não ofertantes de pós-graduação em Computação.

A análise de regressão de Cox foi usada para avaliar as variáveis significativas, de acordo com o método de Kaplan-Meier, determinando o risco de cada uma em relação à sobrevivência. Algumas considerações para a interpretação dos resultados das Tabelas 4.13 e 4.14: A coluna B traz os coeficientes de regressão. Valores positivos nesta coluna indicam que a variável correspondente está contribuindo como previsor positivo para a evasão, ou seja, valores positivos de B indicam que para aquela variável observa-se evasão maior. $\text{Exp}(B)$ é a razão de chances, ou seja, a razão de risco para uma cada variável. Também temos nas tabelas o erro padrão de B (SE), o seu valor de significância de teste de Wald, os graus de liberdade (df), e o valor de significância do coeficiente (p-valor).

A Tabela 4.13 mostra os resultados da análise multivariada feita a partir da regressão de Cox, considerando-se as instituições privadas:

Tabela 4.13: Análise multivariada da sobrevivência (instituições privadas)

Variáveis	B	SE	Wald	df	Exp(B)	Intervalo de confiança de 95%	p-valor
ingresso_vestibular (SIM)	0,130	0,012	114,925	1	1,139	[1,112; 1,167]	0,000
ingresso_enem (SIM)	-0,143	0,015	86,384	1	0,866	[0,841; 0,893]	0,000
cotista (SIM)	-0,061	0,095	0,414	1	0,941	[0,782; 1,133]	0,520
sexo_aluno (feminino)	0,030	0,013	5,575	1	1,030	[1,005; 1,056]	0,018
pos_graduacao_computacao (SIM)	-0,133	0,021	38,369	1	0,875	[0,839; 0,913]	0,000

A Tabela 4.13 mostra que os fatores identificados pelo método de Kaplan-Meier como significativos para prever a sobrevivência, também se mostraram significativos pela regressão de Cox, com exceção de cotista. Os ingressos por vestibular apresentaram risco

cerca de 1,14 vezes maior (valor de $\text{Exp}(B)$) de evasão em relação aos ingressos por ENEM/SISU, enquanto os ingressos por ENEM apresentaram risco 13,4% menor (1 menos o valor 0,866 de $\text{Exp}(B)$ multiplicado por 100) em relação aos ingressos por vestibular. Os alunos do sexo feminino apresentaram risco cerca de 3% maior de evasão do que os de sexo masculino e os alunos estudantes de instituições que ofertam pós-graduação em Computação apresentaram risco de evasão 12,5% menor que os estudantes que não estudaram em instituições ofertantes de pós-graduação em Computação.

A Tabela 4.14 mostra os resultados da análise multivariada feita a partir da regressão de Cox, considerando-se as instituições públicas:

Tabela 4.14: Análise multivariada da sobrevivência (instituições públicas)

Variáveis	B	SE	Wald	df	Exp(B)	Intervalo de confiança de 95%	p-valor
ingresso_vestibular (SIM)	-0,120	0,023	26,407	1	0,887	[0,847; 0,928]	0,000
ingresso_enem (SIM)	0,209	0,023	84,040	1	1,232	[1,178; 1,288]	0,000
cotista (SIM)	-0,204	0,033	38,881	1	0,816	[0,765; 0,870]	0,000
sexo_aluno (feminino)	0,053	0,022	6,034	1	1,055	[1,011; 1,100]	0,014
pos_graduacao_computacao (SIM)	-0,292	0,018	256,976	1	0,746	[0,720; 0,774]	0,000

A Tabela 4.14 mostra que todos os fatores identificados pelo método de Kaplan-Meier como significativos para prever a sobrevivência, também se mostraram significativos pela regressão de Cox. Nas instituições públicas os fatores associados a menores riscos de evasão são o ingresso por vestibular (11,3% menor em relação ao ENEM), o ingresso por cota (18,4% menor em relação ao não cotista) e o fato da instituição ofertar pós-graduação em Computação (23,6% menor em relação à instituição que não oferta pós-graduação em Computação). Os fatores associados a maiores riscos de evasão são o ingresso por ENEM/SISU (23,2% maior que o ingresso por vestibular) e o aluno ser do sexo feminino (5,5% maior que o sexo masculino).

No Capítulo 1, foi levantada a hipótese de que o sexo, a forma de ingresso na instituição e o fato de ser ou não cotista afetam as taxas de evasões nos cursos da área de Computação. Os gráficos de Análise de Sobrevivência até aqui apresentados, mostram uma evidência de que a evasão varia em função desses fatores.

4.2 Estudo dos Dados da UnB

A seguir são apresentados os estudos feitos utilizando os dados da UnB, empregando-se técnicas de Análise de Sobrevivência, mineração de regras de associação com algoritmo

Apriori e análise dos formulários aplicados aos alunos evadidos da UnB para levantamento de causas de evasão.

4.2.1 Análise de Sobrevivência

As curvas de sobrevivência apresentadas a seguir foram obtidas a partir dos registros de 2875 alunos, que ingressaram na UnB no período 2005-2015 nos cursos superiores de Ciência da Computação, Licenciatura em Computação, Engenharia de Computação e Engenharia de Software.

A Figura 4.22 exibe as curvas de sobrevivência comparando os sexos masculino e feminino:

Figura 4.22: Curvas de sobrevivência (sexo) - UnB

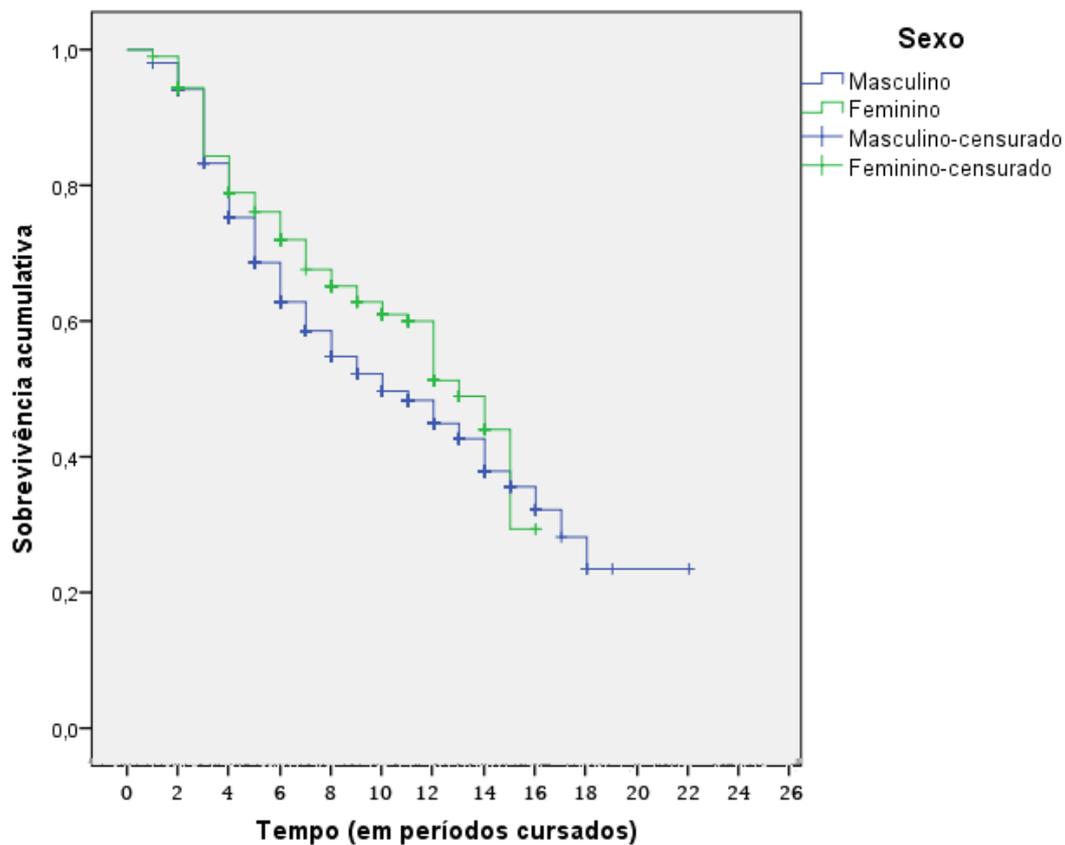


Tabela 4.15: Estatísticas - curvas de sobrevivência (sexo) - UnB

	Qui-quadrado	Graus de liberdade	Sig. (p-valor)
Log Rank (Mantel-Cox)	6,869	1	0,009
Breslow (Generalized Wilcoxon)	5,801	1	0,016
Tarone-Ware	6,897	1	0,009

A Figura 4.22 mostra as curvas de sobrevivência para os sexos masculino e feminino. Foram usados 2573 registros de estudantes do sexo masculino e 302 registros de estudantes do sexo feminino. A curva correspondente ao sexo feminino encontra-se acima da curva do sexo masculino para a maioria dos períodos avaliados, indicando uma sobrevivência superior quanto ao evento evasão. Os 3 testes estatísticos empregados revelam que há diferenças estatísticas significativas entre a sobrevivência apresentada entre os sexos. A evasão apresentada nos sexos masculino e feminino é praticamente a mesma até o terceiro período, quando a partir do quarto a evasão começa a se mostrar maior em meio ao público masculino. A evasão no sexo feminino cessa a partir do período 16, enquanto a evasão no sexo masculino cessa a partir do período 18. A decisão pela evasão é mais rápida no sexo masculino. No caso dos homens, no quarto período há cerca de 75% de sobrevivência enquanto para as mulheres a sobrevivência cai a 75% apenas a partir do quinto período. No décimo período verifica-se uma sobrevivência masculina de cerca de 50%, enquanto a sobrevivência feminina chega a 50% apenas no décimo segundo período. O que se verifica é que a evasão acontece mais rapidamente no início do curso (os quatro primeiros períodos) e a medida que os alunos vão avançando no curso tende a ocorrer mais lentamente.

A Figura 4.23 exhibe as curvas de sobrevivência comparando os ingressos por sistema de cotas e aqueles que ingressaram sem cotas:

Figura 4.23: Curvas de sobrevivência (cotista) - UnB

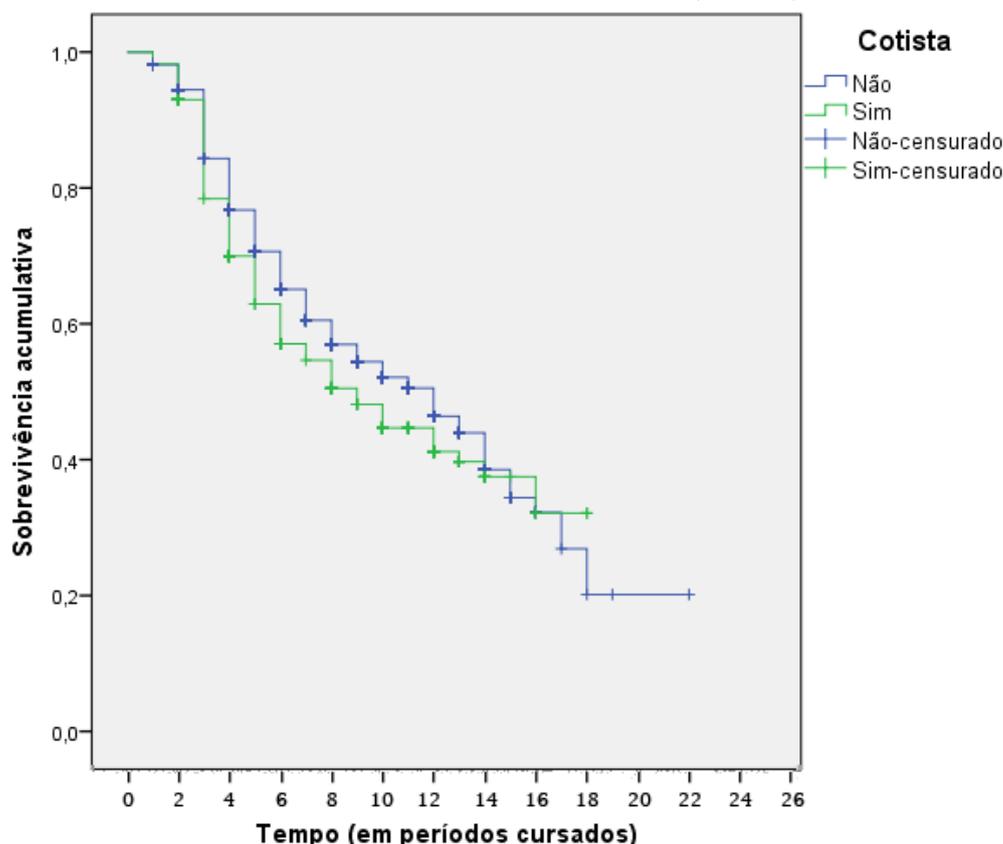


Tabela 4.16: Estatísticas - curvas de sobrevivência (cotista) - UnB

	Qui-quadrado	Graus de liberdade	Sig. (p-valor)
Log Rank (Mantel-Cox)	6,812	1	0,009
Breslow (Generalized Wilcoxon)	9,802	1	0,002
Tarone-Ware	9,180	1	0,002

Para gerar as curvas de sobrevivência da Figura 4.23 foram usados 2361 registros de não cotistas e 514 registros de cotistas. Percebe-se que a curva correspondente aos não cotistas encontra-se acima da curva referente aos cotistas, para a maioria dos períodos avaliados. Esse resultado diverge do resultado encontrado quando se traçou as curvas de sobrevivência para todos os cursos de Computação no país usando os dados do INEP, em que os cotistas mostraram sobrevivência maior à evasão.

A Figura 4.24 exibe as curvas de sobrevivência comparando as diferentes formas de ingresso:

Figura 4.24: Curvas de sobrevivência (ingresso) - UnB

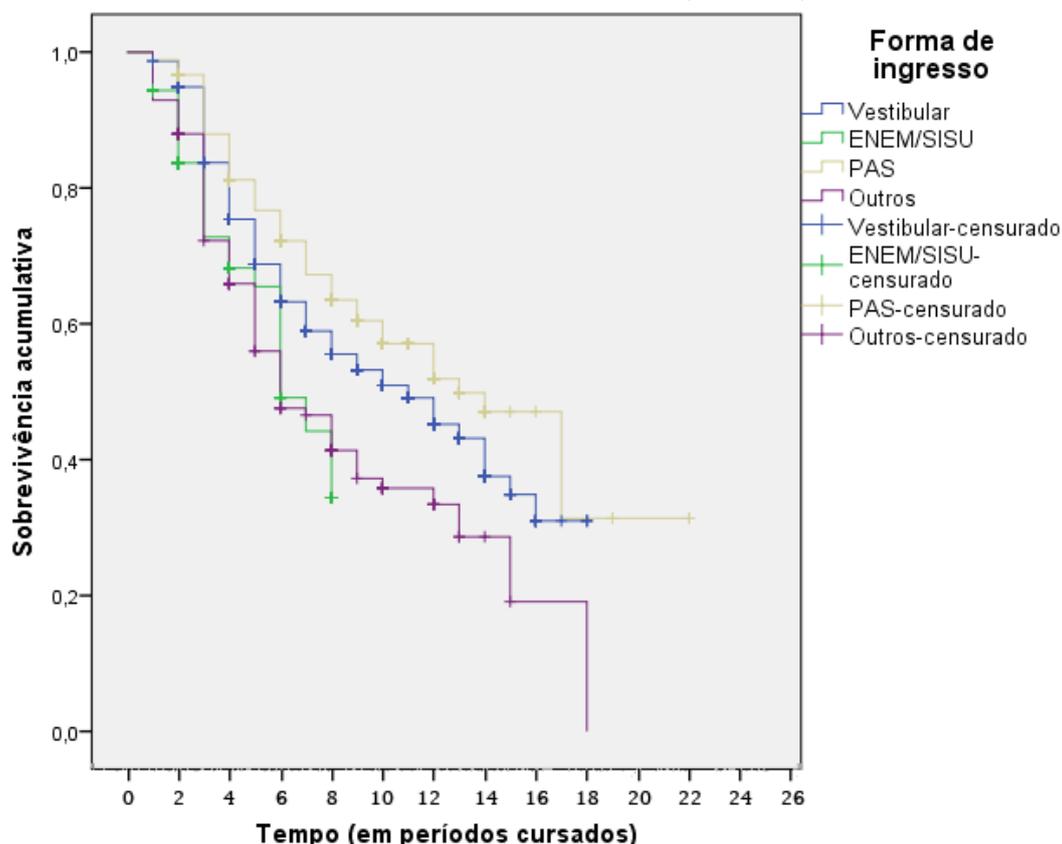


Tabela 4.17: Estatísticas - curvas de sobrevivência (ingresso) - UnB

	Qui-quadrado	Graus de liberdade	Sig. (p-valor)
Log Rank (Mantel-Cox)	47,833	1	0,000
Breslow (Generalized Wilcoxon)	53,975	1	0,000
Tarone-Ware	51,387	1	0,000

Para gerar as curvas de sobrevivência exibidas na Figura 4.24 foram usados 1911 registros de ingressos por vestibular, 195 registros de ingressos por ENEM/SISU, 628 registros de ingressos por PAS e 141 registros de ingressos por outras formas (como transferências). O que se percebe é que a curva correspondente à forma de ingresso por PAS está posicionada acima das demais evidenciando uma sobrevivência superior. As curvas posicionadas mais abaixo indicando menor sobrevivência são as de ENEM/SISU e outras formas de ingresso. A curva de sobrevivência relativa à forma de ingresso por ENEM/SISU só atinge até o oitavo período enquanto as demais se prolongam bem mais. Isso ocorre por que esta forma de ingresso é relativamente recente. Anteriormente a 2010 não houve ingressos

usando o SISU. Como os dados usados compreendem os ingressos no período 2005-2015, é natural que a curva do ENEM/SISU seja mais curta.

A Figura 4.25 exibe as curvas de sobrevivência comparando aqueles discentes que não sofreram nenhuma reprovação na disciplina Cálculo 1 e aqueles que sofreram pelo menos uma reprovação nessa disciplina:

Figura 4.25: Curvas de sobrevivência (disciplina Cálculo 1) - UnB

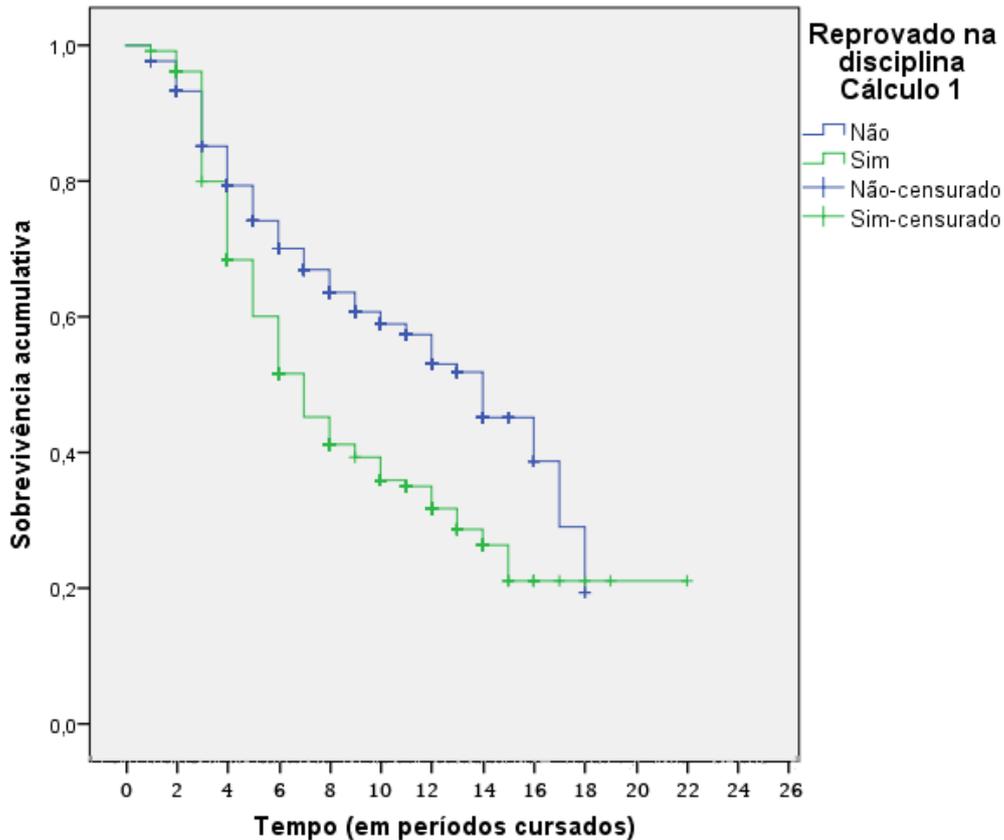


Tabela 4.18: Estatísticas - curvas de sobrevivência (disciplina Cálculo 1) - UnB

	Qui-quadrado	Graus de liberdade	Sig. (p-valor)
Log Rank (Mantel-Cox)	90,552	1	0,000
Breslow (Generalized Wilcoxon)	59,721	1	0,000
Tarone-Ware	77,041	1	0,000

Para gerar as curvas de sobrevivência exibidas na Figura 4.25 foram usados 1939 registros de alunos que não reprovaram em Cálculo 1 e 936 registros de alunos que reprovaram na disciplina Cálculo 1. A reprovação na disciplina de Cálculo 1 foi avaliada

pois esta é uma disciplina geralmente estudada no início do curso (primeiro ou segundo período) e deseja-se analisar se isso tem impacto significativo sobre a evasão. Além disso, há evidências de que o insucesso em disciplinas matemáticas diminui a probabilidade de permanência do aluno no curso. Como se pode perceber os reprovados em Cálculo 1 apresentaram sobrevivência significativamente inferior aos não reprovados. A evasão em meio aos reprovados passa a ser maior a partir do terceiro período. Verifica-se que, a partir do terceiro período, a distância entre as curvas aumenta a medida que os períodos vão passando. Deve-se ressaltar que há disciplinas que são vistas posteriormente, no decorrer do curso como Cálculo 2, que dependem essencialmente do que foi estudado em Cálculo 1 e o adiamento da evasão nos primeiros períodos pode levar a uma situação que o aluno não consegue sustentar quando se depara com outras disciplinas que dependem da matemática.

A Figura 4.26 exibe as curvas de sobrevivência comparando aqueles discentes que não sofreram nenhuma reprovação em disciplinas de algoritmos e aqueles que sofreram pelo menos uma reprovação nessas disciplinas:

Figura 4.26: Curvas de sobrevivência (disciplinas de algoritmos) - UnB

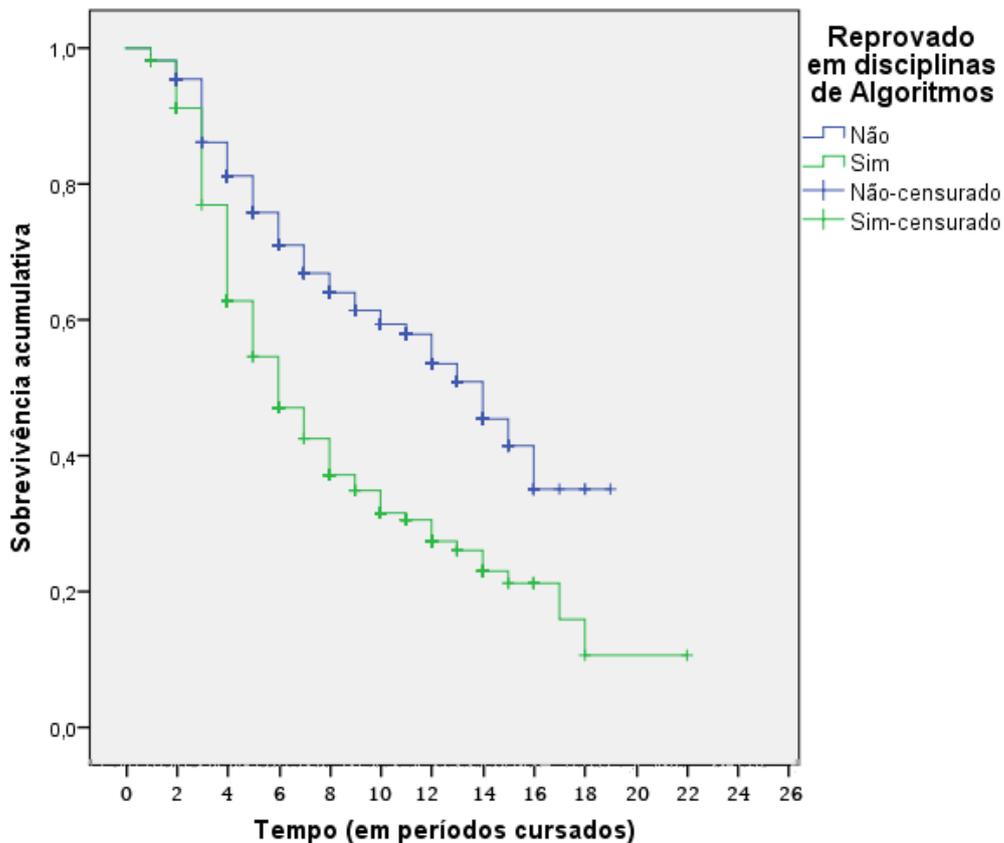


Tabela 4.19: Estatísticas - curvas de sobrevivência (disciplinas de algoritmos) - UnB

	Qui-quadrado	Graus de liberdade	Sig. (p-valor)
Log Rank (Mantel-Cox)	161,578	1	0,000
Breslow (Generalized Wilcoxon)	142,837	1	0,000
Tarone-Ware	157,323	1	0,000

Para exibir as curvas de sobrevivência exibidas na Figura 4.26 foram usados 854 registros de reprovados em disciplinas de algoritmos e 2021 registros de não reprovados. Foram consideradas as disciplinas vistas nos primeiros períodos que tratam de lógica algorítmica e que trazem conceitos essenciais para disciplinas de programação vistas em períodos posteriores do curso. Há evidências levantadas por diversos autores de que o insucesso em disciplinas de algoritmos vistas no início do curso é um fator que aumenta a probabilidade de evasão nos cursos de Computação. Conforme se observa, a curva de sobrevivência correspondente à não reprovação em disciplinas de algoritmos está posicionada acima da curva de reprovação. Há uma diferença estatisticamente significativa entre as duas conforme os testes apresentados na Tabela 4.19. A evasão em meio aos reprovados em disciplinas de algoritmos passa a ser maior já a partir do segundo período, sendo uma evidência de que o mau rendimento em disciplinas de lógica computacional motiva a evasão já no primeiro ano de curso, fato que vai se agravando com o decorrer do curso, pois, conforme se verifica, a medida que os períodos vão avançando a distância entre as curvas aumenta.

A análise de regressão de Cox foi usada para avaliar as variáveis significativas, de acordo com o método de Kaplan-Meier, determinando o risco de cada uma em relação à sobrevivência. Algumas considerações para a interpretação dos resultados da Tabela 4.20: A coluna B traz os coeficientes de regressão. Valores positivos nesta coluna indicam que a variável correspondente está contribuindo como previsor positivo para a evasão, ou seja, valores positivos de B indicam que para aquela variável observa-se evasão maior. $\text{Exp}(B)$ é a razão de chances, ou seja, a razão de risco para uma cada variável. Também temos o erro padrão de B (SE), o seu valor de significância de teste de Wald, os graus de liberdade (df), e o valor de significância do coeficiente (p-valor):

Tabela 4.20: Análise multivariada da sobrevivência (UnB)

Variáveis	B	SE	Wald	df	Exp(B)	Intervalo de confiança de 95%	p-valor
sexo (feminino)	-0,226	0,101	5,005	1	0,798	[0,654; 0,972]	0,025
cotista (sim)	0,044	0,080	0,305	1	1,045	[0,894; 1,222]	0,581
ingresso (vestibular)	0,217	0,079	7,636	1	1,243	[1,065; 1,450]	0,006
ingresso (SISU)	0,686	0,151	20,665	1	1,987	[1,478; 2,671]	0,000
ingresso (outros)	0,769	0,132	33,648	1	2,157	[1,663; 2,796]	0,000
reprovado_calculo_1 (sim)	0,417	0,061	46,421	1	1,517	[1,346; 1,711]	0,000
reprovado_algoritmo (sim)	0,607	0,061	97,851	1	1,835	[1,627; 2,070]	0,000

A Tabela 4.20 mostra que os fatores identificados pelo método de Kaplan-Meier como significativos para prever a sobrevivência, também se mostraram significativos pela regressão de Cox, com exceção de cotista ($p\text{-valor} > 0,05$). O fator associado a um menor risco de evasão é: sexo feminino: 20,2% menos chance de evasão em relação ao sexo masculino. Os fatores associados a um maior risco de evasão são: reprovação na disciplina cálculo 1: aqueles que reprovaram alguma vez nessa disciplina possuem 51,7% mais chances de evadirem em relação ao aluno que nunca reprovou nessa disciplina; reprovação em disciplinas de algoritmos: os reprovados em disciplinas de algoritmos apresentam 83,5% mais chances de evadirem; as formas de ingresso apresentadas na tabela foram comparadas em relação à forma de ingresso via PAS que no gráfico gerado pelo método de Kaplan-Meier apresentou a maior sobrevivência: os ingressos por vestibular apresentaram chances 1,243 vezes maior de evadir em relação aos ingressos por PAS; os ingressos por ENEM/SISU apresentaram chances 1,987 vezes maior de evadir em relação aos ingressos por PAS; os ingressos por outras formas apresentaram chances 2,157 maiores de evadir em relação aos ingressos via PAS.

4.2.2 Mineração de Regras de Associação

O software utilizado para aplicação do algoritmo Apriori foi o R versão 3.1.2. Foram consideradas como variáveis para obtenção das regras de associação as mesmas usadas para obter as curvas de sobrevivência, com exceção da variável que identifica se a instituição oferta pós-graduação em Computação e da variável cotista. Esta variável foi removida por que, de acordo com a análise multivariada usando a regressão de Cox, ela não se mostrou significativa no estudo dos dados da UnB.

Ao executar o algoritmo Apriori via software R, o suporte mínimo foi definido em 1% e a confiança mínima em 70%. Definindo-se um suporte pequeno como este, procurou-se na base de dados regras mais específicas, que estejam em menor evidência. A saída obtida pelo algoritmo Apriori com essas configurações mínimas de suporte em 1% e de confiança em 70% resultou em 723 regras. Contudo, muitas destas regras não trazem novas informações e, portanto, foram omitidas nesse estudo. Por exemplo:

Exemplo 1: INGRESSO=VESTIBULAR => SEXO=MASCULINO

Suporte = 0,60; Confiança = 0,90

Exemplo 2: EVADIDO=SIM => INGRESSO=VESTIBULAR

Suporte = 0,28; Confiança = 0,70

A regra do exemplo 1 diz que quem ingressou pelo Vestibular é também do sexo masculino, com um suporte de 60% e uma confiança de 90%. Essa regra não agrega nenhuma informação relevante a este estudo, pois se pretende elucidar as causas em torno da evasão. Por outro lado, a regra do exemplo 2, ainda que traga uma informação inerente à evasão, não é uma informação nova, pois, como a maioria dos alunos no período 2005-2015 ingressou via Vestibular, é natural que o maior percentual de evadidos seja correspondente a essa forma de ingresso. A regra 2 também poderia levar à dedução equivocada de que os ingressos por vestibular evadem mais que os ingressos por outras formas de seleção.

A coluna *Lift* apresentada na Tabela 4.21 mede o quão frequente o item da esquerda e o item da direita ocorrem juntos se comparados a se fossem estatisticamente independentes. Quando o *lift* é igual a 1 isso indica independência entre o item da direita e o item da esquerda. Um *lift* inferior a 1 indica uma dependência negativa entre o item da direita e o da esquerda e um *lift* superior a 1 indica dependência positiva entre esses itens. Quando essa dependência é positiva a presença do item da direita aumenta as chances de ocorrer o item da esquerda. Sendo assim, valores para o *lift* maiores do que 1 são mais interessantes e na Tabela 4.21 regras com *lift* inferior a 1 foram omitidas.

Dentre as regras obtidas, muitas delas são especificações de regras mais gerais. Por exemplo: para a regra mais geral $A \rightarrow C$ podemos ter regras mais específicas como $A, B \rightarrow C$. De acordo com Burkle [13] se o comportamento descrito pela regra mais geral for uniforme em relação ao comportamento das regras mais específicas então a regra mais geral deve ser mantida e as mais específicas devem ser descartadas. No entanto, se nem todas as regras específicas seguem o mesmo comportamento da regra geral, podemos considerar as regras mais específicas e descartar a mais geral. A Fórmula 4.1 [13] serve

para determinar a cobertura da regra mais geral (CRg):

$$CRg = \sum \text{suporte}(\text{antecedente}(RE)) \div \text{suporte}(\text{antecedente}(RG)) \quad (4.1)$$

onde RE equivale a regra específica e RG equivale a regra geral. A Fórmula 4.1 consiste portanto em somar os suportes dos itens da esquerda das regras mais específicas e dividir pela suporte do item da esquerda da regra mais geral. De acordo com Burkle [13] a medida CRg “pode ser interpretada como a probabilidade condicional de uma instância satisfazer o antecedente de uma das regras mais específicas, dado que a instância satisfaz o antecedente da regra mais geral”. Quanto mais alto for o valor da CRg mais uniforme é o comportamento das regras específicas. Sendo assim, se o valor da CRg for alto (maior do que o valor da confiança mínima estipulada), o comportamento da regra mais geral é considerado uniforme e as regras específicas são descartadas. Para obter o suporte da regra antecedente (regra da esquerda), basta dividir o suporte da regra pela confiança da regra. A Fórmula 4.1 foi aplicada para verificar a cobertura das regras mais gerais em relação às regras mais específicas. Todos os valores obtidos foram superiores à confiança mínima estipulada (70%), sendo mantidas as regras mais gerais e descartadas as regras mais específicas. Um exemplo de regra geral e regras mais específicas é apresentado a seguir:

Regra geral (RG):

REPROVADO_CALCULO_1=NÃO, PERIODOS_1_A_4=NÃO => EVADIDO=NAO
Suporte=0,30; Confiança=0,75

Regras específicas (RE):

RE₁: REPROVADO_CALCULO_1=NÃO, PERIODOS_1_A_4=NÃO, SEXO= FEMININO => EVADIDO=NAO
Suporte=0,04; Confiança = 0,80

RE₂: REPROVADO_CALCULO_1=NÃO, PERIODOS_1_A_4=NÃO, SEXO= MASCULINO => EVADIDO=NAO
Suporte=0,26; Confiança=0,74

Aplicando a Fórmula 4.1 para verificar se RE₁ e RE₂ são uniformes em relação à regra geral, obtém-se o valor 1, que é superior à confiança mínima (70%). Isso mostra que o comportamento descrito pela regra geral é uniforme em relação ao sexo do aluno.

Uma forma de verificar a significância estatística da regra é aplicar um teste de distribuição binomial, conforme proposto por Hamalainen e Nykanen [51]. O p_{random} , apresen-

tado na Fórmula 4.2 [51], corresponde à probabilidade de se obter, ao acaso, o conjunto de transações que cobrem a regra. Para as regras geradas nesse estudo, é considerado o nível de significância de 5%. As regras com valores superiores ao nível de significância estipulado devem ser descartadas. De acordo com Hamalainen e Nykanen [51] a significância estatística obtida com o teste binomial pode ser usada da seguinte forma: define-se um nível de significância (valor máximo de p_{random}) e considera-se as regras com valores para p_{random} abaixo do máximo estipulado. O teste binomial pode ser realizado de acordo com a Fórmula 4.2 [51]:

$$p_{random} = \sum_{i=m(X,A)}^{m(X)} \binom{n}{i} (P(X)P(A))^i (1 - P(X)P(A))^{n-i} \quad (4.2)$$

Onde: $m(X)$ = número de linhas em que ocorre o item da esquerda; i = número de linhas em que ocorrem os itens da esquerda e da direita; n = número de casos, que nesse estudo pode ser obtido a partir do suporte do item da esquerda; $P(X)P(A)$ = probabilidade de ocorrência do item da esquerda e da direita, que é obtido a partir da confiança da regra.

Na Tabela 4.21 é possível visualizar as regras com os respectivos valores de suporte, confiança, lift e p_{random} :

Tabela 4.21: Regras de associação (cursos de Computação da UnB)

Num regra	Item da direita	Item da esquerda	Sup.	Conf.	Lift	p_{random}
1	{EVADIDO=SIM, INGRESSO=OUTROS}	{REPROVADO_CALCULO1=NÃO}	0,02	0,73	1,09	0,10
2	{EVADIDO=SIM, INGRESSO=OUTROS}	{REPROVADO_ALGORITMO=NÃO}	0,02	0,72	1,03	0,10
3	{SEXO=FEMININO, PERIODOS_1_A_4=NÃO}	{EVADIDO=NÃO}	0,05	0,75	1,26	0,07
4	{SEXO=FEMININO}	{REPROVADO_ALGORITMO=NÃO}	0,08	0,73	1,04	0,05
5	{INGRESSO=PAS}	{REPROVADO_ALGORITMO=NÃO}	0,16	0,71	1,02	0,03
6	{EVADIDO=SIM, REPROVADO_ALGORITMO=SIM}	{SEXO=MASCULINO}	0,16	0,92	1,03	0,06
7	{EVADIDO=SIM, REPROVADO_CALCULO1=SIM}	{SEXO=MASCULINO}	0,16	0,90	1,01	0,06
8	{REPROVADO_CALCULO1=NÃO, PERIODOS_1_A_4=NÃO}	{EVADIDO=NÃO}	0,30	0,75	1,26	0,03
9	{REPROVADO_ALGORITMO=NÃO, PERIODOS_1_A_4=NÃO}	{EVADIDO=NÃO}	0,31	0,73	1,24	0,00
10	{PERIODOS_1_A_4=NÃO}	{REPROVADO_ALGORITMO=NÃO}	0,43	0,75	1,07	0,02
11	{EVADIDO=NÃO}	{REPROVADO_CALCULO1=NÃO}	0,45	0,75	1,12	0,02
12	{EVADIDO=NÃO}	{REPROVADO_ALGORITMO=NÃO}	0,47	0,79	1,12	0,02
13	{REPROVADO_CALCULO1=NÃO}	{REPROVADO_ALGORITMO=NÃO}	0,51	0,76	1,08	0,02
14	{REPROVADO_ALGORITMO=NÃO}	{REPROVADO_CALCULO1=NÃO}	0,51	0,73	1,08	0,02

A significância estatística das regras é exibida na coluna p_{random} da Tabela 4.21. Algumas regras apresentaram um valor para p_{random} superior ao limite estabelecido, sendo as seguintes: 1, 2, 3, 4, 6 e 7. Sendo assim, essas regras não são comentadas:

- Regra 5: 71% dos que ingressam por PAS não reprovam em Algoritmos.
- Regra 8: 75% dos que não reprovam em Cálculo 1 e que cursam mais de 4 períodos também não evadem. Isso mostra que o sucesso na disciplina Cálculo 1 aumenta a probabilidade de permanência no curso.
- Regra 9: 73% dos não reprovados em Algoritmos que cursam mais de 4 períodos também não evadem. Isso mostra que o sucesso em disciplinas de algoritmos aumenta a probabilidade de permanência no curso.
- Regra 10: 75% dos que cursam mais de 4 períodos não reprovam em Algoritmos. Isso mostra que quem permanece no curso por mais tempo, em sua maioria, são os não reprovados em Algoritmos.
- Regra 11: 75% dos não evadidos não reprovam em Cálculo 1. Isso mostra como o sucesso nesta disciplina contribui para a permanência do aluno no curso.
- Regra 12: 76% dos não evadidos não reprovam em Algoritmos. O sucesso em Algoritmos é portanto importante para manter o aluno no curso.
- Regra 13: 76% dos não reprovados em Cálculo 1, também não reprovam em Algoritmos. Disciplinas de algoritmos tem uma relação estreita com matemática e um mau rendimento em disciplinas como Cálculo 1 pode levar a um mau rendimento em Algoritmos.
- Regra 14: 73% dos não reprovados em Algoritmos, também não reprovam em Cálculo 1. Tal como foi falado para a regra 17, Algoritmos e Matemática estão intimamente relacionados.

As regras obtidas com o algoritmo Apriori apontam principalmente para o fato de que o êxito em disciplinas de algoritmos e na disciplina Cálculo 1 contribui para a não evasão discente. O êxito em Cálculo 1 também contribui para o êxito em disciplinas de algoritmos e vice-versa. Constatou-se que o percentual de ingressantes por PAS que não reprovam em Algoritmos é maior do que o de ingressantes por ENEM/SISU e por vestibular que não reprovam em Algoritmos, visto que para as formas de ingresso por ENEM/SISU e por Vestibular não se encontrou uma regra desse tipo. Um detalhamento das causas de evasão de acordo com os próprios alunos evadidos é feito na próxima subseção.

4.2.3 Causas de Evasão

Nesta subseção são apresentadas as análises das respostas dos alunos evadidos nos seguintes cursos superiores de Computação ofertados pela UnB: Bacharelado em Ciência da Computação, Licenciatura em Computação, Engenharia de Computação e Engenharia de Software.

Foram enviados questionários aos alunos evadidos nos cursos de Computação da UnB, no período compreendido entre 2005 e 2015. Um dos parâmetros para a fórmula do cálculo amostral é a prevalência dos motivos relacionados a evasão. Assumiu-se nesse caso, que se está trabalhando com uma população heterogênea. Sendo assim, estimou-se a prevalência das razões de evasão em 50%. Para cálculo do tamanho das amostras em cada curso, a Fórmula 4.3 foi utilizada [38]:

$$n_0 = \left[Z_{\alpha/2}^2 \cdot p(1-p) / E^2 \right] \quad (4.3)$$

Em que, $p=50,0\%$, que corresponde à prevalência estimada das razões de evasão; $Z_{\alpha/2} = 1,645$, que é o valor crítico que corresponde ao grau de confiança adotado de 90%; $E = 10\%$, que corresponde à margem de erro adotada. Após obter n_0 , foi feita a correção para uma população finita, em que N corresponde ao total de evadidos em cada curso. A Fórmula 4.4 foi empregada para cada curso, definindo-se o tamanho da amostra em cada um deles [38]:

$$n = n_0 \cdot \{N / [n_0 + (N - 1)]\} \quad (4.4)$$

A Tabela 4.22 mostra a quantidade de alunos em cada curso no período 2005-2015, a quantidade de evadidos em cada curso, o tamanho mínimo definido para a amostra e a quantidade de questionários respondidos:

Tabela 4.22: Dimensionamento da amostra

Curso	Alunos no período	Evadidos (N)	Amostra (n)	Questionários respondidos
Ciência da Computação	903	368	58	93
Licenciatura em Computação	916	509	60	85
Engenharia da Computação	531	183	50	53
Engenharia de Software	525	112	43	35

A quantidade de questionários respondidos correspondente ao curso de Engenharia de Software está destacado pois, infelizmente, para esse curso não se alcançou o tamanho mínimo da amostra de acordo com os parâmetros definidos. Apesar do contato feito com os alunos via e-mail e via telefone, muitos não responderam o questionário. Portanto,

os resultados da pesquisa para o curso de Engenharia de Software são exibidos com esta ressalva.

Na Tabela 4.23 podem ser vistos os dados demográficos dos respondentes da pesquisa em cada curso, onde CC corresponde à Ciência da Computação, LC à Licenciatura em Computação, ES à Engenharia de Software e EC à Engenharia de Computação:

Tabela 4.23: Dados demográficos da amostra (% de respondentes)

	CC	LC	ES	EC
Sexo				
Feminino	7,53	8,24	8,57	13,21
Masculino	92,47	91,76	91,43	86,79
Cotas				
Cotistas	16,13	9,41	14,29	18,87
Não cotistas	83,87	90,59	85,71	81,13
Ingresso				
ENEM/SISU	5,38	12,94	0,00	5,66
Vestibular	70,97	67,06	71,43	79,25
PAS	16,13	16,47	28,57	15,09
Outros	7,53	3,53	0,00	0,00
Recebimento de auxílio estudantil				
SIM	5,38	8,24	14,29	15,09
NÃO	94,62	91,76	85,71	84,91
Primeira formação superior				
Sim	73,12	63,53	82,86	86,79
Não	26,88	36,47	17,14	13,21
Período de desligamento				
1º Período	8,60	14,12	0,00	15,09
2º Período	10,75	12,94	0,00	5,66
3º Período	29,03	11,76	11,43	15,09
4º Período	12,90	20,00	14,29	28,30
5º Período	13,98	14,12	22,86	13,21
6º Período	10,75	7,06	17,14	11,32
7º Período	4,30	1,18	17,14	1,89
8º Período	1,08	4,71	2,86	9,43
9º Período	1,08	2,35	2,86	0,00
10º Período	1,08	5,88	2,86	0,00
Após 10º Período	6,45	5,88	8,57	0,00

Como se pode perceber pela análise da Tabela 4.23, a maioria dos respondentes em

todos os cursos é do sexo masculino, o que é natural, visto que a maioria das matrículas nos cursos de Computação são de homens. Verifica-se também que a maior parte dos respondentes ingressou por acesso universal (sem cotas), via vestibular, não recebeu auxílio estudantil ao longo do curso e o curso de Computação feito na UnB foi sua primeira formação superior. Quanto ao período em que evadiu do curso, a maioria (mais de 50%) dos alunos dos cursos de Ciência da Computação, Licenciatura em Computação e Engenharia de Computação respondeu ter evadido entre o primeiro e o quarto períodos. Isso está de acordo com o verificado nas curvas de sobrevivência, em que há picos maiores de saída no terceiro e quarto períodos. Para o curso de Engenharia de Software, percebeu-se um comportamento diferente quanto ao período de saída, sendo que não houve respondentes afirmando terem saído nos primeiros períodos. Quanto a essa diferença, deve-se levar em consideração que no curso de Engenharia de Software da UnB, os alunos cursam inicialmente diversas disciplinas comuns a várias engenharias, conforme pode ser visto no fluxo do curso apresentado no portal da UnB (<https://matriculaweb.unb.br/graduacao/fluxo.aspx?cod=6009>). São 4 períodos cursando disciplinas de formação geral, comuns a várias engenharias. Após esses 4 períodos o foco se volta para a formação em Computação. E é justamente a partir do quarto período até o sexto período que a maioria dos respondentes diz ter evadido.

Além das informações presentes na Tabela 4.23, foram levantadas as causas de evasão por meio de um formulário de escala tipo Likert, em que cada item foi pontuado pelos respondentes, em uma escala de 0 a 5, como cada um deles influenciou a decisão de evadir do curso. A nota mais baixa, zero, corresponde à nenhuma influência e a nota mais alta, cinco, corresponde à grande influência do item sobre a decisão de evadir. Os itens presentes no formulário likert podem ser visualizados na Tabela 4.24:

Tabela 4.24: Fatores relacionados à evasão

Classificação	Número do item	Descrição do item
Questões Sócio-econômicas	1	Impossibilidade de conciliar o serviço com o estudo
	2	Dificuldades financeiras
	15	Não recebimento de auxílio estudantil
Questões Vocacionais	3	Mudança para um curso de área diferente de Computação
	4	Falta de afinidade com o curso
	5	Dedicação a outros estudos
	6	Falta de identificação com profissionais dessa área
Questões Institucionais/ Acadêmicas	7	Critérios de avaliação dos discentes usados são inadequados ou muito rígidos
	8	Dificuldades em disciplinas do curso ligadas ao ensino de programação e Algoritmos
	9	Dificuldades em disciplinas do curso ligadas à matemática
	10	Dificuldades em outras disciplinas do curso
	11	Falta de informações sobre o que é abordado no curso
	12	Infraestrutura física da instituição inadequada
	13	Dificuldades de relacionamento com os professores/funcionários
	14	Dificuldade dos professores em repassar de maneira compreensível os conteúdos
Questões Familiares/Pessoais/ de Saúde	17	Distância da família
	18	Problemas afetivos
	20	Dificuldade de relacionamento com os colegas
	21	Problemas de saúde (seja seu ou de algum familiar)
Outras Questões	16	Transferência para outra instituição (continuou cursando o mesmo curso em outra instituição)
	19	Mudança de município

Os itens utilizados na pesquisa de causas de evasão foram classificados em questões sócio-econômicas, questões vocacionais, questões institucionais/acadêmicas, questões familiares/pessoais/de saúde e outras questões. Essa classificação é baseada na classificação adotada por Barroso e Falcão (2004) [10], que classifica os principais motivos que levam à evasão em cursos superiores em sócio-econômicos, vocacionais e institucionais e procura elencar também outras questões que não se encaixam bem nessas 3 classificações. A Tabela 4.24 trás também a numeração dos itens, que corresponde à ordem em que os itens foram apresentados aos alunos evadidos ao responderem a pesquisa.

Para verificar se as classificações de itens usados na pesquisa estão satisfatoriamente relacionados à classificação adotada para eles, usou-se o coeficiente alfa de Cronbach. O Alfa de Cronbach “mede a correlação entre respostas em um questionário através da análise do perfil das respostas dadas pelos respondentes. Trata-se de uma correlação média entre perguntas” [26]. “o índice α estima quão uniformemente os itens contribuem para a soma não ponderada do instrumento, variando numa escala de 0 a 1” [65]. Foram

testados um total de 266 casos, correspondentes a soma dos entrevistados nos cursos de Ciências da Computação (93 casos), Licenciatura em Computação (85 casos), Engenharia de Computação (53 casos) e Engenharia de Software (35 casos). Na Tabela 4.25 são exibidos os valores de Alfa de Cronbach usados para cada classificação:

Tabela 4.25: Alfa de Cronbach para cada classificação usada na pesquisa

Classificações	Alfa de Cronbach	Nº de itens
Sócio-econômicas	0,636	3
Vocacionais	0,717	4
Institucionais/Acadêmicas	0,800	8
Familiares/pessoais/de Saúde	0,644	4
Outras Questões	0,457	2

De acordo com Hair Jr. et al.¹ (2005 apud Jardimino et al., 2010) [52] em relação aos valores do Alfa de Cronbach “as medidas mais altas indicam maior confiabilidade entre os indicadores, sendo 0.60 o limite inferior de aceitabilidade para pesquisas de cunho exploratório e 0.70 para os demais tipos de pesquisas”. Assume-se que a investigação feita nesse trabalho é do tipo exploratória. Segundo Godoy² (1995 apud Jardimino et al., 2010) [52] a pesquisa é exploratória pois “se propôs aumentar o conhecimento dos pesquisadores acerca do fenômeno-alvo da pesquisa”. Na Tabela 4.25, percebe-se que todas as classificações empregadas, exceto “Outras Questões” apresentaram Alfa de Cronbach superior à 0,60. Como a classificação “Outras Questões” envolve itens não associados a um tema específico, é compreensível que seu valor tenha se mostrado baixo. No entanto, as classificações mais específicas mostraram valores adequados para uma pesquisa exploratória, mostrando que, na análise geral de todos os entrevistados, as causas de evasão podem ser classificadas de acordo com o que foi proposto.

A seguir os valores obtidos para cada item do formulário de escala Likert, foram comparados pelo teste de Kruskal Wallis, usando o teste post hoc de Dunn para agrupar os grupos homogêneos, com intervalo de confiança a 95%. O teste não paramétrico foi adotado pois as variáveis não atenderam a normalidade, o que foi verificado por meio do teste de Kolmogorov-Smirnov, sendo obtido um valor de $p < 0,05$, rejeitando-se a hipótese nula de que os dados das variáveis se originam de uma distribuição normal.

A Tabela 4.26 mostra o resultado do teste de Kruskal Wallis para o curso de Ciência da Computação, mostrando os grupos homogêneos para os itens de causas de evasão:

¹Hair, Joseph F. et al. Multivariate data analysis. 5th ed. New Jersey: Prentice Hall, 2005.

²Godoy, Arlida Schmidt. Introdução à pesquisa qualitativa e suas possibilidades. Revista de Administração de Empresas, São Paulo, v. 35, n. 2, p. 57-63, 1995.

Tabela 4.26: Teste de Kruskal Wallis (Ciência da Computação)

Postos de Média e Subconjuntos Homogêneos para os itens									
¹ Itens	Subconjunto								
	1	2	3	4	5	6	7	8	9
3	622,586								
17	632,72								
15	696,758	696,758							
16	710,823	710,823	710,823						
2		790,887	790,887	790,887					
20		818,048	818,048	818,048	818,048				
21		819,301	819,301	819,301	819,301				
18			869,726	869,726	869,726				
10				912,613	912,613	912,613			
8				937,242	937,242	937,242	937,242		
13					984,285	984,285	984,285	984,285	
3						1.093,33	1.093,33	1.093,33	1.093,33
6							1.104,43	1.104,43	1.104,43
12							1.113,15	1.113,15	1.113,15
1							1.115,05	1.115,05	1.115,05
11								1.165,84	1.165,84
4								1.174,51	1.174,51
5								1.215,00	1.215,00
9								1.225,69	1.225,69
14									1.253,26
7									1.261,75
Estatística de teste	7,599	9,011	7,836	8,886	10,661	6,689	10,969	13,817	9,247
Sig. (teste de 2 lado(s))	0,055	0,061	0,098	0,114	0,059	0,082	0,052	0,087	0,415
Subconjuntos homogêneos são baseados em significâncias assintóticas. O nível de significância é 0,05.									
¹ Cada célula mostra a posição média de amostra para a variável itens.									

A Tabela 4.26 exibe um teste feito com um N (número total de itens) de 1953 (correspondente a 21 itens x 93 respondentes, que foi o número de respondentes para o curso de Ciência da Computação). A estatística do teste qui-quadrado foi 335,607, com 20 graus de liberdade e um p-valor $< 0,05$, mostrando que há diferenças estatisticamente significativas de pontuação entre os itens. Na Tabela 4.26 vemos um total de 9 subconjuntos homogêneos, onde o maior subconjunto, correspondente aos itens que tiveram maior pontuação na pesquisa para o curso de Ciência da Computação, é o 9. Na Tabela 4.27 o subconjunto 9 é exibido:

Tabela 4.27: Itens do maior subconjunto homogêneo (Ciências da Computação)

Classificação	Itens
Vocacionais	Mudança para um curso de área diferente de Computação
	Falta de identificação com profissionais dessa área
	Falta de afinidade com o curso
	Dedicação a outros estudos
Institucionais	Infraestrutura física da instituição inadequada
	Falta de informações sobre o que é abordado no curso
	Dificuldades em disciplinas do curso ligadas à matemática
	Dificuldade dos professores em repassar de maneira compreensível os conteúdos
	CrITÉrios de avaliação dos discentes usados são inadequados ou muito rÍgidos.
Sócio-econômico	Impossibilidade de conciliar o serviço com o estudo

A Tabela 4.28 mostra o resultado do teste de Kruskal Wallis para o curso de Licenciatura em Computação, mostrando os grupos homogêneos para os itens de causas de evasão:

Tabela 4.28: Teste de Kruskal Wallis (Licenciatura em Computação)

Postos de Média e Subconjuntos Homogêneos para os itens										
¹ Itens	Subconjunto									
	1	2	3	4	5	6	7	8	9	10
17	627,524									
19	652,465									
20	662,147									
16	667,253									
15	669,7									
21	701,071	701,071								
18	779,1	779,1	779,1							
2		828,641	828,641	828,641						
10		848	848	848	848					
6			889,006	889,006	889,006	889,006				
8			891,641	891,641	891,641	891,641				
3			896,941	896,941	896,941	896,941	896,941			
13			921,859	921,859	921,859	921,859	921,859	921,859		
4			975,835	975,835	975,835	975,835	975,835	975,835	975,835	
12				994,847	994,847	994,847	994,847	994,847	994,847	
11					1.041,87	1.041,87	1.041,87	1.041,87	1.041,87	
5						1.055,62	1.055,62	1.055,62	1.055,62	
14							1.098,22	1.098,22	1.098,22	
9								1.116,47	1.116,47	
7									1.159,87	
1										1.274,92
Estatística de teste	9,958	7,623	10,692	10,077	12,778	12,491	10,928	11,304	10,897	. ²
Sig. (teste de 2 lado(s))	0,126	0,054	0,153	0,184	0,078	0,086	0,091	0,079	0,092	.
Subconjuntos homogêneos são baseados em significâncias assintóticas. O nível de significância é ,05.										
¹ Cada célula mostra a posição média de amostra para a variável itens.										
² Não é possível calcular, pois o subconjunto contém somente uma amostra.										

A Tabela 4.28 exibe um teste feito com um N (número total de itens) de 1785 (correspondente a 21 itens x 85 respondentes, que foi o número de respondentes para o curso de Licenciatura em Computação). A estatística do teste qui-quadrado foi 278,360, com 20 graus de liberdade e um p-valor $< 0,05$, mostrando que há diferenças estatisticamente significativas de pontuação entre os itens. Na Tabela 4.28 vemos um total de 9 subconjuntos homogêneos, onde o maior subconjunto, correspondente aos itens que tiveram maior pontuação na pesquisa para o curso de Licenciatura em Computação, é o 10. Contudo, o grupo 10 é formado por apenas um item. Sendo assim, na Tabela 4.29 os subconjuntos 9 e 10 são exibidos:

Tabela 4.29: Itens do maior subconjunto homogêneo (Licenciatura em Computação)

Classificação	Itens
Vocacionais	Falta de afinidade com o curso
	Dedicação a outros estudos
Institucionais	Critérios de avaliação dos discentes usados são inadequados ou muito rígidos.
	Dificuldades em disciplinas do curso ligadas à matemática
	Falta de informações sobre o que é abordado no curso.
	Infraestrutura física da instituição inadequada
Sócio-econômico	Dificuldade dos professores em repassar de maneira compreensível os conteúdos
	Impossibilidade de conciliar o serviço com o estudo

A Tabela 4.30 mostra o resultado do teste de Kruskal Wallis para o curso de Engenharia em Computação, mostrando os grupos homogêneos para os itens de causas de evasão:

Tabela 4.30: Teste de Kruskal Wallis (Engenharia de Computação)

Postos de Média e Subconjuntos Homogêneos para os itens									
¹ Itens	Subconjunto								
	1	2	3	4	5	6	7	8	9
17	317,434								
19	349,17	349,17							
15	361,321	361,321							
16	371,66	371,66							
2	388,755	388,755							
21	409,896	409,896	409,896						
18	428,377	428,377	428,377						
1		440,528	440,528	440,528					
20			522,198	522,198	522,198				
10				536,198	536,198				
9					608,792	608,792			
12					614,113	614,113	614,113		
5					630,679	630,679	630,679	630,679	
8					638,311	638,311	638,311	638,311	638,311
13					644,925	644,925	644,925	644,925	644,925
7						711,604	711,604	711,604	711,604
3						714,019	714,019	714,019	714,019
11						732,934	732,934	732,934	732,934
14							741,774	741,774	741,774
6								765,509	765,509
4									768,802
Estatística de teste	11,559	9,197	7,276	4,666	10,545	13,513	13,771	12,74	12,008
Sig. (teste de 2 lado(s))	0,073	0,163	0,064	0,097	0,104	0,061	0,055	0,079	0,1
Subconjuntos homogêneos são baseados em significâncias assintóticas. O nível de significância é ,05.									
¹ Cada célula mostra a posição média de amostra para a variável itens.									

A Tabela 4.30 exibe um teste feito com um N (número total de itens) de 1113 (correspondente a 21 itens x 53 respondentes, que foi o número de respondentes para o curso de Engenharia de Computação). A estatística do teste qui-quadrado foi 292,670, com 20 graus de liberdade e um p-valor $< 0,05$, mostrando que há diferenças estatisticamente significativas de pontuação entre os itens. Na Tabela 4.30 vemos um total de 9 subconjuntos homogêneos, onde o maior subconjunto, correspondente aos itens que tiveram maior pontuação na pesquisa para o curso de Engenharia de Computação, é o 9. Na Tabela 4.31 o subconjunto 9 é exibido:

Tabela 4.31: Itens do maior subconjunto homogêneo (Engenharia de Computação)

Classificação	Itens
Vocacionais	Mudança para um curso de área diferente de Computação
	Falta de afinidade com o curso
	Falta de identificação com profissionais dessa área
Institucionais	Critérios de avaliação dos discentes usados são inadequados ou muito rígidos.
	Dificuldades em disciplinas do curso ligadas ao ensino de programação e Algoritmos
	Falta de informações sobre o que é abordado no curso.
	Dificuldades de relacionamento com os professores/funcionários
	Dificuldade dos professores em repassar de maneira compreensível os conteúdos

A Tabela 4.32 mostra o resultado do teste de Kruskal Wallis para o curso de Engenharia de Software, mostrando os grupos homogêneos para os itens de causas de evasão:

Tabela 4.32: Teste de Kruskal Wallis (Engenharia de Software)

Postos de Média e Subconjuntos Homogêneos para os itens						
¹ Itens	Subconjunto					
	1	2	3	4	5	6
19	221,043					
17	235,414	235,414				
20	247,914	247,914				
16	254,5	254,5				
18	267,729	267,729	267,729			
15	303,729	303,729	303,729	303,729		
21		310,429	310,429	310,429	310,429	
2			375,129	375,129	375,129	375,129
13				383,129	383,129	383,129
8				388,443	388,443	388,443
3				390,3	390,3	390,3
10				391,771	391,771	391,771
5				393,071	393,071	393,071
6				404,271	404,271	404,271
1				407,557	407,557	407,557
9				413,643	413,643	413,643
4				451,7	451,7	451,7
12					459,457	459,457
11					462,671	462,671
14					477,871	477,871
7						488,229
Estatística de teste	9,02	8,128	6,6	17,036	21,942	17,087
Sig. (teste de 2 lado(s))	0,108	0,149	0,086	0,107	0,056	0,195
Subconjuntos homogêneos são baseados em significâncias assintóticas. O nível de significância é ,05.						
¹ Cada célula mostra a posição média de amostra para a variável itens.						

A Tabela 4.32 exibe um teste feito com um N (número total de itens) de 735 (correspondente a 21 itens x 35 respondentes, que foi o número de respondentes para o curso de Engenharia de Software). A estatística do teste qui-quadrado foi 292,670, com 20 graus de liberdade e um p-valor $< 0,05$, mostrando que há diferenças estatisticamente significativas

de pontuação entre os itens. Na Tabela 4.32 vemos um total de 6 subconjuntos homogêneos, onde o maior subconjunto, correspondente aos itens que tiveram maior pontuação na pesquisa para o curso de Engenharia de Software, é o 6. Na Tabela 4.33 o subconjunto 6 é exibido:

Tabela 4.33: Itens do maior subconjunto homogêneo (Engenharia de Software)

Classificação	Itens
Vocacionais	Mudança para um curso de área diferente de Computação
	Falta de afinidade com o curso
	Dedicação a outros estudos
	Falta de identificação com profissionais dessa área
Institucionais	Critérios de avaliação dos discentes usados são inadequados ou muito rígidos.
	Dificuldades em disciplinas do curso ligadas ao ensino de programação e Algoritmos
	Dificuldades em disciplinas do curso ligadas à matemática
	Dificuldades em outras disciplinas do curso
	Falta de informações sobre o que é abordado no curso.
	Infraestrutura física da instituição inadequada
	Dificuldades de relacionamento com os professores/funcionários
	Dificuldade dos professores em repassar de maneira compreensível os conteúdos
Sócio-econômico	Impossibilidade de conciliar o serviço com o estudo
	Dificuldades financeiras

Analisando-se as questões de maior influência sobre a evasão, de acordo com a percepção dos alunos evadidos, percebe-se a predominância das questões vocacionais, institucionais/acadêmicas e sócio-econômicas. Questões de ordem familiar, pessoal, de saúde e outras questões não tiveram grande pontuação. Algumas questões aparecem em todos os cursos como grandes influenciadoras da evasão, sendo elas:

1. Dificuldade dos professores em repassar de maneira compreensível o conteúdo
2. Falta de informações sobre o que é abordado no curso
3. Critérios de avaliação dos discentes usados são inadequados ou muito rígidos
4. Falta de afinidade com o curso

Destes itens, o 1,2 e 3 são institucionais e o 4 é vocacional. As dificuldades em disciplinas matemáticas também são bem citadas, figurando entre as principais causas nos curso de Ciência da Computação, Licenciatura em Computação e Engenharia de Software. As dificuldades em disciplinas do curso ligadas ao ensino de programação/algoritmos é citada como uma das principais causas de evasão pelos alunos de Engenharia de Computação. As questões sócio-econômicas também aparecem nos cursos de Ciência da Computação, Li-

enciatura em Computação e Engenharia de Software, sendo que o item “Impossibilidade de conciliar o serviço com o estudo” figura entre as principais causas nestes 3 cursos.

Além dos itens do formulário de escala Likert, os alunos também puderam acrescentar algum outro motivo relevante para a sua saída que não houvesse sido descrito na pesquisa. Os motivos citados encontram-se relacionados no Apêndice B desse documento. Alguns alunos ao apontarem outros motivos de evasão na verdade apenas detalharam um item já existente no formulário de escala. Por exemplo, especificou um problema de saúde que possuía. A Tabela 4.34 exhibe a classificação dos motivos descritos pelos alunos:

Tabela 4.34: Porcentagem das respostas conforme classificação adotada

Classificação das questões	Cursos			
	Ciência da Computação	Licenciatura em Computação	Engenharia de Computação	Engenharia de Software
Vocacionais	7,69	3,70	0,00	0,00
Sócio-econômicas	0,00	14,81	5,00	7,14
Familiares/pessoais/de saúde	7,69	14,81	5,00	21,43
Institucionais/Acadêmicas	42,31	62,96	60,00	42,86
Outras	42,31	14,81	30,00	35,71

Grande parte das razões acrescentadas pelos discentes evadidos como causas de evasão podem ser classificadas como Institucionais/acadêmicas. Dos 93 alunos de Ciência da Computação que responderam a pesquisa, 26 acrescentam motivos para evasão; dos 85 alunos de Licenciatura em Computação respondentes, 27 acrescentaram motivos para evasão; dos 53 alunos de Engenharia de Computação respondentes, 20 acrescentaram motivos para evasão; na Engenharia de Software, dos 35 alunos respondentes, 14 acrescentaram motivos para evasão. Conforme dito, algumas razões colocadas pelos alunos foram apenas um detalhamento dos itens já previstos no formulário Likert. Contudo, algumas questões apontadas por eles não estavam previstas e estão relacionadas na Tabela 4.35, onde CC = Ciência da Computação, LC = Licenciatura em Computação, EC = Engenharia de Computação, ES = Engenharia de Software:

Tabela 4.35: Itens acrescentados pelos alunos como causas de evasão

Causas	Cursos			
	CC	LC	EC	ES
Baixa qualidade docente	X	X		X
professores com baixo interesse		X	X	
Professores querem competir com os alunos				X
Excesso de cobrança dos professores				X
Falta de base estudantil		X	X	X
Currículo desatualizado/ementa inadequada	X	X	X	X
Disciplinas teóricas com pouca relevância para a área/Falta de abordagem prática	X		X	
Falta de apoio e orientação acadêmica		X	X	X
conteúdo trivial/superfial	X	X		
Pré-requisitos desnecessários	X			
Falta de adequação da academia com o mercado de trabalho	X	X	X	
Falta de flexibilidade no abono de faltas		X		
Professores dão maior ênfase à pesquisa do que ao ensino			X	
Disciplina apenas focadas em provas e não na aprendizagem	X			
Exigência excessiva em disciplinas como Física e Cálculo	X			
Problemas com trâmites institucionais	X			
Reprovações sucessivas	X	X	X	
Dificuldade de fazer matrícula em disciplinas	X			
Falta de estímulo para o TCC / Dificuldade em concluir o TCC		X		
Burocracia nos serviços de secretaria		X	X	
Falta de um curso de bacharelado noturno		X		
falta de grupos de estudo		X		
Falta de Computador	X		X	
Dificuldade de transporte para a UnB	X			X
Conhecimentos suficientes sem ter que fazer o curso	X			
Já possuía outro curso superior de Computação	X			
Já tinha uma graduação	X			
Procrastinação	X			
Aceito em universidade americana	X			
Mudança para curso da mesma área	X		X	
Mudança de curso para área correlata	X		X	
Greves		X	X	
Má remuneração na área		X	X	

Na Tabela 4.35 as células assinaladas com “X” correspondem ao item que foi citado como causa de evasão para aquele curso. Como mencionado anteriormente, maior parte dos itens acrescentados pelos alunos como causas de evasão são por razões institucionais/acadêmicas. Alguns se destacam na Tabela 4.35 por terem sido citados na maioria dos cursos (em mais de dois deles):

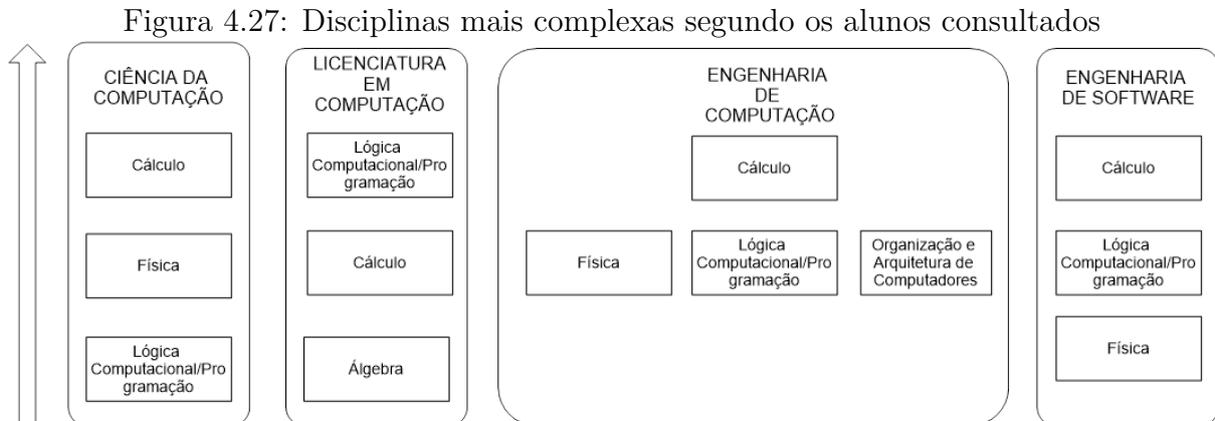
- Baixa qualidade docente
- Falta de base estudantil
- Currículo desatualizado/ementa inadequada
- Falta de apoio e orientação acadêmica
- Falta de adequação da academia com o mercado de trabalho
- Reprovações sucessivas

Estes itens exigem uma alerta quanto a abordagem docente em sala de aula, pois na percepção de muitos alunos ela tem sido inadequada. De acordo com muitos discentes o currículo dos cursos precisa ser atualizado em consonância com o mercado de trabalho. É necessário, portanto, que o professor avalie de que forma a sua abordagem em sala de aula pode se tornar mais atraente, procurando fazer sempre uma interseção entre teoria e cobranças no mercado de trabalho. Também é interessante incentivar as atividades de monitoria, visto que muitos alunos consideram que lhes faltam as bases que deveriam ter sido adquiridas no ensino médio e fundamental.

Os alunos foram questionados também sobre os motivos que os levaram a escolher um curso de graduação em Computação. As respostas podem ser acompanhadas no Apêndice B deste documento. Mais de 50% dos alunos em cada curso afirmaram ter sido motivados principalmente pela afinidade, curiosidade e interesse por computadores e pela área de Computação. É interessante notar que a falta de afinidade com o curso foi uma das questões mais pontuadas no formulário Likert, em todos os cursos, como causa de evasão. Isso é uma evidência de que muitos alunos não tem uma visão adequada do que é estudado no curso de Computação e confundem habilidades em algum software específico ou o gosto pela navegação na Internet com aptidão para o estudo aprofundado da Computação. Muitas vezes também o aluno já gosta de questões mais complexas como programação, mas acabam se decepcionando com muitas disciplinas vistas nos cursos superiores, por acharem o conteúdo muito maçante ou inadequado em relação ao que o mercado de trabalho espera de um profissional de Computação.

Os discentes foram consultados também quanto à disciplina que julgavam ter sido mais complexa ao longo do curso. Todas as disciplinas mencionadas pelos alunos podem ser

conferidas no Apêndice B. A Figura 4.27 apresenta as disciplinas mencionadas como mais complexas em cada curso:



Na Figura 4.27, as disciplinas na base foram menos citadas (dentre as 3 da figura) e as do topo as mais citadas. O curso de Engenharia de Computação apresenta 4 disciplinas pois 3 delas foram igualmente citadas pelos alunos. Percebe-se que disciplinas relacionadas a algoritmos/programação e disciplinas relacionadas a Cálculos estão entre as mais citadas, em todos os cursos. Conforme visto anteriormente, as dificuldades em disciplinas do curso ligadas a matemática foram consideradas por muitos alunos que responderam à pesquisa como uma das principais causas de evasão. No curso de Engenharia de Computação, as dificuldades em disciplinas do curso ligadas ao ensino de programação e algoritmos foi uma das principais causas de evasão segundo os alunos evadidos respondentes da pesquisa. Recordando o que foi visto no estudo de Análise de sobrevivência, a sobrevivência dos alunos que reprovaram alguma vez em Cálculo 1 ou em Algoritmos foi significativamente inferior à sobrevivência dos alunos que não tiveram nenhuma reprovação nessas disciplinas. Na aplicação do algoritmo Apriori, também foi visto que o êxito nessas disciplinas contribui para a permanência do aluno no curso.

Ao levantarmos as causas de evasão nos cursos superiores de Computação, vimos, portanto, que se trata de um problema com múltiplas causas associadas. Contudo algumas causas se destacaram, principalmente fatores institucionais/acadêmicos. É importante ressaltar que essa pesquisa foi conduzida em uma universidade pública (que é a UnB). Os motivos para evasão poderiam variar se a instituição adotada para a pesquisa tivesse sido uma instituição privada. Fatores sócio-econômicos, como dificuldades financeiras, poderiam ter exercido um impacto maior nesse caso. Nesta pesquisa, embora não se tenha contemplado todos os motivos relacionados à evasão dada a complexidade do problema, mostrou-se que existem diversos fatores inerentes à evasão que podem ser combatidos a

partir de um esforço institucional para assim mitigar o grave problema que é a evasão escolar.

Capítulo 5

Conclusões

A evasão nos cursos superiores de Computação se mostrou um problema complexo, que sofre a influência de diversas variáveis. Além dos fatores institucionais/acadêmicos, vocacionais, sócio-econômicos, pessoais, familiares e de saúde, a evasão se mostra diferente em função das formas de ingresso, mantenedoras (se pública ou privada), sexo do aluno e se a instituição oferta ou não pós-graduação em Computação. A Tabela 5.1 apresenta um resumo sobre aspectos da evasão vistos ao longo desse trabalho:

Tabela 5.1: Fatores relacionados à evasão nos cursos superiores de Computação

Aspecto	Descrição	Instituições	Dados Usados
Grande área de Ciências, Matemática e Computação	Juntamente com a grande área de Serviços, apresentou as maiores taxas de evasão (2010-2014)	Privadas e Públicas	INEP
Concorrência	Concorrências maiores estão associadas a menores taxas de evasão	Privadas e Públicas	INEP
Sexo Masculino	Maior sobrevivência à evasão	Públicas e Privadas	INEP
	Menor sobrevivência à evasão	UnB	UnB
Ingresso ENEM/SISU	Maior Sobrevivência à evasão (menor sobrevivência para o ingresso por Vestibular)	Privadas	INEP
	Menor sobrevivência à evasão (maior sobrevivência para o ingresso por PAS)	Públicas / UnB	INEP / UnB
Oferta de curso de pós-graduação em Computação	Maior sobrevivência à evasão	Públicas e Privadas	INEP/CAPES
Cotistas	Maior sobrevivência à evasão	Públicas	INEP
Não reprovação em Algoritmos e em Cálculo 1	Maior sobrevivência à evasão	UnB	UnB
Não reprovação em Cálculo 1	Diminui reprovação em Algoritmos	UnB	UnB
Evasão ocorre mais rapidamente	Quatro primeiros períodos	UnB	UnB

Na pesquisa feita na UnB para levantamento das causas de evasão, destacaram-se as

questões institucionais/acadêmicas. Os alunos evadidos indicaram como causas: baixa qualidade das aulas, currículos dos cursos em desacordo com o que esperavam, critérios de avaliação muito rígidos, dificuldades em disciplinas que requerem conhecimentos matemáticos e em disciplinas relacionadas à algoritmos/programação, etc. Essas opiniões devem levar a uma reflexão por parte das universidades, sobre os métodos de ensino que estão adotando. Sabe-se que muitos alunos chegam à universidade com uma base deficiente proveniente do ensino fundamental e médio. Contudo, as universidades podem buscar mecanismos para mitigar essas deficiências e tornar a aprendizagem nos cursos de Computação algo mais prazeroso. Conforme visto ao longo dessa dissertação a reprovação em disciplinas de Algoritmos e de Cálculo 1, vistas nos períodos iniciais do curso, contribuem para a evasão discente e essa evasão ocorre mais rapidamente nos primeiros períodos, com picos maiores no terceiro e quarto períodos. Isso é reforçado no Apêndice C desse trabalho, onde é exibido um artigo apresentado na 4ª Conferência Ibero Americana em Computação Aplicada - Lisboa, Portugal, sobre a influência da reprovação nas disciplinas de algoritmos e de Cálculo 1 em relação à evasão.

Uma causa citada como motivo para a evasão é a falta de base em conceitos do ensino médio. Uma medida que pode dar bons resultados é a adoção de aulas de reforço no início do curso para alunos com dificuldades em disciplinas como Cálculo 1 e Algoritmos. A aprendizagem precisa ser mais didática e divertida, motivando o aluno a ele mesmo buscar a aprendizagem sem se limitar à sala de aula. Outra medida que pode ajudar é incentivar os alunos a formarem grupos de estudos com reuniões periódicas, contando inclusive, com a participação de professores e monitores nessas reuniões. Estimular os trabalhos de monitoria com concessão de benefícios que possam ir além do recebimento de certificados, pode favorecer também a manutenção dos alunos com dificuldades na universidade.

No que tange ao ensino de programação, é interessante que o discente, no início do curso, tenha acesso a ferramentas simples, sem configurações avançadas que possam assustá-lo. Um exemplo seria o Portugol Studio. O Portugol “é uma notação utilizada para escrever programas em uma mescla de português e símbolos comuns em linguagens de programação de alto nível (operadores relacionais, operadores aritméticos, entre outros)” [71].

A escolha de um curso superior não costuma ser algo trivial. Além de ser uma decisão que definirá o futuro profissional do graduando, o aluno pode acabar ingressando em um curso sem conhecer o seu foco ou pode acabar sendo influenciado pela família ou por amigos. A partir da pesquisa conduzida na UnB percebeu-se que, embora muitos alunos tenham declarado a escolha pelo curso de Computação por afinidade com computadores, muitos apontam como principais causas para desistência razões vocacionais. Isso mostra

que muitos discentes tem uma percepção errada do que é estudado nos cursos superiores de Computação e também da utilidade de determinadas disciplinas em sua formação. Falta muitas vezes conhecimento do aluno em relação ao curso escolhido. Nesse sentido, a UFPR desenvolveu um projeto intitulado “Repensando a Escolha Profissional” [40], desenvolvido por estagiários do curso de psicologia e supervisionados por uma psicóloga. Entre os participantes deste projeto tem-se observado que cerca de metade deles decide permanecer no curso e outra metade decide migrar de curso. Um trabalho desse tipo pode ajudar a conduzir o aluno para um curso com o qual ele tenha uma grande afinidade, ou pode ainda ajudar a elucidar, nos casos dos cursos de Computação, que todas as disciplinas estudadas, como Cálculo e Física, tem relevância na sua formação profissional. É importante que essa orientação vocacional seja conduzida ainda nos primeiros períodos, visto que esse estudo revelou picos maiores de saída nos períodos iniciais nos cursos de Computação.

Outros fatores associados à evasão também são preocupantes: dificuldades dos professores em repassar de maneira compreensível o conteúdo, falta de informações sobre o que é abordado no curso e critérios de avaliação dos discentes inadequados ou muito rígidos. Esses fatores institucionais estão entre os mais citados pelos alunos como causas de evasão em todos os cursos de Computação pesquisados na UnB. Isso revela um distanciamento no relacionamento professor-aluno. O professor deve atentar para o fato de que a sua função como docente não é meramente conteudista. O docente deve ser motivador, um facilitador do raciocínio crítico, deve levar o discente a trilhar o seu próprio caminho. Além de avaliar o aluno, deve fazer parte das políticas institucionais que os docentes façam uma autoavaliação e que eles sejam avaliados pelos discentes. Os professores precisam ter um *feedback* constante em relação ao seu trabalho. O trecho abaixo descreve bem a postura que os professores devem assumir para promover uma formação integral do aluno, como profissionais que não apenas recebem mas buscam autonomamente o conhecimento:

[...]O professor precisa estar imbuído da necessidade de uma imersão não somente na verticalidade epistemológica do saber especializado e técnico, mas deve, na horizontalidade da abordagem didática, sair do centro propulsor e responsável do processo ensino aprendizagem para se tornar um facilitador, um animador, um estudante pronto a romper, a transgredir e a lançar desafios [57].

É importante ressaltar que o professor não deve recorrer à pedagogia do terror para tentar pressionar os alunos a estudarem a partir do medo. O aluno deve se inspirar no professor como alguém que, além de capacitado e com bom domínio do conteúdo, se interessa pelo seu progresso, tira dúvidas e o leva a ter um interesse genuíno pela disciplina. Uma outra política educacional interessante para ser empregada nas universidades seria ofertar cursos de didática aos professores, especialmente para aqueles docentes que não possuem graduação em licenciatura.

Algo que serve de alerta para os gestores acadêmicos, como os coordenadores de curso, é o item Currículo desatualizado/ementa inadequada ter sido citado por diversos alunos como causa de evasão. Disciplinas matemáticas como Cálculo 1, podem se tornar desinteressantes para os alunos ao terem uma abordagem muito teórica e não se encontrar no início do curso relação clara entre os cálculos e as demandas profissionais de um profissional da área de Computação. É inegável a importância do conhecimento teórico matemático para que o aluno saiba quando e como empregar os cálculos. O uso de laboratórios de cálculos com softwares específicos podem auxiliar na construção da ponte teoria-prática. Um estudo conduzido na PUCRS [57] sobre o uso do computador no ensino da matemática revela um bom potencial dos computadores como ferramenta para ensino de cálculo:

[...] à medida que os alunos executam atividades, aumenta o nível de conversa e, às vezes, surge um grande interesse na execução da atividade; são momentos que criam no laboratório um ambiente de trabalho. A permanência no laboratório é ótima, com poucas saídas ou deslocamentos de alunos. A disciplina é excelente, não há brincadeiras ou perturbações da ordem.

Se essas discussões em torno das atividades de cálculo que são promovidas nos laboratórios forem associadas à explicação dos conteúdos e à uma maior interação professor aluno, isso pode reverter a visão dos discentes de que os professores tem dificuldades em repassar de maneira compreensível o conteúdo. O professor que não esteja familiarizado com ferramentas de software para ensino de Cálculo deve procurar se qualificar nesse sentido.

A troca de informações entre professores, discentes e núcleos de assistência estudantil também deve fazer parte da rotina das universidades. Por isso é tão relevante o estreitamento dos laços entre professores e alunos. Se o professor comunica precocemente um problema que acomete algum aluno, seja de natureza pessoal, familiar, de saúde ou sócio-econômica, os núcleos de assistência estudantil podem agir precocemente no sentido de tornar a evasão reversível. O incentivo à prática esportiva, seja como lazer ou para competições oficiais, além do incentivo à cultura e às artes, é também muito relevante, pois o aluno passa bastante tempo na universidade e necessita de momentos de entretenimento.

Os alunos devem também ser levados a participarem como colaboradores em projetos científicos ou de extensão, para que possam compreender a relação existente entre ensino, pesquisa, análise e produção de conhecimentos que são indissociáveis. Envolver nessa pesquisa alunos com diferentes perfis pode auxiliar na criação dos laços entre os discentes na universidade, além de propiciar ao aluno enxergar uma maior utilidade em relação às disciplinas estudadas, já que um dos motivos citados para a evasão é a falta de adequação da academia com o mercado de trabalho. Trabalhar em equipe também é um requisito essencial no mercado de trabalho. O discente aprenderá a ter responsabilidade, ser líder, independente. Isso tornará todo o processo ensino-aprendizagem menos maçante e mais significativo. Um aluno motivado terá menos motivos para evadir. É importante que

o professor acompanhe e oriente com frequencia os alunos envolvidos nos projetos, sane dúvidas, lance provocações sadias para serem discutidas dentro do grupo, oriente sobre as principais tendências, tecnologias e ferramentas, mas enfatize que cabe a eles as decisões, distribuições de papeis e desenvolvimento dos projetos.

Outro fator que impacta a evasão nos cursos de Computação, conforme visto ao longo desse estudo, é a forma de ingresso. O vestibular em 3 etapas implantado na UFPE [42], no Centro de Ciências Exatas e da Natureza, mostra-se uma política interessante na diminuição da evasão:

[...]A primeira igual a toda a universidade. A segunda etapa consta apenas da prova de português. Os 120 primeiros alunos passam para terceira fase, frequentando por seis meses as disciplinas de física, química e matemática na nossa UFPE. Findo este período, ingressarão nos cursos de física, química e matemática os 40 primeiros colocados, em condições de acompanhar as disciplinas curriculares, pelo fato de ter submetido previamente a esta capacitação. Outras Instituições de Ensino Superior (IES) que implantaram este sistema conseguiram reduzir significativamente a evasão e a retenção dos seus estudantes.

Com essa alteração na forma de ingresso pelo vestibular é possível fazer o nivelamento dos alunos já na seleção, sem necessidade de investir um tempo maior suprindo deficiências de base dos alunos depois que eles ingressam na universidade.

A abordagem usada nesse trabalho usando Análise de Sobrevida e mineração de regras de associação com o algoritmo Apriori mostrou um bom potencial para análise de evasão. Com a Análise de Sobrevida é possível analisar a evasão levando-se em consideração o tempo que o aluno leva para evadir e é possível analisar como a evasão se manifesta diante de diferentes fatores. O algoritmo Apriori, por sua vez, permite verificar quais fatores se associam de forma mais contundente à evasão. As técnicas de mineração de dados e a Análise de Sobrevida podem fornecer importantes subsídios para orientar as políticas educacionais em uma instituição de ensino superior. Seu uso não precisam se limitar à análise de evasão nos cursos de Computação, mas pode ser expandido para quaisquer cursos dentro da instituição.

Como sugestões para trabalhos futuros, seria interessante verificar em uma instituição de ensino superior privada as causas de evasão nos cursos de Computação, consultando-se os alunos evadidos e comparando os resultados com os obtidos no estudo de caso feito na UnB, visto que, de acordo com a mantenedora institucional (se pública ou privada), as causas de evasão podem variar. Também seria interessante estabelecer uma correlação mais ampla entre desempenho e evasão, já que nesse estudo limitou-se a verificar a reprovação ou não na disciplina de Cálculo 1 e em disciplinas de Algoritmos, mas não se analisou o desempenho do aluno.

Esse trabalho não esgota o tema da evasão nos cursos de Computação. No entanto, várias causas que estão por trás da evasão discente foram aqui elencadas. Espera-se,

a partir desse trabalho, não apenas informar mas também conscientizar professores e dirigentes em instituições de ensino superior, de que é possível coibir a evasão em muitas circunstâncias. Trabalhar para a permanência do aluno é um compromisso que toda a instituição de ensino deve assumir com a formação profissional e cidadã dos seus alunos.

Referências

- [1] Ghadeer S. Abu-Oda e Alaa M. El-Halees. Data mining in higher education: University student dropout case study. *International Journal of Data Mining & Knowledge Management Process*, 5(1):15–27, 2015. 22
- [2] Ana Amélia Chaves Teixeira Adachi. Evasão e evadidos nos cursos de graduação da Universidade Federal de Minas Gerais. Dissertação (Mestrado), Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais - Brasil. 9, 13, 15
- [3] Onília Cristina de Souza de Almeida. Evasão em cursos à distância: Análise dos motivos de desistência. In *14 Congresso Internacional ABED de Educação a Distância. Santos – São Paulo, Brasil*, 2008. 7
- [4] ANDIFES. Grupo de trabalho sobre evasão e retenção da Associação Nacional dos Dirigentes das Instituições Nacionais de Ensino Superior (Andifes). Disponível em: < <http://site.ufsm.br/noticias/exibir/8800>>. Acesso em: 18/10/2015. 17
- [5] Wagner Bandeira Andriola, Cristiany Gomes Andriola, e Cristiane Pascoall Moura. Opiniões de docentes e de coordenadores acerca do fenômeno da evasão discente dos cursos de graduação da Universidade Federal do Ceará (UFC). *Revista Ensaio: aval. pol. públ. Educ.*, 14(52):365–382, 2006. 15
- [6] Cristiane Aparecida dos Santos Baggi e Doraci Alves Lopes. Evasão e avaliação institucional no ensino superior: uma discussão bibliográfica. *Avaliação*, 16(2):355–374, 2011. 14
- [7] Ryan Shaun Joazeiro de Baker, Seiji Isotani, e Adriana Maria Joazeiro Baker de Carvalho. Mineração de dados educacionais; oportunidades para o Brasil. *Revista Brasileira de Informática na Educação*, 19(2):3–11, 2011. 36
- [8] Gerson Henrique Baptista, Adriano Herbert H. K. Gonçalves de Andrade, e Suely Ruiz Giolo. Fatores associados à duração do aleitamento materno em crianças de famílias de baixa renda da região sul da cidade de Curitiba, Paraná, Brasil. *Cad. saúde pública*, 25(3):596–604, 2009. 33
- [9] Thiago Schumacher Barcelos e Ismar Frango Silveira. Pensamento computacional e educação matemática: Relações para o ensino de computação na educação básica. In *XX Workshop sobre Educação em Computação, Curitiba. Anais do XXXII CSBC*, 2012. 2, 23

- [10] M. F. Barroso e E. B. M. Falcão. Evasão universitária: O caso do instituto de física da UFRJ. In *IX Encontro Nacional de Pesquisa em Ensino de Física*, 2004. 15, 25, 92
- [11] Francisco Botelho, Carlos Silva, e Francisco Cruz. Epidemiologia explicada – análise de sobrevivência. *Acta Urológica*, 26(4):33–38, 2009. 30, 31
- [12] Portal Brasil. Evasão de bolsistas do Prouni é inferior à média das universidades privadas. Disponível em: <<http://www.brasil.gov.br/educacao/2011/08/evasao-de-bolsistas-do-prouni-e-menor-do-que-a-de-universitarios-do-sistema-privado>>. Acesso em: 30/10/2016, 2011. 18
- [13] Paula Yamada Burkle. Um método de pós-processamento de regras de associação com base nas relações de dependência entre os atributos. Dissertação (Mestrado), Universidade Federal Fluminense, Niteroi, Rio de Janeiro - Brasil. 84, 85
- [14] Maria Teres Bustamante-Teixeira, Eduardo Faerstein, e Latorre do Rosário. Técnicas de análise de sobrevida - survival analysis techniques. *Cad. Saúde Pública*, 18(3):579–594, 2002. 31, 33
- [15] Marisa Veiga Capela e Jorge M. V. Capela. Elaboração de gráficos Box-plot em planilhas de cálculo. In *CONGRESSO DE MATEMÁTICA APLICADA E COMPUTACIONAL DA REGIÃO SUDESTE–CNMAC Sudeste. Vol. 1*, 2011. 26, 28
- [16] Claudete Batista Cardoso. Efeitos da política de cotas na Universidade de Brasília: uma análise do rendimento e da evasão. Dissertação (Mestrado), Universidade de Brasília - Brasil, Março 2008. 19
- [17] Pete Chapman, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer, e Rüdiger Wirth. CRISP-DM 1.0: Step-by-step data mining guide. Disponível em: <<https://www.the-modeling-agency.com/crisp-dm.pdf>>. Acesso em: 06/11/2016, 2000. 35
- [18] Evandro Costa, Ryan Shaun Joazeiro de Baker, Lucas Amorim, Jonathas Magalhães, e Tarsis Marinho. Mineração de dados educacionais: Conceitos, técnicas, ferramentas e aplicações. *Jornada de Atualização em Informática na Educação*, 1(1):1–29, 2013. 34
- [19] Armando Jorge Ribeiro da Cruz. Data mining via redes neuronais artificiais e máquinas de vectores de suporte. Dissertação (Mestrado), Universidade do Minho, Braga, Portugal, 2007. 35
- [20] Toby S. Cubitt, Jens Eisert, e Michael M. Wolf. Extracting dynamical equations from experimental data is np hard. *Physical review letters*, 108(12):120503, 2012. 14
- [21] Jacqueline Veneroso Alves Da Cunha, Eduardo Mendes Nascimento, e Samuel de Oliveira Durso. Razões e influências para a evasão universitária: um estudo com estudantes ingressantes nos cursos de ciências contábeis de instituições públicas federais da região sudeste. In *XIV Congresso de Controladoria e Contabilidade - Novas Perspectivas na Pesquisa Contábil. Universidade de São Paulo (USP)*, 2014. 3

- [22] Ministério da Educação (MEC). Programa de apoio a planos de reestruturação e expansão das universidades federais (Reuni). Disponível em: <<http://reuni.mec.gov.br/o-que-e-o-reuni>>. Acesso em: 15/03/2016, 2010. 16
- [23] Ministério da Educação (MEC). Perguntas frequentes. Disponível em: <<http://portal.mec.gov.br/cotas/perguntas-frequentes.html>>. Acesso em: 30/10/2016, 2012. 19
- [24] Ministério da Educação (MEC). O que é o PROUNI. Disponível em: <<http://siteprouni.mec.gov.br/>>. Acesso em: 30/10/2016, 2016. 18
- [25] BRASIL. SESu/MEC. Ministério da Educação. Secretaria de Educação Superior. Comissão especial de estudos sobre a evasão nas universidades públicas brasileiras. Disponível em: <<http://www.dominiopublico.gov.br/download/texto/me002240.pdf>>. Acesso em: 29/10/2016, Outubro 1997. 8, 9, 19
- [26] Henrique Rego Monteiro da Hora, Gina Torres Rego Monteiro, e José Arica. Confiabilidade em questionários para qualidade: um estudo com o coeficiente Alfa de Cronbach. *Produto & Produção*, 11(2):85–103, 2010. 92
- [27] Camila da Silva Oliveira e Luciano Nadler Lins. Identificação das causas da evasão e retenção prolongada de estudantes do ensino superior a partir de dados censurados. In *XLIII Simpósio Brasileiro de Pesquisa Operacional. Ubatuba - São Paulo*, 2011. 30, 33
- [28] Christine Dancey e John Reidy. *Estatística Sem Matemática para Psicologia: Usando SPSS para Windows*. Artmed, Porto Alegre, 3 edition, 2006. 56
- [29] Sindicato das Entidades Mantenedoras de Estabelecimentos de Ensino Superior no Estado de São Paulo (Brasil). Índice de evasão de alunos é maior na área de tecnologia da informação. Disponível em: <<http://g1.globo.com/sp/sao-carlos-regiao/noticia/2012/09/indice-de-evasao-de-alunos-e-maior-na-area-de-tecnologia-da-informacao.html>>. Acesso em: 04/08/2015, Setembro 2012. 1
- [30] Sandra de Amo. Técnicas de mineração de dados. In *XXIV Congresso da Sociedade Brasileira de Computação, vol. 1–1, Jul/Ago*, 2004. 38, 39, 40
- [31] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Obtenção de dados sobre evasão nos cursos superiores de graduação em computação. Comunicação pessoal Via Lei de Acesso a Informação. Resposta obtida em 06/07/2015. 13
- [32] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Informe estatístico do MEC revela melhoria do rendimento escolar. Disponível em: <http://portal.inep.gov.br/rss_censo-escolar/-/asset_publisher/oV0H/content/id/19145>. Acesso em: 18/02/2016, 1998. 7

- [33] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Classificação internacional EUROSTAT/UNESCO/OCDE. Disponível em: <http://download.inep.gov.br/download/superior/2009/Tabela_OCDE_2009.pdf>. Acesso em: 07/11/2015, 2000. 41, 51
- [34] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Resumo técnico do censo da educação superior de 2009. Disponível em: <http://download.inep.gov.br/download/superior/censo/2009/resumo_tecnico_2009.pdf>. Acesso em: 15/02/2016, 2010. 12, 43
- [35] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Resumo Técnico Censo da Educação Superior de 2010. Disponível em: <http://download.inep.gov.br/download/superior/censo/2010/resumo_tecnico_censo_educacao_superior_2010.pdf>. Acesso em: 02/01/2016, 2012. 44
- [36] Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP). Censo da Educação Superior 2014. Glossário. Disponível em: <http://download.inep.gov.br/educacao_superior/censo_superior/questionarios_e_manuais/2014/glossario_curso_2014.pdf>. Acesso em: 07/11/2016, 2015. 46
- [37] Pró-Reitoria de Graduação da Universidade Federal de São Paulo (UNIFESP). Dados Comparados Sobre Evasão Nos Cursos de Graduação em Universidades Federais – Ano Referência 2014. Disponível em: <<http://www.unifesp.br/reitoria/prograd/pro-reitoria-de-graduacao/informacoes-institucionais/graduacao-em-numeros?download=600:evasao-2014>>. Acesso em: 02/10/2016. 10
- [38] Gilmar Mercês de Jesus e Éric Fernando Almeida de Jesus. Nível de atividade física e barreiras percebidas para a prática de atividades físicas entre policiais militares. *Revista Brasileira de Ciências do Esporte*, 34(2):433–448, 2011. 88
- [39] Maria Madalena Dias. Parâmetros na escolha de técnicas e ferramentas de mineração de dados. *Acta Scientiarum. Technology*, 24(6):1715–1725, 2002. 34
- [40] Universidade Federal do Paraná. Projeto Repensando a Escolha Profissional orienta estudante indecisos. Disponível em: <<http://www.ufpr.br/portalufpr/blog/noticias/projeto-repensando-a-escolha-profissional-orienta-estudantes-indecisos/>>. Acesso em: 18/11/2016. 108
- [41] Universidade Tecnológica Federal do Paraná (UTFPR). Sisu amplia mobilidade, mas evasão cresce. Disponível em: <<http://www.utfpr.edu.br/estrutura-universitaria/diretorias-de-gestao/dircom/noticias/clipping-1/16-02-2014/sisu-amplia-mobilidade-mas-evasio-cresce-tribuna-do-norte>>. Acesso em: 15/03/2016, 2014. 18
- [42] Gilson Edmar. A evasão e a retenção. Disponível em: <https://www.ufpe.br/agencia/index.php?option=com_content&view=article&id=36622:a-evasio-e-a-retencao&catid=7&Itemid=80>. Acesso em: 18/11/2016. 110

- [43] Marillia Gabriella Duarte Fialho. A evasão escolar e a gestão universitária: O caso da Universidade Federal da Paraíba. Dissertação (Mestrado), Universidade Federal da Paraíba - Brasil, 2014. 7, 8, 10
- [44] Silvia Figini. *Data Mining for Lifetime Value Estimation. Encyclopedia of Data Warehousing and Mining.*, pages 431–437. Information Science Reference, Hershey - New York, second edition, 2009. 29
- [45] Natalícia Pacheco de Lacerda Gaioso. Evasão discente na educação superior: a perspectiva dos dirigentes e dos alunos. Dissertação (Mestrado), Universidade Católica de Brasília - Brasília, Brasil, 2005. 6, 13
- [46] Renato de Sousa Porto Gilioli. Evasão em instituições federais de ensino superior no Brasil: Expansão da rede, Sisu e desafios. Disponível em: <http://www2.camara.leg.br/documentos-e-pesquisa/publicacoes/estnottec/areas-da-conle/tema11/2016_7371_evasao-em-instituicoes-de-ensino-superior_renato-gilioli>. Acesso em: 15/10/2016, 2016. 17, 18
- [47] Lucia Maria Martins Giraffa e Michael da Costa Mora. Mineração de dados educacionais; oportunidades para o Brasil. In *Tercera Conferencia LatinoAmericana Sobre El Abandono En La Educación Superior, Rio Grande do Sul - Brasil*, 2013. 15
- [48] Maria José Gomes, Mariana Monteiro, Anderson Medeiros Damasceno, Tereza Jacy Silva Almeida, e Raquel Baroni de Carvalho. Evasão acadêmica no ensino superior: Estudo na Área da saúde. *Revista Brasileira de Pesquisa em Saúde/Brazilian Journal of Health Research*, 12(1):6–13, 2010. 20
- [49] Stefan Gruner. On the future of computer science in south africa: A survey amongst students at university. In *44th Annual Southern African Computer Lecturers Association 2015 (SACLA 2015)*. Johan-nesburg - South Africa, 2006. 24
- [50] Ornélio Hinterholz Jr. Tepequém: uma nova ferramenta para o ensino de algoritmos nos cursos superiores em computação. In *XVII-Anais do Workshop sobre Educação em Informática*, 2009. 20, 21
- [51] Wilhelmiina Hämäläinen e Matti Nykänen. Efficient discovery of statistically significant association rules. In *2008 Eighth IEEE International Conference on Data Mining*, pages 203–212, 2008. 85, 86
- [52] José Rubens Lima Jardimino, Derly Jardim do Amaral, e Delmário Ferreira Lima. A interação professor-aluno em sala de aula no ensino superior: o curso de administração de empresas. *Revista Diálogo Educ.*, Curitiba, 10(29):101–119, 2010. 93
- [53] Alvaro Alberto Ferreira Mendes Junior. A evasão no ensino superior brasileiro. *Revista Ensaio: Avaliação e Políticas Públicas em Educação*, 22(82):31–56, 2014. 19
- [54] José Roberto Kassai, Amanda Yamashiro Campos de Oliveira, Carlos Eduardo Alves da Silva, e Luiz Nelson Guedes de Carvalho. Reflexões sobre o nível de evasão e o custo anual per capita das unidades de ensino da USP com base no método inquired balance sheet. In *XXXIV Encontro da ANPAD. Rio de Janeiro (RJ)*., 2010. 3

- [55] Päivi Kinnunen e Lauri Malmi. Why students drop out cs1 course? In *Proceedings of the second international work-shop on Computing education research*. ACM, p. 97-108, 2006. 1
- [56] Marcelo Kisahleitner. Análise de técnicas de data mining na aquisição de clientes de cartão de crédito não correntistas. Dissertação (Mestrado), Fundação Getulio Vargas Escola de Administração de Empresas de São Paulo. São Paulo – SP., 2009. 29
- [57] João Bosco Laudares e Jonas Lachini. O uso do computador no ensino de matemática na graduação. In *23ª Reunião Anual da Associação Nacional de Pós-Graduação e Pesquisa em Educação*, pages 32–43, 2000. 108, 109
- [58] João Batista Gomes de Lima. Estudo da evasão escolar do ponto de vista econômico: o caso dos centros universitários camilianos do brasil. Dissertação (Mestrado), Fundação Instituto Capixaba de Pesquisa em Contabilidade, Economia e Finanças. Espírito Santo, Vitória, 2008. 16
- [59] Paulo Lima Junior, Nilce Santos de Melo, e Mauro Rabelo. Decanato de Ensino de Graduação. Análise de Trajetória dos Alunos da UnB. Disponível em: <<http://unb2.unb.br/administracao/decanatos/deg/trajetoria/trajetoria.htm>>. Acesso em: 29/10/2016. 10
- [60] Paulo Roberto Menezes Lima Junior. *Evasão do ensino superior de Física segundo a tradição disposicionalista em sociologia da educação*. Tese (Doutorado), Instituto de Física, Universidade Federal do Rio Grande do Sul, Rio Grande do Sul, 2013. 3
- [61] Paulo Lima Júnior, Fernando Lang da Silveira, e Fernanda Ostermann. Análise de sobrevivência aplicada ao estudo do fluxo escolar nos cursos de graduação em física: um exemplo de uma universidade brasileira. *Revista brasileira de ensino de física*, 34(1):10 p, 2012. 29, 30, 31, 32
- [62] Miriam Liston, Denise Frawley, e Vivienne Patterson. A study of progression in Irish higher education. Technical report, Higher Education Authority, Dublin, 2016. 1
- [63] Raquel Cardoso Machado e Eduardo Luiz Dias Cavalcanti. Desempenho acadêmico e sucesso/insucesso escolar dos estudantes do curso de química: relações possíveis. In *XV Encontro Nacional de Ensino de Química (XV ENEQ) – Brasília, DF, Brasil*, 2010. 14
- [64] Laci Mary Barbosa Manhães, Sérgio Manuel Serra Da Cruz, e Raimundo J. Macário Costa. Previsão de estudantes com risco de evasão utilizando técnicas de mineração de dados. *Anais do XXII SBIE - XVII WIE*, 1(1):150–159, 2011. 36, 37
- [65] João Maroco e Teresa Garcia-Marques. Qual a fiabilidade do Alfa de Cronbach? questões antigas e soluções modernas? *Laboratório de Psicologia*, 4(1):65–90, 2013. 92
- [66] Elizabeth Mercuri e Camila Alves Fior. Análise dos fatores preditivos da evasão em uma universidade confessional. In *Segunda Conferencia Latinoamericana sobre el abandono de la educacion superior (Eds). Libro de actas.*, pages 178–189, 2012. 15

- [67] Oliver Mooney, Vivienne Patterson, Muiris O' Connor, e Abigail Chantler. A study of progression in Irish higher education. Technical report, Higher Education Authority, Dublin, 2010. 1
- [68] José Uibson Pereira Moraes, Renato Santos Araujo, e Deise Miranda Vianna. Estudo sobre o curso de Licenciatura em Física: o que dizem os dados estatísticos. In *X Escola de Verão em Educação em Química, UFS, São Cristóvão, SE*, 2014. 10
- [69] Wille Muriel. Evasão escolar nas instituições de ensino superior. Disponível em: <<http://www2.cartaconsulta.com.br/espacodocoordenador/evasao-escolar-nas-instituicoes-de-ensino-superior>>. Acesso em: 02/04/2015. 15
- [70] Grupo Assessor nomeado pela Portaria nº 552 SESu/MEC de 25 de junho de 2007. Diretrizes Gerais do Programa de Apoio a Planos de Reestruturação e Expansão das Universidades Federais – REUNI. Disponível em: <<http://portal.mec.gov.br/sesu/arquivos/pdf/diretrizesreuni.pdf>>. Acesso em: 18/10/2015. 17
- [71] Luis Fernando Noschang, Elieser A. Fillipi Pelz, e André L. A. Raabe. Portugol Studio: Uma IDE para iniciantes em programação. In *Workshop de Educação em Computação, Anais do Congresso Anual da Sociedade Brasileira de Computação*, 2014. 107
- [72] Daltro José Nunes. Estatísticas da educação superior: Área da computação. Technical report, Sociedade Brasileira de Computação, 2008. 61
- [73] Ângela Tavares Paes. Itens essenciais em bioestatística. *Arquivos brasileiros de cardiologia*, 71(4):1998, 575-580. 43
- [74] Leda Queiróz Paula, Dilermando Piva Júnior, e Ricardo Luis Freitas. A leitura e a abstração do problema no processo de formação do raciocínio lógico-abstrato em alunos de computação. *Reverte-Revista de Estudos e Reflexões Tecnológicas da Faculdade de Indaiatuba*, n.7, 2009. 22
- [75] Dilermando Piva Jr e Ricardo L. Freitas. Estratégias para melhorar os processos de abstração na disciplina de algoritmos. In *XXXI Congresso da Sociedade Brasileira de Computação, Vol. 1. No. 1*, 2011. 22
- [76] Juliane Prediger, Luana Berwanger, e Marlete Finke Mors. Relação entre aluno e matemática: Reflexões sobre o desinteresse dos estudantes pela aprendizagem desta disciplina. *Destques Acadêmicos*, 1(4):151–159, 2013. 14
- [77] Soraia Silva Prietch e Tatiana Annoni Pazeto. Estudo sobre a evasão em um curso de licenciatura em informática e considerações para melhorias. In *Workshop de Educação em Informática Bahia-Alagoas-Sergipe (WEIBASE), Maceió/AL*, 2010. 21, 22, 23, 24
- [78] Wesley Romão. *Descoberta de Conhecimento Relevante em Banco de Dados sobre Ciência e Tecnologia*. Tese (Doutorado), Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina, 2002. 37, 39

- [79] Wesley Romão, Carlos A. P. Niederauer, Alejandro Martins, Aran Tcholakian, Roberto C. S. Pacheco, e Ricardo M. Barcia. Extração de regras de associação em C&T: O algoritmo Apriori. In *XIX Encontro Nacional em Engenharia de Produção*, 1999. 34, 37, 39, 40
- [80] Eduarde Rosa. A evasão no ensino superior : (um estudo sobre a Universidade Federal de Goiás). Dissertação (Mestrado), Universidade Federal de Goiás - Brasil, 1977. 19
- [81] Fabrícia Damando Santos, Magda Bercht, Leandro Krug Wives, e Silvio César Cazella. Análise de evidências do estado de ânimo desanimado de alunos de um AVEA: uma proposta a partir da aplicação de regras de associação. *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, 4(1):1054–1063, 2015. 37
- [82] Átila Pires dos Santos, Sandra Isaelle Figueiredo Dos Santos, e Vandor Roberto Vilarde Rissoli. A predição da evasão de estudantes de graduação como recurso de apoio fornecido por um assistente inteligente. In *10º Simpósio de Excelência em Gestão e Tecnologia (SEGeT)*, vol. 10, 2013. 13
- [83] Antônio Simões Silva. Comparação entre modelos de taxas de evasão. In *Cobenge 2014 - XLII Congresso Brasileiro de Educação em Engenharia (COBENGE)*, pages 1–5, 2014. 9, 10
- [84] Leandro A. Silva e Luciano Silva. Fundamentos de mineração de dados educacionais. *Anais dos Workshops do Congresso Brasileiro de Informática na Educação*, 3(1):568–581, 2004. 34
- [85] Roberto Leal Lobo e Silva Filho e Maria Beatriz de Carvalho Melo Lobo. Como a mudança na metodologia do INEP altera o cálculo da evasão. Disponível em <http://www.institutolobo.org.br/imagens/pdf/artigos/art_079.pdf>. Acesso em: 02/09/2016. 11
- [86] Roberto Leal Lobo e Silva Filho e Maria Beatriz de Carvalho Melo Lobo. Esclarecimentos metodológicos sobre os cálculos de evasão. Disponível em <http://www.institutolobo.org.br/imagens/pdf/artigos/art_078.pdf>. Acesso em: 02/09/2016. 11
- [87] Roberto Leal Lobo e Silva Filho e Maria Beatriz de Carvalho Melo Lobo. Instituto lobo para o desenvolvimento da educação, da ciência e da tecnologia. Disponível em: <<http://institutolobo.org.br/paginas/institutolobo.php?v=1>>. Acesso em: 29/10/2016. 10
- [88] Roberto Leal Lobo e Silva Filho, Paulo Roberto Montejunas, Oscar Hipólito, e Maria Beatriz de Carvalho Melo Lobo. A evasão no ensino superior brasileiro. *Cadernos de Pesquisa*, 37(132):641–659, 2007. 9, 11, 12, 19, 52, 122
- [89] Site Nacional de Empregos (SINE). Média salarial. Disponível em: <<http://www.sine.com.br/media-salarial>>. Acesso em: 18/11/2016. 23

- [90] Clair Teresinha de Souza, Caroline da Silva Petró, e Rosana Maria Gessinger. Um estudo sobre evasão no ensino superior do Brasil nos últimos dez anos: As possíveis causas e fatores que influenciam no abandono. prevendo o risco do abandono. In *CONFERÊNCIA LATINOAMERICANA SOBRE EL ABANDONO EM LA EDUCACION SUPERIOR-CLABES*, volume 2, 2012. 14
- [91] Joselaine Valaski e Emerson Cabrera Paraiso. Limitações da utilização do Alice no ensino de programação para alunos de graduação. In *Anais do Simpósio Brasileiro de Informática na Educação, Vol. 23, No. 1*, 2012. 21
- [92] Michalis Xenos, Christos Pierrakeas, e Panagiotis Pintelas. A survey on student dropout rates and dropout causes concerning the students in the course of informatics of the hellenic open university. *Computers & Education*, 39(4):361–377, 2002. 25

Apêndice A

Taxas de Evasões Anuais

Tabela A.1: Evasão anual e concorrência nas grandes áreas do conhecimento

Grande Área	Ocorrência	Instituições Privadas					Instituições Públicas				
		Ano					Ano				
		2010	2011	2012	2013	2014	2010	2011	2012	2013	2014
Educação	Concorrência	0,55	0,68	0,80	0,79	0,70	4,09	5,88	6,74	6,93	6,61
	Evasão (%)	16,06	17,91	17,72	16,32	17,89	13,83	13,19	14,82	15,16	13,98
Humanidade e Artes	Concorrência	0,92	1,18	1,16	1,09	1,00	6,13	7,29	9,97	8,87	7,95
	Evasão (%)	16,07	18,00	16,76	17,40	18,29	14,64	13,46	12,47	14,91	12,47
Ciências Sociais, Negócios e Direito	Concorrência	0,94	1,20	1,22	1,16	1,06	7,29	10,61	13,93	13,61	13,87
	Evasão (%)	15,60	17,44	17,76	16,76	17,55	10,43	10,09	10,35	10,86	10,59
Ciências, Matemática e Computação	Concorrência	0,74	1,09	1,07	1,00	0,87	6,37	8,95	12,39	9,78	9,58
	Evasão (%)	18,72	20,22	19,65	20,22	20,53	14,97	14,40	15,46	15,28	16,01
Engenharia, Produção e Construção	Concorrência	1,33	1,66	1,85	1,69	1,35	8,98	13,42	16,72	16,32	15,23
	Evasão (%)	16,77	17,26	16,20	16,66	15,22	11,63	10,76	11,20	10,24	10,87
Agricultura e Veterinária	Concorrência	1,11	1,40	1,69	1,68	1,05	7,58	10,47	12,94	12,02	12,41
	Evasão (%)	12,71	13,46	13,86	13,83	12,37	10,29	9,80	11,04	10,45	11,05
Saúde e Bem Estar Social	Concorrência	1,45	1,82	2,11	2,05	1,94	15,28	19,50	25,91	25,91	26,84
	Evasão (%)	13,80	14,57	13,63	14,49	13,89	7,92	6,93	6,27	7,30	7,46
Serviços	Concorrência	0,88	1,17	1,03	1,24	0,85	8,51	13,19	16,46	15,39	15,01
	Evasão (%)	17,12	20,77	18,97	18,84	21,50	13,57	13,07	14,41	15,44	16,07

Tabela A.2: Evasão anual média nas grandes áreas do conhecimento (2001 a 2005)

Grande Área	2001	2002	2003	2004	2005	Média
Educação	19	17	16	21	15	18
Humanidade e Artes	22	23	23	24	25	23
Ciências Sociais, Negócios e Direito	23	24	25	27	24	25
Ciências, Matemática e Computação	29	27	27	29	28	28
Engenharia, Produção e Construção	21	21	22	22	20	21
Agricultura e Veterinária	17	17	22	16	13	17
Saúde e Bem Estar Social	18	17	20	19	19	19
Serviços	36	24	29	30	28	29

Fonte: Silva Filho et al. (2007) [88]

Tabela A.3: Evasão anual por sexo nas grandes áreas do conhecimento (instituições privadas)

Grande Área	Evasão (%) - Instituições Privadas									
	Masculino					Feminino				
	Ano					Ano				
	2010	2011	2012	2013	2014	2010	2011	2012	2013	2014
Educação	20,15	21,47	20,94	19,67	21,28	14,78	16,84	16,76	15,32	16,83
Humanidades e Artes	16,09	18,38	16,75	17,41	19,07	16,05	17,68	16,77	17,39	17,64
Ciências Sociais, Negócios e Direito	16,75	18,77	19,1	17,9	18,76	14,65	16,36	16,69	15,88	16,63
Ciências, Matemática e Computação	19,23	20,71	20,02	20,87	21,38	17,25	18,85	18,64	18,47	18,34
Engenharia, Produção e Construção	17,52	17,97	16,89	17,27	15,81	14,61	15,33	14,42	15,21	13,85
Agricultura e Veterinária	13,06	13,44	14,07	14,01	12,89	12,19	13,48	13,56	13,61	11,75
Saúde e Bem Estar Social	14,38	15,88	14,82	15,43	14,98	13,62	14,18	13,28	14,21	13,57
Serviços	17,7	21,62	20,37	19,58	23,05	16,67	20,14	18,03	18,36	20,47

Tabela A.4: Evasão anual por sexo nas grandes áreas do conhecimento (instituições públicas)

Grande Área	Evasão (%) - Instituições Públicas									
	Masculino					Feminino				
	Ano					Ano				
	2010	2011	2012	2013	2014	2010	2011	2012	2013	2014
Educação	16,66	15,67	17,66	17,89	16,22	12,07	11,67	13,10	13,49	12,62
Humanidades e Artes	16,20	14,53	13,68	15,87	13,28	13,39	12,63	11,52	14,15	12,36
Ciências Sociais, Negócios e Direito	11,58	11,25	11,61	12,03	11,95	9,26	8,96	9,14	9,76	9,32
Ciências, Matemática e Computação	15,47	14,91	16,11	15,65	16,28	13,97	13,42	14,22	14,57	15,48
Engenharia, Produção e Construção	12,42	11,29	11,82	10,69	11,34	9,79	9,61	9,90	9,32	9,94
Agricultura e Veterinária	10,79	10,27	11,54	10,86	11,63	9,63	9,20	10,45	9,98	10,42
Saúde e Bem Estar Social	8,70	7,68	7,67	8,20	8,54	7,61	6,64	5,73	6,94	7,01
Serviços	15,06	15,47	16,64	17,30	18,49	12,71	11,64	13,08	14,29	14,51

Tabela A.5: Evasão anual na grande área de Ciências, Matemática e Computação

Área	Evasão (%)									
	Instituições Privadas					Instituições Públicas				
	Ano					Ano				
	2010	2011	2012	2013	2014	2010	2011	2012	2013	2014
Ciências da Vida (Cursos Gerais)	-	-	-	-	-	-	20,73	29,89	22,75	-
Biologia e Bioquímica	16,38	18,76	18,25	17,18	16,74	13,86	12,72	11,59	11,81	14,52
Ciências Ambientais	10,50	19,07	12,93	10,77	5,67	10,18	14,91	18,07	17,59	19,18
Ciências Físicas (Cursos Gerais)	11,02	-	-	-	-	19,03	15,17	16,09	13,88	16,15
Física	21,26	35,52	18,91	17,56	18,05	21,16	25,97	20,65	24,09	21,74
Química	13,44	15,08	14,59	18,14	14,26	15,24	12,65	14,64	16,21	14,08
Ciências da Terra	12,31	17,61	20,02	18,46	13,86	10,75	11,15	11,47	12,54	13,13
Matemática	5,60	29,01	26,32	22,06	19,55	28,39	27,49	24,23	26,90	26,53
Estatística	19,81	22,20	20,62	17,37	18,68	14,87	14,45	17,51	18,39	18,19
Computação	19,42	20,74	20,16	21,10	21,89	14,23	13,87	16,18	15,12	15,88

Tabela A.6: Evasão anual nos cursos de Computação

Curso	Evasão (%)									
	Instituições Privadas					Instituições Públicas				
	Ano					Ano				
	2010	2011	2012	2013	2014	2010	2011	2012	2013	2014
Ciência da Computação	19,70	20,62	21,67	21,20	19,06	13,49	13,32	14,01	13,66	14,27
Engenharia de Computação	21,03	23,01	20,30	19,69	19,52	10,85	10,43	12,01	10,91	11,48
Engenharia de Software	-	-	-	16,39	4,39	9,87	10,14	13,07	9,06	13,66
Sistemas de Informação	17,83	18,50	17,78	19,26	19,42	16,38	13,36	13,74	16,50	15,54
Licenciatura em Computação	21,41	23,15	14,14	22,39	26,69	12,32	17,96	17,90	19,46	17,64
Tecnólogos	20,66	22,58	21,29	22,41	24,46	14,21	14,81	17,77	15,20	16,96

Tabela A.7: Evasão anual por sexo nos cursos de Computação (instituições privadas)

Curso	Evasão (%) - Instituições Privadas									
	Masculino					Feminino				
	Ano					Ano				
	2010	2011	2012	2013	2014	2010	2011	2012	2013	2014
Ciência da Computação	19,72	20,77	21,52	21,15	19,10	19,57	19,61	22,65	21,50	18,73
Engenharia de Computação	20,81	23,01	20,33	19,74	19,54	22,71	23,00	20,11	19,32	19,32
Engenharia de Software	-	-	-	13,64	4,40	-	-	-	41,67	4,35
Sistemas de Informação	17,81	18,68	17,72	19,31	19,37	17,95	17,6	18,11	19,03	19,73
Licenciatura em Computação	22,12	23,99	14,25	23,12	28,13	19,76	20,99	13,87	20,34	23,15
Tecnólogos	20,63	22,48	21,10	22,27	24,21	20,83	23,26	22,56	23,35	26,17

Tabela A.8: Evasão anual por sexo nos cursos de Computação (instituições públicas)

Curso	Evasão (%) - Instituições Públicas									
	Masculino					Feminino				
	Ano					Ano				
	2010	2011	2012	2013	2014	2010	2011	2012	2013	2014
Ciência da Computação	13,51	13,06	13,90	13,47	14,01	13,37	15,12	14,73	14,87	15,93
Engenharia de Computação	11,44	10,41	11,95	11,02	11,44	5,96	10,63	12,45	10,15	11,70
Engenharia de Software	9,22	10,05	13,53	9,17	13,45	15,63	10,71	10,53	8,51	14,75
Sistemas de Informação	16,35	13,51	13,88	16,34	15,12	16,49	12,82	13,26	17,09	17,08
Licenciatura em Computação	13,48	17,70	17,98	20,71	18,19	10,05	18,43	17,75	17,23	16,68
Tecnólogos	13,75	14,46	17,42	14,76	16,68	15,81	16,09	19,10	17,00	18,13

Apêndice B

Respostas dos Evadidos - Pesquisa de Evasão

B.1 Outros Motivos para ter Evadido

A seguir são apresentados outros motivos que tenham influenciado o aluno na decisão de evadir, que não foram elencados no formulário de escala Likert. Cada item corresponde a resposta de um aluno entrevistado.

B.1.1 Ciência da Computação

1. Conhecimentos e habilidades suficientes para fazer carreira na área mesmo sem concluir o curso.
2. Após já ter concluído o curso de Licenciatura Plena em computação, eu demoraria mais 10 semestres para concluir o bacharelado (por causa de questões de horário das disciplinas ofertadas, pré-requisitos, etc), e ter que lidar novamente com calouros recém-egressos do ensino médio enquanto tive aulas com professores alunos do mestrado que entraram na UnB depois de mim me motivaram a deixar o bacharelado de lado.
3. Fui aceito numa universidade americana.
4. Decidi sair e mudei de curso.
5. Resolvi fazer o curso de sistemas de informação.
6. Mudei para o curso de Engenharia Mecatrônica, que além de ter grande parte de sua base na área de computação, também aborda mecânica e elétrica.

7. O maior problema do curso (além dos péssimos professores) é o currículo completamente desatualizado. Poderíamos ter os melhores professores (o que está muito longe da realidade) que com o currículo atual o curso seria ruim. É uma perda de tempo em matérias inúteis ou com ementas ruins pra ganhar um pedaço de papel sem valor. E depois disso entra a péssima qualidade da maioria dos professores. Que se recusam a se atualizarem, mudar conceitos, método de ensino e por aí vai.
8. Conteúdo trivial, quantidade absurda de pré-requisitos desnecessários, currículo desatualizado.
9. Já fui desligado diversas vezes devido à depressão, entretanto, nesse tempo que fiquei fora da faculdade, trabalhei com programação e estudei muito por conta própria, o que resultou em grande desânimo no meu retorno aos estudos, pois além de já almejar trabalhar logo, eu tinha que chegar em sala para ver assuntos que já dominava, às vezes até tendo que corrigir professor, então não via mais sentido perder tempo ali, ainda que minha saída signifique não obter um diploma... Atualmente estudo opções de ensino à distância para poder obter um diploma, visto que tenho interesse também em dar aulas no futuro.
10. Disciplinas com foco em provas e não nos trabalhos onde se aprende mais
11. Reprovei 3 vezes em Estatística
12. Falta de conhecimento meu sobre o curso, entrei achando que seria algo totalmente diferente.
13. Eu entrei para a UnB em 01/2006. Devido à uma doença grave na família (minha mãe, que veio a falecer), eu acabei por, ao longo de 3 semestres, reprovar muitas disciplinas. Sabendo que o primeiro vestibular de 2009 seria o último vestibular que tiraria todas as reprovações do histórico e confiando que eu poderia passar no vestibular uma terceira vez, eu passei novamente para CiC, e, com a cabeça no lugar, concluí o curso sem maiores problemas. =)
14. A ciência da computação era meu segundo curso superior depois da licenciatura em computação e eu percebi que não existiria ganhos para mim em completar o bacharelado
15. Eu achei que faltou um pouco de companheirismo entre os colegas para se ajudar.
16. Concluí o curso de Licenciatura em Computação e iniciei um Pos-Graduação na Católica. Além disso tive uma serie de problemas administrativos na UnB, devido uma falha do sistema. Fui considerado em condição no ultimo semestre de licenciatura, no mesmo semestre consegui o duplo curso para o Bacharelado. Porém

considerando as regras da época e seria desligado do curso de Bacharelado. O que acredito ter ocorrido, pois nunca fui notificado de nada. Além disso antes iniciar o curso de computação era aluno de matemática Bacharelado. Hoje sou servidor publico no meu terceiro concurso todos de TI.

17. Muitas matérias de física e, principalmente, matemática teórica, com pouca ou nenhuma relevância para minha área. Acabei saindo para montar um portfólio, concentrando meus estudos autônomos no que era realmente produtivo e relevante, e começar a trabalhar como programador profissional, sem o diploma. Me arrependo por ter demorado tanto para fazer isso.
18. Na verdade, sempre que vou jubilar faço novo vestibular, pois não tenho computador para efetivamente colocar um site no ar. Calcular padrões para números primos etc. Deixar algo executando por meses. E aprender indubitavelmente.
19. Professores tendem a achar que todo mundo mora no plano e fazem provas em final de semana quando não tem ônibus
20. Física e cálculo, professores que exigem que se faça exercícios em excesso, nos obrigando a se dedicar quase que por completo a área deles e acabam dificultando o aprendizado sobre a área computacional e de automação
21. Distanciamento da academia com a realidade do mercado de trabalho
22. Era minha segunda graduação, vi que não ia me ajudar futuramente no meu emprego e larguei.
23. Procrastinação
24. Nunca quis fazer Ciências da Computação, só fiz por quê meus pai falaram para fazer, sempre quis fazer educação física.
25. Dificuldade de se matricular em disciplinas
26. Percebi que, apesar de gostar bastante do curso, não me identificava bem com a área, e mudei para outra com a qual eu era mais compatível.

B.1.2 Licenciatura em Computação

1. Jubilamento antecipado por baixo aproveitamento.
2. Entrar na universidade vindo de escola publica sem base estudantil e nao ter nenhum tipo de apoio. Greves, gerando transtorno, perda de conteudo, reposicoes de aulas em horarios incompativeis.

3. Ter que fazer outra monografia tendo acabado de concluir uma.
4. Currículo ultrapassado (na época) e muito maçante.
5. Baixa qualidade dos docentes de Humanas.
6. Conteúdo abordado de maneira superficial e pouca relação com projetos práticos.
7. Problemas familiares no semestre de desligamento
8. Falta de estímulo para iniciar o TCC
9. Dificuldade de passar na disciplina lógica computacional 1, turmas muito cheias complicado desenvolvimento e relacionamento com o professor, muito conteúdo nas disciplinas de matemática, não tem material didático , 12 créditos exigidos tornam a grade sufocante.
10. No semestre de desligamento estava passando por problemas afetivos e financeiros. Então, desisti do curso.
11. Dificuldades em Cálculo e Bolsa de Estudos no IESB
12. Falta de grupos ou tutoriais de ensino/estudo com menos pessoas ministrados por alunos do curso para ajudar no entendimento e ajudar alunos que não tiveram uma base de cálculo boa na escola para que estes pudessem nivelar com os demais alunos. Falta também um curso de bacharelado noturno e maior oferta de matérias a noite. Acho lastimável um curso noturno ter disciplinas as 16 horas ou choque de horários constantes impossibilitando que alguém sem flexibilidade no trabalho possa ter acesso aos estudos de nível superior.
13. Foi basicamente o serviço e a moradia. Morava na Ceilândia e trabalho em Santa Maria. No final do dia ir para UnB à noite seria uma rotina complicada. Atualmente estou morando no Gama e trabalhando em Santa Maria então fica mais fácil conciliar algum curso na FGA.
14. Má remuneração na área. Curso desatualizado para o mercado atual e muito longo para outras áreas não específicas. Professores com baixo interesse em ensinar didaticamente a matéria. Falta de incentivo dos professores e monitores para retirar dúvidas.
15. Não me via trabalhando a vida inteira com isso.
16. Dificuldade em concluir o trabalho de conclusão de curso.

17. Falta flexibilidade para a questão da presença, principalmente para um curso noturno onde, em teoria, seria voltado para quem já trabalha. Alguns professores não entendem isso.
18. Obrigatoriedade de fazer 2 TCCs em caso de duplo curso. Eu me formei em ciência da computação e não tinha condições físicas, mentais e temporais de escrever outro TCC em 6 meses.
19. Nenhum apoio e orientação acadêmico para ajudar a enfrentar as dificuldades naturais para conciliar o curso, trabalho e família.
20. Motivação pessoal, falta de disciplina e desgosto.
21. Extrema dificuldade para se matricular ou resolver outras coisas com as secretarias, burocracia extrema e desorganização total. Cada um empurra para outro falando que outro setor resolve e você acaba voltando para algum dos primeiros.
22. O principal problema foi a conciliação dos horários de trabalho com os horários da faculdade. Chegar muito tarde em casa, por volta de meia-noite, ter dois filhos para cuidar (hoje tenho 3), além de trabalhar. Foi um mau planejamento da minha parte ao não pensar seriamente sobre isso quando me matriculei. Atualmente faço Análise e Desenvolvimento de Sistemas à distância, curso tecnológico de 2 anos, o que é o ideal para mim. Acho que a UnB deveria investir mais em cursos à distância, pois nem todos querem seguir a carreira acadêmica ou desenvolver pesquisas/estudos. No meu caso eu tinha e tenho intenção de crescer profissionalmente na empresa onde trabalho, sou bancário da Caixa. Tenho formação em Licenciatura Plena em Física pela UFU em 2010 e Pós-Graduação em Educação Matemática em 2014, sem dificuldade em exatas. Apesar disso não tive equivalência permitida em Cálculo 2, mas somente em Cálculo 1 e isso desestimula também, ou seja, ter que refazer novamente Cálculo 2 desanima. Uma crítica que faço à UnB em geral é que, posso estar errado sobre isso, como disse, ela dá mais atenção à pesquisa em detrimento do público que quer apenas progredir profissionalmente no trabalho. Isso aconteceu também na Universidade Federal de Uberlândia, apesar de ser formado em Licenciatura, tive somente 2 disciplinas pedagógicas durante todo o curso - Psicologia da Educação e Didática. Talvez seja por isso que hoje não consigo lecionar Física em nenhuma escola particular, tenho muita dificuldade em dar uma boa aula que possa animar os professores avaliadores e ser contratado. A ênfase do meu curso foi demasiada em cálculos, mas sem nenhuma preocupação em como transmitir da melhor forma possível os conhecimentos de Física para os alunos. Há pesquisas que dizem que a carência de professores com formação em Física no Brasil é muito

grande. No entanto, muitos colégios particulares não contratam o professor somente pelo fato de ele ter um diploma de Licenciatura em Física. Há muitos professores de Física que estão lecionando sem terem diploma de Física, mas sim diplomas de Matemática ou de Química. Um curso de Licenciatura deveria, no meu ver, dar mais ênfase a disciplinas pedagógicas, senão o profissional fica muito despreparado para o mercado, que é a sala de aula. Posso ter desviado um pouco do assunto principal desta pesquisa, mas acredito que essa situação por mim mencionada pode ser equivalente ao que ocorre com a Licenciatura em Computação. Agradeço esta oportunidade de poder fazer essa pesquisa, acho muito válida e parabéns pela ideia.

23. Na época, mudei-me para Goiânia.
24. Dificuldade de estudar as matérias e o trabalho que não conseguia conciliar.
25. TDH
26. Falta de reconhecimento do grau de “licenciado” no mercado. Geralmente as empresas reconhecem mais o bacharel e o curso de Ciência da Computação.
27. A área da computação com a qual eu já havia estado trabalhando, web, não era contemplada no curso

B.1.3 Engenharia de Computação

1. Professores que acham que estão fazendo um favor em dar aula e que não estão dispostos a ensinar.
2. O suporte que o departamento fornecia era próximo de zero. O aluno calouro pouco recebe qualquer ajuda dos seus "peers" e professores coordenadores.
3. Professores mais preocupados com pesquisa do que ensinar; Muitos conteúdos pouco relevantes para o mercado de trabalho atual; Não é necessário ter diploma para ser um bom desenvolvedor ou conseguir um trabalho na área; Falta de muitos conteúdos muito importantes para o mercado de trabalho (pelo menos até onde eu fui e conversar com as pessoas), como TDD, Git (controle de versão)
4. Aproveite o espaço para dar uma resumida e facilitar o colhimento de dados para vocês. Comecei a cursar engenharia de computação em 2011, o curso era bem diferente do que eu imaginava, em especial a unb. A matéria do primeiro semestre era interessante e algo completamente novo para mim, nunca tinha vista nada assim na vida. Entretanto, não consegui acompanhar o ritmo de matérias dadas, e acabei por reprovar algumas matérias nos primeiros semestres, mas tentei continuar cursando para ver se melhorava. Não houve melhoria do rendimento e as matérias

ficavam cada vez mais complexas, e claro que reprovacoes desestimulam e ao longo do tempo foi se tornando algo desagradavel frequentar a unb. No 6 semestre decidi tentar outro curso e estou gostando bastante, nao me arrependo de ter trocado.

5. Depressão severa e morte de ente próximo.
6. Incentivo por parte dos professores.
7. Por ter muita dificuldade com a disciplinas de programação acabei sendo desligado. Gostaria de ter continuado tentando, mas não foi possível.
8. Me desliguei de engenharia da computação para ingressar em ciência da computação por maior afinidade e interesse.
9. Não conseguia ver aplicação prática na maior parte do conteúdo estudado em disciplinas em sala. Muita ênfase numa disciplina que me parecia pouco aplicável no dia a dia do profissional. O ambiente acadêmico no qual me inseri não me cativou e não me motivou a correr atrás de um aperfeiçoamento. Matérias nas quais eu tirava boas notas não me motivavam a buscar uma pesquisa ou projeto. Algumas provas incoerentes com a ementa, prefere-se medir a atenção do aluno por meio de “peguinhas” ao invés do conhecimento. Colegas de pouca confiança e alguns poucos professores de difícil convivência me desmotivaram. Mas vale ressaltar que a maioria dos professores foram bons, mas os ruins marcaram. Hoje, não consigo pisar na Unb com o mesmo vigor e motivação da época de calouro. Espero conseguir um dia.
10. Baixa expectativa salarial e profissional após o término do curso.
11. Passado em concurso federal
12. Na verdade queria trocar de curso logo depois do primeiro semestre, logo depois que ingressei na UnB meu pai ficou desempregado e com depressão porque ficou muito tempo sem trabalho, por motivos de eu não ter um notebook, pois isso influenciava muito em relação as materias de programação, e a dificuldade que tinha nas materias por falta de uma base melhor no ensino medio, pois sou de escola publica e sentir uma dificuldade enorme principalmente nas disciplinas que tinha matematica.
13. Burocracia - por greve e morosidade dos coordenadores do curso e de outros cursos na análise do currículo a ser aproveitado (ja tinha graduação em matemática), estava sendo "obrigado" a cursar disciplinas em que ja estava habilitado.
14. Meu motivo de ter saído de Engenharia da Computação não foi por achar o curso difícil, foi porque eu gostava muito mais de matemática e física e eu queria ver matemática e física aplicada, algo que não encontrei na Engenharia da Computação,

onde eu vi praticamente programação. Ao mudar para Engenharia Mecânica eu fiquei realizado. E também eu já sabia que não queria fazer Engenharia de Computação, quando passei no vestibular, eu tinha escolhido esse curso porque era mais fácil de passar e o começo de todas as engenharias era bem parecido. Fiquei um tempo em computação porque como eu ainda tava em dúvida sobre qual engenharia eu queria cursar, mas sabia que queria fazer engenharia, então fui ver se me encontrava na engenharia da computação, o que não aconteceu. Porque minha ideia de engenharia sempre foi a de um curso para quem gosta bastante de matemática e física e quer ver elas sendo aplicadas, mas ao chegar na computação, percebi que o curso se tratava quase todo de conhecimentos novos estabelecidos do zero, sem usar quase nada de matemática e física como base, muitas matérias e programação, que era algo completamente novo. Não achei o curso ruim, programar era legal, mas não era o que eu queria fazer, eu gostava mais de matemática e física e queria ver novas matérias usando aquilo como base e não ver matérias novas completamente aleatórias. Mas como já tinha entrado com o objetivo de não ficar em engenharia da computação, isso não foi tanto um problema.

15. Opções do mercado de trabalho
16. Desanima muito, chegar na sala, no primeiro dia de aula de um semestre e escutar: Então gente, 80% de vocês vão reprovar. Chegar em uma disciplina onde tem 2 professores dando a matéria, ambos com 30 vagas. O primeiro tem 37 alunos cadastrados e uma fila de espera enorme, o segundo tem 4 pessoas cadastradas. Escutar a frase Não pega com professor tal que você vai reprovar, ele passa uns 2 por semestre só.
17. Preconceito de alguns professores pelo fato de ser mulher, marcação.
18. Esperava que o curso realmente me ensinasse a desenvolver computadores e sistemas computadorizados, mas senti que na verdade ele apenas ensinava ciência da computação (e geralmente de maneira muito maçante e padronizada. Chegava a aprender mais usando o Google do que indo às aulas) com um pouco a mais de engenharia. Uma pesquisa rápida na minha época de graduando me mostrou que tem curso técnico aqui no DF que segue mais a formação de engenheiro da computação que o currículo da graduação e hoje conheço um colega de curso que está fazendo uma graduação adicional em Goiânia pra poder suprir as deficiências da grade da UnB. A grade que não nos dá valor no mercado de trabalho faz com que os formandos tenham que “roubar” profissões de outros cursos, principalmente os de ciência da computação (algo que ainda vejo acontecer constantemente no curso), justamente pela falta de uma formação adequada. Isso chegou ao cúmulo de eu

ver veteranos meus (que estavam perto de formar, diga-se de passagem) tendo que levar o computador pro conserto numa loja qualquer porque não sabiam arrumá-los, o que obviamente é um absurdo se parar pra pensar. É como se um médico tivesse que pedir ajuda para um enfermeiro porque não sabe fazer o diagnóstico de um paciente com gripe. Se já não bastasse isso, via com muita evidência que os departamentos "desprezavam" nosso curso; não conseguiam ver o graduando como aluno de engenharia da computação, mas sim como um "genérico" de cientista da computação e engenheiro elétrico. Acabávamos recebendo pouquíssima atenção por causa disso, e via meu curso sendo tratado como um problema incômodo ao invés de curso universitário. O corpo docente, apesar de ter boa qualificação, não cultivava o potencial dos alunos. Não importava se você se excedia ou não, se você fez seus trabalhos ou não; eles estavam ali só pra dar a ementa e pronto. Chegou ao ponto de eu achar que isso era algo normal, até entrar no meu curso atual e ver que nada disso ocorria. Isso prejudica muito os alunos, pois eles seguem essa mentalidade e passam a ver o curso apenas como uma etapa pra bater o ponto e ganhar um diploma, e não como uma oportunidade de desenvolver pesquisas, criar produtos e se desenvolver como indivíduo.

19. Problemas de ordem emocional, que me fizeram não conseguir continuar no curso.
20. Diversos professores da matemática tem critérios absurdos de correção e muitos escolhem a dedo quem passa e quem não passa. Inúmeras vezes fui reprovado por critério de criatividade ou levar zero numa questão pq fiz certo mas não como ele queria sendo que metodologia nenhuma havia sido especificada. Na tentativa de reclamação desses acontecimentos a única resposta que recebi e que o professor em tem o direito de fazer o que quiser com aula dele. Já peguei a matéria cálculo 2 o professor chegava bêbado e com cheiro de pinga. Levei diversas testemunhas do ocorrido e a única medida tomada foi perguntar para ele se ele tinha ido a aula bebado...me perdoe mas só um idiota afirmaria essa acusação. Eu saí da UNB principalmente pelo desgosto de além de lidar com a total marginalização do meu curso com ainda tinha de lidar com universidades que se importa mais com um professor do que com os alunos que era largados a mercê do passasse na cabeça deles já que não há nenhum regulador de qualidade dentre os professores e nada que assegure que eles façam um bom trabalho. Sinceramente se e um favor estudar na UnB e não um investimento prefiro me dedicar a algo o que possa ser valorizado e não ridicularizado.

B.1.4 Engenharia de Software

1. Dificuldade de aprendizado devido ao TDHA
2. Problemas alérgicos graves, pois o campi do Gama, é extremamente empoeirado mantendo-me constantemente com crises de renite aguda. Outro fator importantíssimo é o fato de uma baixa aderência do que é cobrado nas matérias matemáticas com a realidade de alunos oriundos do ensino público, as matérias como Álgebra linear e Calculo, são levadas a mão de ferro, seria prudente no primeiro semestre a matéria de pré-cálculo e não Calculo 1 e Álgebra Linear pela definição.
3. Curso mal planejado, engenharia de software com matérias que não tem relação com software incluídas na grade, e te fazem perder tempo. Infraestrutura deficiente, professores que não sabem ensinar.
4. Tenho a sensação que os professores da UnB parecem querer competir com os alunos ao invés de ajudar. Totalmente contrário tanto a pedagogia quanto à ciência.
5. No meu caso fui nomeado em um concurso público federal e solicitei a antecipação de outorga de diploma para poder tomar posse e assumir o cargo, porém a unb não apresentou rapidez no processo nem com mandato de segurança, por isso tive que paralelamente fazer aproveitamento das disciplinas e me formar em outra universidade. Como após isso obtive o diploma de graduação não havia motivo para dar continuidade ao curso na unb.
6. Distância do campus para a minha residência.
7. Na desistência do curso de Engenharia de Software, a distância do campus também afetou a decisão. Sou morador do Plano Piloto e o curso pertence ao campus do Gama.
8. Falta de motivação e desproporcionalidade de cobrança por parte dos professores.
9. Apoio estudantil voltado para jovens, nenhum tipo de apoio para adultos
10. Transporte para a universidade
11. Falta de aderência com a realidade do que é a engenharia de software, impondo disciplinas da área da matemática e química.
12. Tenho um grande problema em relação as faltas, por trabalhar e por ter como prioridade o meu trabalho essa sempre foi uma constante na minha vida acadêmica, por várias vezes quase pedi transferência para o UniCeub. A inflexibilidade de muitos professores em relação à alunos que trabalham e/ou que são capazes de

estudar e aprender sozinhos fora de sala de aula e “aparecem” apenas para cumprir com os exercícios e provas demandadas sempre foi algo que me incomodou muito. Acho que se eu estudasse no Campus Darcy Ribeiro ao invés do Campus do Gama não teria tantos problemas com faltas, por ser um campus bem mais acessível e muito próximo ao meu trabalho, mas prefiro estar no Gama por me identificar mais com o curso de Engenharia de Software do que com o curso de Ciência da Computação.

13. Problemas Psicológico: Depressão.
14. Roubaram meu carro financiado no campus do gama, sai pra trabalhar e pagar a porra do carro.

B.2 Motivos para Escolha do Curso de Computação

A seguir são apresentados os motivos para escolha do curso de Computação, de acordo com a pesquisa feita com os alunos evadidos nos cursos de Ciência da Computação, Licenciatura em Computação, Engenharia de Computação e Engenharia de Software da UnB. Cada item corresponde a resposta de um aluno entrevistado.

B.2.1 Ciência da Computação

1. Pressão familiar
2. Mercado de trabalho
3. Afinidade com a área.
4. Escolhi por ter afinidade com a área desde a adolescência, mas fui mal assessorado na escolha, talvez tivesse feito alguma engenharia (elétrica ou mecânica) se tivesse uma melhor compreensão do mercado e do conteúdo de todos os cursos antes.
5. Gosto de programação.
6. Procurava um curso na área de exatas e após várias tentativas de ingresso em engenharia civil sem obter sucesso, fui aprovado para o bacharelado em ciência da computação
7. Mercado de trabalho, identificação com a área.
8. Já tinha me formado em outro curso na UnB, mas sempre tive contato com computação e programação desde a adolescência. Foi uma oportunidade pra buscar algo que sempre quis estudar. No entanto, não foi possível conciliar com o trabalho e com estudos pra progredir no serviço.

9. Não tinha nenhum curso definido, optei por algo que envolvesse tecnologia.
10. Falta de opção
11. Afinidade com a área e paixão por programação
12. O curso que eu realmente queria não possui, em geral, um bom retorno financeiro ou reconhecimento. Então escolhi este pois me identificava com as matérias de exatas, fiz o vestibular e passei. Porém o curso não despertou meu interesse ou alguma motivação para continua-lo então fiz o vestibular novamente e ingressei no curso que queria desde o início.
13. Ainda gosto bastante de computação, mas tive que mudar pois tive problemas pessoais e não consegui ter bons rendimentos. Então a Universidade me expulsou do curso.
14. influência de professores no ensino medio
15. Jogos de computador
16. Afinidade com a área de informática.
17. Escolhi por afinidade a área
18. Interesse e experiência prévia em programação.
19. Interesse em "criar" e "inovar", utilizando a programação de computadores.
20. Afinidade com computadores e redes de informação
21. Computação é uma das áreas mais interessantes que conheço. Sempre gostei da forma que computadores funcionavam e a ideia de resolver problemas através deles. Criar algoritmos, calcular e pensar de forma lógica é divertido.
22. Falta de opções no meu nível acadêmico da época já que não pude escolher ir direto para Eng Civil devido a baixa nota do ENEM. Além disso, ainda não sei qual carreira seguir
23. Afinidade
24. Aptidão
25. Afinidade a computação desde muito novo, interesse na área de segurança e pentesting em geral.
26. Efeito manada
27. Exatas

28. Falta de opções
29. Afinidade com uso de computadores e vontade de aprender sobre computação.
30. Não sabia ao certo que curso fazer e este curso foi o pareceu que eu mais me identificava
31. Gosto de computação
32. Futuro profissional
33. Sempre tive interesse pela área de Tecnologia da Informação.
34. Me afinava com área e exatas e com computadores
35. Práticas anteriores em programação e interesse em computação gráfica.
36. Sempre tive interesse em trabalhar com tecnologia
37. Opção familiar
38. Sempre gostei de computador, trabalhar com o mesmo. Acreditei que seria interessante compreender como tudo funcionava, como criar programas etc.
39. Interesse em programação
40. Achei que seria prazeroso trabalhar na área.
41. Na verdade não pensei da forma que deveria antes de ingressar, aproveitei a chance para sair antecipadamente do ensino médio. Hoje eu preferiria ter avaliado melhor qual curso escolher.
42. Para me auxiliar nas atividades do trabalho
43. Afinidade com computadores, logica de programação, matemática e curiosidade sobre criação de jogos.
44. Influências familiares, pelo fato do curso "dar dinheiro"
45. A pessoa entra no curso pensando que ele é uma coisa, mas com o tempo vê que se trata de outra coisa completamente diferente, programar um jogo é completamente de jogar um jogo ou saber utilizar aquele programa na prática.
46. Afinidade com informática e jogos eletrônicos
47. Escolhi pela afinidade por computadores e inovação.
48. (não respondeu)
49. Afinidade.

50. Desde pequeno já gostava de jogos e tecnologia. Sempre fui muito curioso em relação à como as coisas funcionavam, ávido leitor de Scientific American/Popular Mechanics/etc... e conforme fui tendo acesso a tecnologias mais novas, mais tinha interesse em como eram implementadas.
51. Devido a curiosidade e gosto por tecnologia
52. Afinidade com programação e cálculo
53. Por afinidade com a área
54. Afinidade Com o tema do curso
55. Não sabia o que realmente queria fazer
56. Eu queria algo que eu me identificasse e que me abrisse portas tanto para concurso público quanto pra trabalhar na área particular.
57. Afinidade com a área. Até hoje intercalo a computação com a minha profissão.
58. Pensava que gostaria do curso
59. Por que eu amo e sou desenvolvedor de jogos. Trabalho como programador em uma empresa de jogos atualmente (Aquiris Game Studio). Só não me dei bem com estatística, que é obrigatória. E tenho mais um game em desenvolvimento www.turbolinkgame.com
60. Escolhi baseado na vontade de trabalhar com programação de jogos,mas sem consultar informações sobre o vídeo
61. Queria trabalhar na área de jogos, porém voltado mais na área de arte e criação, e no curso percebi que só seria estudado o lado de programação.
62. Eu escolhi computação tendo uma vaga (vem vaga) idéia do que era “programar”, e o curso sempre foi interessante pra mim, eu nunca desanimei do curso ou me vi fazendo outra coisa.
63. Eu gosto de computação e estudo desde os 12 anos
64. Escolhi, pois meu pai era formado na área e gostava de jogos eletrônicos, mas me decepcionei com o curso, pois nem sempre conseguia acompanhar a lógica das tarefas
65. Afinidade com computadores
66. Sempre fui muito ligado à computadores e jogos eletrônicos então achei que pudesse me encaixar no perfil de estudante de ciências da computação, o que não aconteceu. Apesar de apresentar extrema facilidade com matérias de exatas nos ensinamentos básicos

e médio, apresentei grande dificuldade nas matérias de física e matemática do curso de computação, o que acabou por auxiliar na desmotivação que sentia em relação ao curso.

67. Afinidade com a área de programação
68. Me identificar muito com área de exatas !
69. Afinidade com a área desde pequeno
70. Porque gosto de programar
71. Sempre quis trabalhar com jogos, e programava mods para jogos e bots automáticos de MMOs durante meu ensino médio. Já tinha uma boa base de programação. Computação pareceu uma escolha óbvia.
72. Eu já tinha criado alguns blogs e acabei de interessando por programação
73. Sou bom em programação e acho que é o melhor caminho.
74. Afinidade com o tema durante a juventude
75. Sempre gostei da área de computação
76. Era um usuário entusiasmado, pensei erroneamente que seria bom na área.
77. Gosto bastante de tecnologia.
78. Desde criança queria criar jogos eletrônicos, ciência da computação pareceu o mais adequado para tal finalidade.
79. Por que via bom potencial para encontrar emprego
80. Afinidade com computação
81. Gostaria de trabalhar com jogos eletrônicos e havia uma professora do departamento de Computação que procurava ajudar nessa formação dos alunos.
82. Sou profissional de TI até hoje. Amo a área. Sou especialista em Gerência de Projetos e Arquitetura de Baixa Plataforma. Não me formei. Tenho vontade de voltar a estudar e completar o curso algum dia.
83. Fiz vestibular quando estava no último semestre do meu outro curso e fiz por fazer e passei.
84. Perspectiva de uma profissão com um bom retorno financeiro.
85. Afinidade

86. Sempre fui bom com eletrônicos, por isso meus pais aconselharam a fazer ciências da computação.
87. Gostava da área, tinha certa afinidade com matérias da área de exatas e no momento do vestibular não trabalhava e com isso possuía tempo para estudar para o curso.
88. Oportunidades de trabalho
89. Afinidade.
90. Eu tinha, e ainda tenho, interesse por programação de computadores.
91. Identificação com a área, afinidade com programação
92. Interesse pessoal na área.
93. Sempre foi algo que me atraiu bastante. Porém, mais como um hobby do que como profissão.

B.2.2 Licenciatura em Computação

1. Afinidade
2. Já era a profissão praticada desde antes.
3. Por ter muito contato com computador, imaginei que o curso me ajudaria a ter uma formação superior na área.
4. Faltavam apenas 30 créditos para completar a dupla-habilitação em Licenciatura e eu tenho vontade de lecionar.
5. Desde pequeno já estava inserido em vários cursos correlatos a área
6. Porque eu gostava da área e ela me proporcionaria um bom futuro.
7. Queria aprender a programar e entrar na área de computação, mas acabei desistindo de fazer isso via o curso.
8. Foi uma decisão natural. Ao longo da adolescência, por gostar de jogos e internet.
9. Possibilidades de carreira
10. Por querer me aprofundar em Ciências da computação
11. Depois de ter me formado em curso de humanas e muito teórico (Letras), queria um curso na área de exatas e mais prático.
12. Era muito novo gostava de computadores e achava que gostava de exatas

13. Eu já tinha formação em outra área e possuía projetos nessa área de formação inicial que dependiam de conhecimentos de programação. No meio do curso eu já tinha desenvolvido o projeto com conhecimento adquirido fora da faculdade estudando por conta própria. No fim das contas o curso de computação foi uma boa experiência mas que serviu mais para conhecimento geral do que efetivamente para formar o conhecimento necessário que eu buscava.
14. Na época queria transferência para Engenharia da Computação
15. Afinidade com a área
16. Afinidade com o curso, porém conteúdos que não condizem
17. Fiz um curso de banco de dados no Órgão onde trabalho e gostei muito.
18. Gosto da área de Redes de Computadores, no entanto, trabalhava durante o dia para cursar Engenharia de Redes. Como minha disponibilidade era apenas no período noturno, escolhi o curso Computação - Licenciatura, mas nada a ver comigo. Então, fui cursar Redes na Universidade Católica de Brasília.
19. Porque achei que ia administrar servidores, criar alguns programas, aprender sobre redes.
20. Vestibular fácil
21. Desde os treze anos de idade, tinha bom relacionamento com computadores, ao me familiarizar com HTML, CSS.
22. Desde os 12 anos de idade me identifico com a área de tecnologia, tanto que estou nela já à 6 anos como profissional, ou seja, desde o meu ingresso no UnB. Antes disso já participava de cursos e a cada um deles eu admirava mais o foco do trabalho em TI. Estou formado hoje em Tecnologia de Redes de Computadores e estou matriculado no curso de Especialização em Segurança de Redes na Faculdade de Tecnologia FT da UnB. Se pudesse, voltaria ao curso de Computação, afinal hoje o meu "know how" é completamente outro, assim não teria tanta dificuldade.
23. Não existe o curso de ciência da computação noturno. Era a opção que tinha maior afinidade com a área profissional que atuo.
24. Porque era a única opção noturna, se o curso de bacharelado em ciência da computação noturno existisse nunca teria feito licenciatura.
25. Na época que estava fazendo física cursei ICC no verão com o professor Ly Freitas. Desde então, sempre que dava certo fazia matérias de computação. Resolvi entrar na computação como meu segundo curso por que acho que programar é algo fantástico.

26. Gostaria de ter feito o Bacharelado em Ciência da Computação, mas tive que optar por um curso noturno, pois trabalho. Não tenho interesse em disciplinas sobre educação, pois não pretendo trabalhar na área. Minha mente se abriu um pouco para o ensino à distância, no entanto.
27. Afinidade com a área de Tecnologia da Informação
28. Interesse e experiência prévia na área de computação.
29. afinidade com a área
30. Foi a opção que me pareceu melhor.
31. De acordo com a conjuntura da época era uma área em franca expansão e um mercado de trabalho bem aquecido. Familiaridade com a área.
32. Afinidade com a área de atuação.
33. Profissional da área de TI
34. Já trabalhava na área.
35. Já trabalho na área
36. Afinidade com TI
37. Único curso de computação no período noturno.
38. Interesse na área
39. Pretendia ir para a área de computação, mas focado na área de software e não em modo geral (como pesquisa)
40. Desejo de conciliar educação e TI
41. Influência familiar
42. Achava interessante e com um bom mercado de trabalho.
43. Amo computacao e tecnologia
44. Importância da área para o desenvolvimento do país.
45. Porquê sempre tive afinidade, parecia um bom mercado de trabalho, etc.
46. Crescer profissional da empresa onde trabalho.
47. afinidade com tecnologia e mercado de trabalho promissor
48. Gosto e trabalho na área.

49. Aliar licenciatura com atividades de T.I.
50. Afinidade com programação
51. Devido o mercado de trabalho que está em expansão.
52. Gosto por informática no geral.
53. Gosto da área de informática e trabalhava o dia todo. Então só podia cursar a noite. Na época só noturno só tinha o licenciamento em Computação
54. Gosto de programar e era um curso noturno
55. Achei ter uma aproximação maior com meus gostos pessoais por tecnologia.
56. Por ser noturno
57. Pela identificação e facilidade com estudos computacionais
58. Gosto de informática desde a infância. Entrei no curso acreditando que receberia uma formação integral (aprender a fazer e pensar sobre o que estou fazendo, para tomar decisões). Entretanto, a maioria dos professores é tecnicista (não pense, faça isso do jeito que estou mandando). Alguns professores empurram tendências com fervor religioso e inadequado à Academia (ex.: computação ubíqua), e outros usam da sala de aula para fazer palanque político (lembro-me de dois professores do departamento que costumam fazer doutrinação liberal em momentos inoportunos). Os alunos estudam Lógica Computacional, mas não sabem defender argumentos quando chegam ao TCC. Implanta-se uma cultura não do diálogo com outras áreas e correntes da mesma área, mas do confronto, em que o importante é encontrar algo para se defender, mesmo que se usem os argumentos mais sujos e desonestos. Isso não é ciência. O departamento não se preocupa em desenvolver cientistas, mas sim religiosos que olham para os colegas de outros grupos de pesquisa como infiéis que devem ser combatidos. (o texto A Ciência como Vocação, de Max Weber, explica como professores universitários podem agir como pregadores, e isso ocorre *muito* no CIC e no Instituto de Exatas em geral - http://www.lusosofia.net/textos/weber_a_ciencia_como_vocacao.pdf) No mais, deixo o texto abaixo, que detalha como a falta do pensar sobre prejudica os cursos de Exatas: <http://vida-estilo.estadao.com.br/blogs/daniel-martins-de-barros/quando-humanas-e-exatas-brigam-todo-mundo-sai-perdendo/>
59. Trabalho com computação
60. Trabalho na área sem a graduação

61. Devido a uma idade prematura para escolha de um graduação, nessa idade, completando o ensino médio havia uma certa pressão minha e dos familiares para o ingresso na universidade de Brasília. Depois de iniciar o curso percebi que dificilmente eu me encaixaria no perfil de um profissional de computação.
62. Trabalho na área, além de gostar muito
63. Por afinidade área do curso e perspectiva profissional motivaram a escolha e foi feita após analisar os cursos relacionados com TI.
64. trabalho na área
65. Já trabalho nesta área há muitos anos.
66. Mercado de trabalho
67. Facilidade com informática e desenvolvimento lógico-racional.
68. Afinidade com a profissão . Trabalho com TI há mais de 20 anos e também gosto de ensinar.
69. Foi escolhido de maneira arbitrária.
70. Sabia que teria nota para passar, no PAS, para o curso, por isso optei pela opção segura. Mudei de opinião sobre o curso pretendido e, depois de passar via vestibular, me formei em C. biológicas e hoje tenho mestrado em Ecologia. Aproveito para dizer que seria interessante ter uma opção “Não se aplica” nas respostas acima (além da escala de 0 a 5)
71. Aptidão com a área.
72. Por afinidade.
73. Gosto da área e trabalho nela
74. Eu acompanhei esse curso desde a sua implantação do mesmo na UnB (em 2000) e achei que seria a melhor opção para mim. Esse curso seria a opção perfeita para mim já que eu gosto muito de computação e também da docência.
75. Trabalho na área desde 2002, na época achava que um diploma pudesse ajudar minha carreira, eu devia saber melhor, já estava “dentro”, a partir daí o que conta não é formação acadêmica, é networking e marketing pessoal.
76. Escolhi porque sempre gostei de mexer com computadores e queria saber como funcionava mas me decepcionei com o q encontrei.
77. Afinidade com a área de tecnologia.

78. Afinidade com informática desde criança
79. Sempre tive afinidade com a computação. Foi uma escolha natural.
80. Por já ter conhecimento prévio e ter trabalhado na área
81. Ingenuidade. Gosto por hardware, software, mormente jogos.
82. Afinidade
83. Trabalho na área de redes e queria ter um nível superior na área
84. Sou formada em matemática, mas comercialmente a computação é melhor para se trabalhar. Aproveitei que já tinha metade do curso para tentar me formar também em computação.
85. Sempre achei a área interessante

B.2.3 Engenharia de Computação

1. Sempre me interessei pelo assunto.
2. Escolhi cursar engenharia da computação por afinidade com o curso e continuo cursando so que em outra universidade.
3. Influência familiar. Não sabia o que cursar e possuo muito parentes engenheiros.
4. Incentivo dos pais para estudar nas áreas de tecnologia para ter um futuro melhor.
5. Porque é uma área de que gosto muito e até hoje ainda sou muito ligado a ela.
6. Pretendia cursar engenharia elétrica, porém não obtive pontuação no vestibular e acreditei que teria afinidade com o curso de engenharia da computação.
7. Eu já havia tido uma experiência com programação e gostado, achei que o curso seria uma continuação da vida acadêmica até o momento (ensino fundamental/médio) - onde os professores são avaliados pela sua capacidade de ensinar, não de pesquisar - voltado para uma área de interesse, e que me formando eu naturalmente me tornaria um bom desenvolvedor com um salário bom
8. Inicialmente, esperava que o curso estivesse mais atrelado a engenharia elétrica e computação, mas na verdade o curso foca basicamente em programação e a parte de elétrica e eletrônica se mostra fraca.
9. Afinidade com computadores
10. Afinidade com programação

11. Teste vocacional errou
12. Dentre os motivos que me influenciavam à época de escolha de curso de graduação, gostava muito de matemática e tinha altíssima afinidade com computadores e sistemas, por isso, achei que fosse a escolha mais assertiva.
13. Ótimo mercado de trabalho
14. Meio que accidental, prestei vestibular e passei
15. Sempre gostei de computadores e programação, mas esperava um curso mais prático.
16. Tinha 17 anos e passei no vestibular no meio do terceiro ano. Curso escolhido como uma forma de testar minha capacidade de passar no vestibular.
17. É o que a nota do ENEM permitia
18. Achava que o curso seria mais pratico com hardware
19. Similaridade com Ciência da Computação
20. Apreciação por programação.
21. Por que é uma área que sempre tive interesse e maior conhecimento.
22. Gostava de computação.
23. Pensava ter afinidade e facilidade.
24. Por interesse gosto muito da área de computação
25. Afinidade com programação
26. Gostava e ainda gosto de tecnologia e de computadores. Era um aluno acima da médias nas disciplinas de matemática e física. Sempre tive um gosto por ciência e tecnologia. Estudo programação hoje por conta própria, o interesse por estudar programação para fins acadêmicos diminuiu muito. Me interesse ainda por áreas como Inteligência Artificial e Redes Neurais.
27. Sempre fui apaixonado por computadores. Eu sabia que queria fazer computação desde que minha família comprou um computador quando eu tinha cerca de 5 anos de idade, desde então me tornei um "viciado" em computadores, procurando entender como funciona, etc. Antes de ter entrado na engenharia de computação, eu nunca havia considerado com seriedade fazer outro curso que não estivesse diretamente relacionado a computadores.
28. Assim como muitos que conheci e que também abandonaram algum curso da área de TI, acabaram escolhendo o curso por gostar de mexer com computadores.

29. Achava que aprenderia a programar
30. Tinha afinidade na area de TI
31. já tinha afinidade com a área
32. Eu escolhi Engenharia de Computação porque eu achava mais fácil de passar, como eu só poderia fazer o vestibular no sistema universal, minha estratégia sempre foi escolher um curso de engenharia mais fácil de passar e depois mudar para Mecânica ou Civil, que era as duas engenharias que eu pensava em fazer. Como o começo de todas as engenharias é bem parecido, não foi tanto problema fazer apenas o começo de engenharia de computação.
33. Desejo do meu pai era que eu cursasse alguma engenharia, eu escolhi computação por achar um pouco mais interessante
34. Saindo do Ensino Médio decidi cursar Engenharia de Computação devido à grande quantidade de emprego na área e à possibilidade de um salário alto.
35. Escolhi olhando o edital de um concurso específico. No fim das contas outras áreas que me agradam mais me habilitam para o mesmo concurso.
36. Queria fazer alguma engenharia e escolhi computação. Nunca estive tão errado
37. Afinidade
38. Afinidade com programação básica
39. Tinha facilidade em exatas,pensei em fazer engenharia, como meu cunhado fazia CIC na época, me influenciou bastante.
40. Perspectiva de mercado e interesse pessoal.
41. Afinidade com a área
42. Influência da família, amigos e análise de mercado
43. Curso do futuro
44. Achei que era um curso com um grande mercado de trabalho e que poderia ter facilidade de entrar.
45. Acabei escolhendo o curso com o auxílio de alguns tutores no meu ensino médio, e escolhi a área porque eu tinha muito interesse na área de computação, visto que é atualmente uma das grandes fronteiras da ciência (algo que muito me interessava).
46. Influência dos pais

47. Sempre tive afinidade com lógica matemática e tecnologia
48. Escolhi como segunda opção, por influências familiares
49. Gosto de lógica, computação e eletrônica.
50. Entrei e não me sentia confortável tanto com os colegas como com algumas matérias, me estressada muito, então decidi procurar outro curso e me identifiquei com cursos de saúde/biológicas
51. Afinidade, grande interesse na área.
52. Novo vestibular para curso com qual há mais afinidade
53. Por mera afinidade. Sempre gostei de computação

B.2.4 Engenharia de Software

1. Por que é um curso voltado para área de programação e que eu tenho afinidade com a área de programação
2. Afinidade com a informática
3. Mudei de engenharia automotiva para de software
4. Sempre quis abrir uma empresa na qual eu poderia ser a única mão de obra suficiente para que ela funcionasse e obtivesse crescimento sem aumentar funcionários. Somente TI me proporcionaria isso.
5. Eu já era técnico em informática, e fascinado pela área de desenvolvimento de software após uma longa análise nos cursos existentes escolhi engenharia de software porquê se enquadrava melhor no meu perfil.
6. Não houve método de escolha. Escolhi o curso pelo fato de ter muitas vagas
7. Por gostar de programação e exatas.
8. Me identifico com lógicas e tecnologias
9. Afinidade
10. tentativa de mudança do curso de engenharia eletrônica
11. Visão preconcebida sobre o curso e influência de amigos que cursavam.
12. Sempre tive facilidade e curiosidade de mexer e aprender assuntos relacionados a computação. Achava, inicialmente, que Ciência da Computação era o caminho

lógico para desenvolver essa curiosidade. Porém, escolhi, inicialmente, o curso de Engenharia de Software por uma mera facilidade de aprovação no mesmo.

13. Não sei.
14. Escolhi o curso por gostar da área de computação e pelo campus ser na cidade onde moro.
15. Influência dos pais e indecisão com o que eu gostaria de fazer
16. Vontade de se formar nessa área
17. Visão imatura sobre o mercado de trabalho, influência de parentes e colegas, proximidade à cidade de residência (Gama)
18. O meu pensamento que justificava a entrada no curso era de implementar ideias inovadoras no mercado. A Engenharia de Software me daria as ferramentas necessárias pra isso. Entretanto, faltou “paixão” pelo curso pra terminá-lo.
19. Gosto e já trabalho na área
20. Pois é uma área de interesse com uma boa expectativa financeira.
21. Estudar fora de casa
22. Afinidade com programação
23. Afinidade e interesse com computador e computação em geral.
24. PAS
25. Sempre gostei da área de computação e tive curiosidade de aprender a elaborar websites. Então, quando tive que decidir que curso faria optei pelo que acreditei que seria mais prático do que teórico.
26. Não me vejo trabalhando com algo que não seja tecnologia, mas por ter uma tendência auto-didata acredito que estaria aprendendo muito mais em um curso de Administração, Direito ou Comunicação Social, enquanto mantenho meus estudos relacionados à tecnologia por conta própria, do que em um curso de Computação. Já pensei seriamente em trocar algumas vezes e cheguei a ser aprovado no vestibular para Comunicação, mas foi no mesmo semestre em que fui contemplado pelo Ciência sem Fronteiras e preferi ir estudar Computação fora, hoje não troco mais por estar perto da formatura mas planejo iniciar uma graduação em Direito assim que terminar Engenharia de Software.
27. Facilidade com matérias de exatas

28. Era muito jovem e acreditava que computação seria algo com o qual iria me identificar
29. Afinidade com a área
30. Afinidade com os conteúdos e interesse na área de trabalho.
31. Por que amo computação
32. Afinidade com a área de informática
33. Pela afinidade que tinha com matemática e por acreditar que as matérias seriam um pouco diferentes do que foram
34. Achei que gostava e tinha habilidade com matemática
35. Gosto e ainda trabalho na área

B.3 Disciplinas Consideradas mais Difíceis

A seguir são apresentadas as disciplinas consideradas mais difíceis pelos alunos. Cada item corresponde à resposta de um aluno entrevistado.

B.3.1 Ciência da Computação

1. Física
2. Circuitos digitais
3. Não considero nenhuma das disciplinas que cursei complexas, levando em conta que a disciplina mais avançada que cursei foi LP. Porém a didática da maioria das disciplinas é muito ruim quando se considera a quantidade de recursos e tecnologias que poderiam ser empregadas atualmente na lição do conteúdo.
4. Teoria da computação e álgebra 1 foram as duas disciplinas mais complexas.
5. Calculo
6. Estrutura de Dados
7. Álgebra 1
8. Álgebra 1 - disciplina fascinante e muito importante para o curso, mas abstrata.
9. Programação Sistemática, pois a aula não ocorreu numa dinâmica muito boa.
10. Físicas, Cálculos, Computação Básica e Físicas Experimentais

11. Compiladores
12. Física
13. Essa pressão que a UnB tem do desligamento é muito ruim, só existe na UnB e não acrescenta em nada.
14. Física 2
15. Autômatos e computabilidade
16. As matérias de física 1, 2 e 3 e experimentais, não se tratam das disciplinas de compreensão mais difícil, mas sim de matérias nas quais não entendi a aplicação no curso. Acho que uma matéria de física 'computacional' com conceitos mais ligados a área de computação seria mais útil.
17. Cálculo 2 e 3 (acho desnecessário para o curso)
18. Cálculos
19. Como só cursei 1 semestre de CiC, a matéria mais difícil pra mim nesse período foi física 1, mas que ainda é relativamente fácil.
20. Cálculo
21. A dificuldade não é a matéria, qualquer coisa se aprende se dada com qualidade. Pra exemplificar, Lógica Computacional não é tão difícil, mas os dois professores da matéria são inadequados para serem professores. Um não fala idioma nenhum e outro se preocupa em manter pose e não consegue resolver os exercícios nem tirar dúvidas. Outra disciplina é a de linguagem de programação onde a principal turma é dada por um machista retrógrado que marca os alunos que discordam dele e os reprova por puro prazer.
22. Calculo
23. Cálculo 1
24. Física
25. Teoria dos Números
26. Análise de Algoritmos
27. Análise de algoritmos
28. Cálculo
29. fisica

30. Computação básica
31. Cálculo
32. Cálculo 3 e álgebra linear
33. Circuitos Digitais e OAC
34. Física 3
35. Cálculo e Física
36. Não cheguei a cursar qualquer disciplina
37. não me recordo
38. Computação Básica
39. Física 3
40. Cálculo
41. Cálculo
42. Computação Básica
43. Computação Basica, o passo do curso segue num ritmo desmotivador, seria interessante deixar ja abertos os trabalhos finais para que os alunos interessados pudessem inicia-los previamente, para evitar conflitos com outras matérias.
44. Estrutura de Dados foi a mais complexa que cursei, haja vista que não há um bom professor dessa matéria no curso.
45. As matérias de matemática e física em geral são bem difíceis, exigindo do aluno muitos conhecimentos da época do ensino médio, e que nem sempre são tratados com profundidade nas escolas.
46. Cálculo 3
47. Lógica Computacional e Álgebra 1
48. Autômatos
49. OAC
50. Até onde fui não achei nenhuma matéria de outro mundo.
51. Organização e arquitetura de computadores
52. Física Experimental 1

53. Cálculo 3
54. Lógica de Programação
55. Física
56. Organização e Arquitetura de Computadores.
57. Estrutura de dados. Fiz em java. Tinha apenas três alunos matriculados na turma.
As aulas eram na sala do professor.
58. Física
59. Estatística
60. Não houve matérias consideradas complexas, sendo as dificuldades devidas à procrastinação
61. Estrutura de Dados, Teoria dos Números
62. Tradutores
63. Não se aplica
64. Computação básica
65. Estrutura de dados e algoritmos
66. Calculo 3
67. Probabilidade e estatística
68. Circuitos Digitais, Analises de Algoritmos
69. Calculo 1
70. Disciplinas matemáticas
71. Organização e Arquitetura de Computadores.
72. As físicas
73. Álgebra 1 e álgebra 2
74. Sem resposta
75. Cálculo 3
76. Programação
77. Cálculo.

78. Software Básico
79. Cálculo 3
80. Organização de arquivos e banco de dados
81. Física. O conteúdo era apenas o do Ensino Médio, mais aprofundado; mas os modos de avaliação e de ensino pareciam ter sido criados para dificultar a aprendizagem do aluno.
82. Algebra Linear
83. Só fiz 1 semestre, achei todas de boa, mas a mais complicada foi Álgebra 1.
84. Física
85. Sistemas digitais
86. Física 1, 2 e 3
87. Estrutura de dados
88. Matemática
89. OAC
90. Eu não cheguei à alcançar esta disciplina na grade, mas com um pouco de leitura prévia a disciplina que é mais complexa é a que lida com linguagens formais e autômatos.
91. Calculo
92. Cálculo no geral, o curso é muito voltado pra área de matemática.
93. Física

B.3.2 Licenciatura em Computação

1. Cálculo
2. Calculo 1
3. Calculo 1 Materia muito dificil, principalmente para alunos vindos de escolas publicas, professores que nao se comunicam corretamente com os alunos, proprio professor deixa claro que os alunos tem que ser auto didatas, nao ha nenhum tipo de estímulo para o enterece do aluno com a materia.
4. Análise de algoritmos

5. Teoria da Computação
6. Programação e Lógica. Os professores abordam conteúdos como se já soubéssemos da base, algumas pessoas já eram programadores mas metade de minha turma entrou pelo PAS, ou seja, recém saídos do ensino médio, não teve nenhum tipo de nivelamento.
7. Lógica Computacional foi uma disciplina péssima porque foi dada no primeiro semestre por uma professora totalmente sem paciência com os alunos, e o grau de dificuldade que ela colocou foi muito desestimulante.
8. Cálculo e as do eixo de matemática
9. Álgebra 1
10. Lógica
11. Lógica Computacional
12. Calculo
13. Arquitetura de Processamento de Dados (APD)
14. Cálculo 1 e Lógica Computacional
15. Análise de algoritmos
16. Cálculo 1
17. Algebra
18. Como já faz bastante tempo que desisti do curso, não lembro o nome, tampouco o assunto da disciplina; só lembro que tinha que usar o LaTeX.
19. Lógica Computacional 1, e horrível, muito difícil de entender professores não ajudam.
20. Cálculo
21. Linguagens de programação.
22. Cálculo, pois não foi definido um único livro utilizado na base da ementa disciplinar, além disso o mais usado, do Thomas, apenas em Inglês e média harmônica para base da nota final acredito que não é a forma mais justa.
23. Estruturas de Dados
24. Lógica e Cálculo

25. Álgebra 1, sem dúvidas é uma grande barreira.
26. Lógica Computacional
27. Grade curricular confusa, com pouco foco em Computação
28. Probabilidade e estatística. Computação Básica. Cálculo.
29. Lógica
30. Mudei de faculdade antes do início desse curso.
31. programação. dificultou demais o professor ser mestrando e muito rígido.
32. Software Básico
33. Cálculo
34. Probabilidade e estatística
35. Teoria da Computação
36. Análise de Algoritmos
37. Cálculo 1 e Probabilidade e Estatística.
38. Lógica Computacional
39. Cálculo 1
40. Cálculo
41. Probabilidade e estatística
42. Cálculo
43. Álgebra
44. Introdução a Microinformática (Obs: a professora não ensinou absolutamente nada. Apenas disse os temas dos trabalhos em grupos e deixou por conta de cada um. Suporte zero.)
45. Estrutura de dados, porém não cursei muitas disciplinas que sei que são mais difíceis.
46. Não se aplica.
47. Lógica computacional e computação básica
48. Desliguei-me antes de cursar matéria.
49. Estrutura de dados e estatística, mais ao rigor das avaliações destas e outras disciplinas

50. Álgebra 1
51. Lógica
52. Lógica Computacional
53. Abandonei no segundo período. Não tinha dificuldades em relação à disciplinas
54. Uma relacionada à psicologia
55. Engenharia de software.
56. Calculo 2
57. Álgebra
58. Álgebra 1 (a exigência da disciplina não se justifica quando se estuda Teoria da Computação)
59. Matemática
60. OAC
61. Matéria sobre Lógica
62. Algoritmos
63. Álgebra / Teoria da computação
64. Lógica
65. Cálculo I e II e Lógica
66. Não lembro o nome, era a primeira disciplina que tratava diretamente de programação, no segundo semestre do curso, o professor resolveu ensinar phyton uma linguagem de programação de difícil compreensão e com pouquíssimas referencias para estudo, nem os monitores sabiam direito a matéria.
67. Cálculo.
68. Não me lembro de alguma disciplina específica mas a falta de prática com a matemática do ensino médio (tenho 41 anos e faz tempo que estudei os detalhes matemática) me fez temer as disciplinas de cálculo. Acho que se tivesse tempo para rever a matemática do ensino médio não teria problemas. De qualquer maneira o ano em que abandonei o curso foi um pouco atribulado para mim e pensei em terminar primeiro um outro curso que tinha começado (Segurança de TI na Católica Virtual-UCB) antes de pensar a voltar ao curso na UNB. Talvez um dia eu tente novamente.
69. Lógica computacional

70. Não cheguei a cursar nenhuma aula. Desisti para estudar para o vestibular.
71. Análise de Algoritmos com o Claus.
72. Teoria da Computação
73. Programação orientada a objetos
74. As disciplinas de calculo I e II. E a elaboração do Projeto Final.
75. Tinha um professor que baseava todo o curso dele em uma plataforma proprietária de aula a distância, tipo um joomla off-line, ele distribuía os disquetes com as aulas no início do semestre e ficava executando esse programa durante as aulas. Era terrivelmente chato, a matéria era sobre algoritmos, não era particularmente difícil, acho que era feita antes das matérias de linguagem de programação, tentei fazer umas 2x antes de sair. Outra matéria que achei insuportável era psicologia da educação, que deveria se chamar história da psicologia da educação no século XX, também não era difícil mas tentei fazer por 3 ou 4x sem sucesso, simplesmente não conseguia ficar assistindo às aulas, ficar aprendendo teorias que já tinham sido consideradas falhas ou pior por estudos mais recentes não é coisa que consiga me manter acordado.
76. Cálculo
77. Lógica Computacional
78. Álgebra 1
79. Teoria da Computação
80. Lógica Computacional
81. Lógica Computacional
82. Lógica Computacional
83. Linguagem de programação
84. Teoria da Computação
85. Lógica computacional

B.3.3 Engenharia de Computação

1. Nenhuma
2. Software Básico

3. Não lembro de nenhuma disciplina muito complexa.
4. Programação orientada a objeto
5. Física
6. Equações Diferenciais
7. Linguagens de Programação
8. Organizacao e Arquitetura de Computadores
9. OAC
10. Físicas
11. Estrutura de Dados
12. Introdução a Álgebra Linear
13. Cálculo
14. Calculo e materias mais especificas de programacao
15. A maioria das matérias de matemática, principalmente por serem puramente teóricas sem nenhum exemplo prático de uso das mesmas.
16. Organização e arquitetura de computadores, linguagens de programação, sinais e sistemas, projeto e análise de algoritmos
17. Computação básica
18. OAC
19. IAL, Estrutura de Dados
20. Eletromagnetismo.
21. Física, Calculo.
22. Métodos de Programação
23. Cálculo 1
24. Física 1
25. Sistemas digitais
26. Organização e Arquitetura de Computadores.
27. Sinais e Sistemas ou Eletromagnetismo 1

28. Autômatos
29. Nenhuma
30. Cálculo, física, computação básica e algoritmo e estrutura de dados.
31. circuitos elétricos 2
32. Não achei nenhuma disciplina difícil, eu era um bom aluno, mesmo não querendo cursar o curso, eu tirava boas notas.
33. Computação Básica
34. Introdução à Álgebra Linear
35. Nenhuma foi realmente um desafio
36. Programação Orientada a Objetos
37. Física 1. Eu já fiz até OAC e achei tranquila, mas é simplesmente ridículo provas com fator de correção, lugar marcado, feitas no estilo vestibular para induzir ao erro. Onde atualmente requer mais tempo que cálculo 1 e álgebra linear somadas.
38. Estrutura de dados
39. Computação Básica
40. Cálculo, Álgebra
41. Introdução a Geometria Analítica
42. Eletromagnetismo
43. Física 1
44. OAC
45. Cálculo 3, Arquitetura de Computadores,
46. Programação Orientada Objeto - POO
47. Cálculo
48. Física
49. IDJ
50. Física
51. Física 1.
52. Organização e arquitetura de computadores

53. Qualquer materia pertencentes ao departamento de matematica. Pela falta de organização e criterio fixo. A matéria não era difícil. Difícil era saber o que os professores queriam.

B.3.4 Engenharia de Software

1. Cálculo
2. Álgebra Linear, Calculo e Química (essa ultima considerava um absurdo a cursas, em Engenharia de Software, só pelo fato de ser um curso de engenharia)
3. Física classica. Não entendo o porque de engenharia de software ter essa disciplina como obrigatória. Em nenhum outro lugar que tenha esse curso essa disciplina faz parte da grade.
4. Física(s)
5. Tcc, no curso de engenharia de software na unb do gama não existe disciplina na grade curricular obrigatória que ofereça bom embasamento na metodologia científica como em outros cursos. Sendo assim no geral muitos têm dificuldade nessa disciplina, sendo que nem sempre os orientadores dão essa base ou realmente participam ativamente.
6. Orientação a Objetos
7. Física
8. Quimica, na primeira vez que entrei era matéria obrigatória do primeiro semestre do curso de engenharia de software
9. Estruturas matemática para computação (emc)
10. Não lembro o nome, mas o professor não era dos melhores.
11. Nenhuma.
12. Das disciplinas cursadas, enquanto aluno de Engenharia de Software, nenhuma disciplina apresentava um conteúdo muito complexo. O meu problema era não ver (e continuo sem ver) a importância de cursar Cálculo 1-3, Física 1-3, Física Experimental 1-3, entre outras disciplinas de matemática. Por não ver a importância, não conseguia ter paciência para estudar as disciplinas e tive várias reprovações nas mesmas, o que gerou uma frustração com o curso todo.
13. Programação.
14. Física e Estrutura de Dados e Algoritmos

15. Química, sistemas digitais
16. Não havia maior complexidade relevante
17. Introdução a ciência da computação. Tive um professor nada didático e, considerando que era algo completamente novo para quem não teve contato com computação antes, foi uma experiência nada prazerosa.
18. Até o 6º período, em relação ao fluxograma do curso na época (2014), a disciplina mais difícil era Métodos de Desenvolvimento de Software. Não é a mais difícil do curso. Só posso afirmar que era a mais difícil até o 6º período.
19. Cálculo, Química
20. Sistemas digitais
21. GPP- gestão de projeto e portfólio
22. Física
23. Estruturas de dados e algoritmos
24. Estrutura de dados por um professor que faltou 2 meses e ainda deu zero numa prova que tinha acertado
25. Os cursos de programação, creio que se as turmas fossem menores e incentivassem o estudo em grupo (como agora o fazem nas turmas de física 1 unificada e cálculo 1 na FGA) ficaria mais fácil de assimilar todo o conteúdo.
26. Nunca tive nenhum problema grave com alguma disciplina que mereça ser mencionado. As que reprovei foi por falta de dedicação minha ou não adequação aos métodos dos professores, não por alta complexidade.
27. Engenharia e Ambiente
28. Cálculo 2
29. Tronco básico do curso de engenharias
30. As disciplinas de documentação que cobravam programas em linguagens ainda não ensinadas no curso.
31. Química Geral Teórica
32. Matemática e Programação
33. Acredito que a matéria de ICC seja a mais complexa, para quem está acabando de entrar em um curso de computação e não conhece nem entende nada de programação. As minhas experiências foram péssimas, pois, ao meu ver, os professores não

estavam muito interessados em ensinar do 0 para quem não sabia nada e sim tentar ensinar algo mais avançado para os que já sabiam.

34. Cálculo

35. N/A

Apêndice C

Artigo Apresentado na 4^a CIACA -
Lisboa, Portugal

ANÁLISE DE EVASÃO NOS CURSOS SUPERIORES DE COMPUTAÇÃO: UMA ABORDAGEM USANDO ANÁLISE DE SOBREVIVÊNCIA E ALGORITMO APRIORI

Raphael Magalhães Hoed

Instituto Federal de Educação, Ciência e Tecnologia do Norte de Minas Gerais
raphael.hoed@ifnmg.edu.br

Marcelo Ladeira

Universidade de Brasília
mladeira@unb.br

RESUMO

A técnica estatística de Análise de Sobrevivência e a mineração de regras de associação via algoritmo *Apriori* foram aplicadas neste artigo, usando registros dos alunos dos cursos de Bacharelado em Ciência da Computação, Licenciatura em Computação, Engenharia de Computação e Engenharia de Software da Universidade de Brasília (UnB). Esse artigo verifica o comportamento da evasão nos cursos de Computação ao longo dos períodos cursados, a influência que o gênero e a reprovação na disciplina de Cálculo 1 e em disciplinas iniciais de Algoritmos exerce sobre a evasão. A partir do emprego de técnicas estatísticas e de mineração de regras de associação, espera-se obter maior entendimento do fenômeno da evasão nos cursos superiores de Computação.

PALAVRAS-CHAVE

Evasão, Computação, Mineração, Sobrevivência.

1. INTRODUÇÃO

A evasão nos cursos superiores de Computação é um problema que atinge o Brasil e vários outros países, conforme será mostrado na seção 2 desse trabalho. É sabido que é dever das instituições de ensino prover condições para a permanência dos discentes no curso. Tanto nas instituições privadas quanto nas públicas a evasão leva à perda de receitas. A melhor compreensão do fenômeno de evasão pode auxiliar na adoção de medidas mais eficazes no sentido de mitigar o problema de evasão.

O objetivo geral desta pesquisa é analisar fatores relacionados à evasão nos cursos superiores de Computação a partir de dados fornecidos pela secretaria de registros acadêmicos da Universidade de Brasília (UnB) - Brasil. Nesse estudo considera-se evadido aquele aluno que é desligado do curso antes da sua conclusão por qualquer motivo (exceto por falecimento). A seguinte hipótese será investigada nessa pesquisa: os alunos que reprovam em disciplinas iniciais de Cálculo e Algoritmos tem maior predisposição a evadirem.

São utilizadas nesse trabalho técnicas estatísticas de Análise de Sobrevivência, a partir da função de sobrevivência de *Kaplan-Meier* e mineração de dados usando o algoritmo *Apriori* para minerar regras de associação, analisando-se os seguintes cursos superiores de Computação ofertados pela UnB: Bacharelado em Ciência da Computação, Licenciatura em Computação, Engenharia de Computação e Engenharia de Software. O intuito é verificar a sobrevivência dos alunos considerando-se o evento evasão, e levando em conta os seguintes fatores: gênero, reprovação em disciplinas iniciais de Matemática e reprovação em disciplinas iniciais de algoritmo. Foram considerados no estudo os ingressos nos cursos citados no período 2005 a 2015. Também é verificado a ocorrência de regras de associação entre as variáveis descritas, levando-se em conta também o período em que o aluno evadiu.

Este artigo segue a seguinte organização: A seção 2 mostra a fundamentação teórica, explorando o problema de evasão e discorrendo sobre as técnicas estatísticas e de mineração de dados empregadas; A seção 3 mostra a metodologia empregada no estudo; A seção 4 mostra os estudos realizados e os resultados obtidos. A seção 5 exhibe as conclusões obtidas e sugestões para trabalhos futuros.

2. FUNDAMENTAÇÃO TEÓRICA

Esta seção está dividida em duas subseções: a subseção 2.1 trata do conceito de evasão e do problema de evasão nos cursos superiores de Computação e a subseção 2.2 discorre sobre a técnica estatística de Análise de Sobrevivência e o algoritmo *A priori* para mineração de regras de associação.

2.1 O Problema da Evasão em Cursos de Computação

Antes de discorrer sobre evasão nos cursos de Computação é importante conceituar a evasão. O conceito de evasão pode mudar dependendo da abordagem e do contexto. Em estudo feito pela comissão especial de estudos sobre a evasão nas universidades públicas brasileiras (SESu/MEC - ANDIFES – ABRUEM) (MEC, 1997) são apresentados os conceitos de evasão de curso (desligamento apenas do curso, sem desvinculo com a instituição); evasão da instituição (desligamento da instituição na qual está matriculado); evasão do sistema (abandono do ensino superior). Em relação à adoção do conceito mais apropriado a respectiva comissão afirma que isso deve ser feito em função do objetivo pretendido. Nesse estudo será contemplada a evasão de curso, já que interessa, nesse caso, qualquer desvinculo do aluno com o curso de Computação, ainda que ele tenha permanecido na instituição.

A evasão em cursos de Computação é um problema em vários países. Foi verificado que os cursos da área de Ciência da Computação da Universidade de Tecnologia de Helsinki na Finlândia, tem de 500 a 600 inscrições anuais e as taxas de evasão tem variado entre 30 e 50% (Kinnunen e Malmi, 2006). Uma comparação feita entre a evasão nos cursos de graduação em Computação e Mestrado em Educação ofertados pela Hellenic Open University revelou que dos 1.230 alunos de graduação matriculados no curso de Informática entre 2000 e 2002, 349 desistiram (28,4%) e dos 1.220 alunos matriculados no curso da área de Educação, 173 evadiram (14,2%) (Pierrakeas et al., 2004). Segundo estudo realizado pela Higher Education Authority (HEA) (Mooney et al., 2010) na Irlanda, analisando-se os anos letivos de 2007/08 a 2008/09, verificou-se que os cursos de Ciência da Computação experimentaram a mais alta taxa de evasão, com 27% desses alunos saindo ainda no primeiro ano. Os cursos com a menor taxa de abandono foram medicina, com apenas 2%, seguido por direito, com 3%. Esse estudo também mostrou que os homens são ligeiramente mais propensos a evadir do que as mulheres, com 17% de taxa de evasão entre os homens em comparação com 13% de taxa de evasão entre as mulheres. Outro dado interessante do estudo é que os estudantes com baixos níveis de escolaridade em matemática são mais propensos a abandonar um curso de ciência ou tecnologia. Um estudo publicado também pela HEA em 2016 (Liston et al., 2016), analisando-se os anos letivos de 2012/13 a 2013/14 na Irlanda, mostrou resultados semelhantes ao estudo anterior, com as áreas de Construção e Afins, Serviços, Ciência da Computação e Engenharia com as maiores taxas de evasão e com mulheres apresentando menores taxas de evasão em todos os níveis e setores em relação aos homens. No Brasil, no período 2001-2005, a taxa de evasão nos cursos de Computação foi uma das mais altas do país com média de 32%, sendo que a evasão nacional teve média de 22% no período (Silva Filho et al., 2007).

Em relação às causas de evasão em cursos superiores da área de Computação, um estudo realizado na Hellenic Open University - Patras, Grécia elencou as principais causas de evasão em cursos superiores de Ciência da Computação para os anos letivos de 2000-2001 e 2001-2002: Profissional (62,1%), acadêmico (46,2%), familiar (17,8%), pessoais (8,9%), relacionado a saúde (9,5%) (Xenos et al., 2002). Um estudo realizado na Universidade AL-AQSA – Palestina envolvendo 1290 registros de alunos no período de 2005 a 2011 em cursos de Ciência da Computação, usando técnicas de mineração de dados, constatou que o êxito em disciplinas iniciais do curso relacionadas a algoritmos diminui a probabilidade de abandono do aluno (Abu-Oda e El-Halees 2015). Em relação ao uso da Matemática em cursos superiores de Computação, Gruner (Gruner, 2015) afirma que os estudantes de Ciência da Computação na África do Sul que pretendem tornar-se pesquisadores ou cientistas é atualmente baixa e a popularidade da Matemática abordada no

currículo dos cursos está em queda. Byrne e Lyons (2001 apud Stamouli et al., 2004) sugerem que existe uma relação entre as habilidades de programação e experiência em matemática. De acordo com Savage e Hawkes (2000, apud Stamouli et al., 2004), A preocupação com a queda no domínio de habilidades matemáticas básicas dos alunos levou várias universidades do Reino Unido a estabelecerem Centros de Apoio a Matemática para cursos como Matemática, Computação e Engenharia. As causas de evasão no ensino superior que aparecem com maior frequência no Brasil de acordo com a produção acadêmica sobre o assunto são: falta de condições financeiras, influência familiar, questão vocacional, reprovações em disciplinas que requerem conhecimento matemático, qualidade do curso, localização da instituição, trabalho, idade do aluno com evasão mais alta verificada entre os de maior idade (Souza et al., 2012). Verifica-se que a questão vinculada à aprendizagem em matemática figura entre as principais causas de evasão em cursos superiores.

2.2 Análise de Sobrevivência e Mineração de Regras de Associação

A análise de sobrevivência tem um grande número de aplicações que não se restringem apenas à área da saúde. De acordo com Oliveira e Lins (Oliveira e Lins, 2011) a análise de sobrevivência “é aplicável quando se pretende analisar dados ao longo do tempo, buscando, entre outras informações, o tempo de ocorrência de um dado evento de interesse”. No presente estudo, o evento de interesse se trata da evasão de alunos.

Um outro aspecto importante da análise de sobrevivência é a presença de dados censurados. De acordo com Oliveira e Lins (Oliveira e Lins, 2011), “a análise de sobrevivência possui como característica fundamental a existência de censuras, que são basicamente observações parciais da resposta”. No contexto da análise realizada nesse trabalho, o falecimento de um aluno durante o período avaliado, ou um aluno que continua na instituição findado o período de observação, seriam exemplos de censura, visto que o evento de interesse é a evasão.

Segundo Colosimo e Giolo (2006 apud Oliveira e Lins, 2011), podemos ter os seguintes tipos de censura: Censura à direita do tipo I: O estudo se encerra após um tempo, mesmo que o evento terminal não ocorra; Censura à direita do tipo II: O estudo se encerra após obter um número pré-definido de observações do evento terminal; Censura aleatória: um participante é removido do estudo sem que tenha sido observado a ocorrência do evento terminal. No contexto desse estudo, tem-se como evento terminal a evasão dos alunos.

A curva de Sobrevivência é definida como a probabilidade do evento em estudo não falhar até o tempo t , ou seja, a chance do evento terminal ser observado após o tempo t . Isto é descrito como a equação (1), sendo uma função sempre monotônica decrescente (Oliveira e Lins, 2011):

$$S(t) = P[T > t] \quad (1)$$

De acordo com Oliveira e Lins (Oliveira e Lins, 2011), “a análise de sobrevivência é a técnica ideal para analisar respostas binárias (ter ou não ter um evento) relacionadas aos tempos dos indivíduos”.

As duas principais técnicas de análise de sobrevivência empregadas em estudos são diferenciadas da seguinte forma: O método Atuarial divide o tempo em estudo em intervalos equivalentes, estimando a probabilidade que os participantes sobreviventes ao início de cada intervalo tem de sofrer o evento terminal até final desse intervalo. O método de *Kaplan-Meier* divide o tempo em um seguimento de intervalos, e os limites correspondem ao tempo de seguimento em que houve eventos. A técnica de *Kaplan-Meier* usa a data exata do evento terminal, tornando o resultado mais preciso (Botelho et al., 2009).

Existem vários testes que podem ser usados para verificar a significância estatística em estudos de análise de sobrevivência. O software *SPSS Statistics*, em suas versões mais recentes, disponibiliza os seguintes testes estatísticos para análise de sobrevivência: *Logrank*, *Breslow* e *Tarone-Ware*. Em relação a estes testes, pode-se dizer o seguinte:

De maneira geral, esses testes consistem em calcular uma estatística ponderada que, sob a hipótese nula (hipótese de que os grupos de medidas são amostras originadas da mesma população), tem distribuição conhecida. Entretanto os testes possuem regras de atribuição de peso para melhor discriminar as curvas em determinadas etapas de sua evolução temporal. No teste *Logrank*, os pesos são atribuídos de maneira a enfatizar diferenças ao final das distribuições de tempo de vida. O teste de *Breslow* enfatiza diferenças no início da distribuição do tempo de vida. O teste de *Tarone-Ware*, por sua vez, foi desenvolvido com o objetivo de discriminar distribuições em fases intermediárias do tempo de vida (Lima Júnior et al., 2012).

Além da técnica de *Kaplan-Meier* será usado nesse trabalho também o algoritmo *Apriori* para mineração de regras de associação, para descobrir associações importantes entre os dados e validar as ocorrências verificadas via análise de sobrevivência. A descoberta de regras de associação pode ser decomposta em duas etapas, de acordo com Agrawal et al. (1993 apud Romão et al., 1999): localizar os conjuntos de itens (*itemsets*) que apresentam suporte superior ao mínimo definido à partida; utilizar os *itemsets* obtidos na etapa 1 para gerar as regras de associação do banco de dados. Algumas definições importantes sobre mineração de regras de associação: “A toda regra de associação $A \rightarrow B$ associamos um grau de confiança, denotado por $conf(A \rightarrow B)$ ” (de Amo, 2004). O grau de confiança seria a probabilidade de que uma transação que tenha um item, também contenha o outro item. A equação (2) a seguir, formaliza essa definição (de Amo, 2004):

$$conf(A \rightarrow B) = \frac{\text{número de transações que suportam } (A \cup B)}{\text{número de transações que suportam } A} \quad (2)$$

Uma outra definição importante seria que “a toda regra de associação $A \rightarrow B$ associamos um suporte, denotado por $sup(A \rightarrow B)$ definido como sendo o suporte do *itemset* $A \cup B$ ” (de Amo, 2004). O suporte seria a proporção de transações que contém os itens. Ao definir um grau mínimo de confiança e um grau mínimo de suporte, uma regra de associação interessante seria então aquela que possui um suporte igual ou superior ao mínimo definido e aquela que possui uma confiança igual ou superior ao mínimo definido.

As fases de execução do algoritmo *Apriori* compreendem geração, poda, validação (de Amo, 2004). Resumidamente, sem entrar em detalhes sobre cada fase, na fase de geração são gerados os *itemsets* que tenham alguma chance de serem frequentes, na fase de poda são descartados os *itemsets* sem chances de serem frequentes, e na última é calculado o suporte de cada um dos *itemsets* do conjunto (de Amo, 2004). O funcionamento do algoritmo *Apriori* é descrito da seguinte forma:

Na primeira passagem, o suporte para cada item individual (conjuntos-de-1-item) é contado e todos aqueles que satisfazem o suporte_mínimo são selecionados, constituindo-se os conjuntos-de-1-item frequentes (F_1).

Na segunda iteração, conjuntos-de-2-itens candidatos são gerados pela junção dos conjuntos-de-1-item (a junção é feita através da função *apriori-gen*) e seus suportes são determinados pela pesquisa no banco de dados, sendo, assim, encontrados os conjuntos-de-2-itens frequentes. O algoritmo *Apriori* prossegue iterativamente, até que o conjunto-de-k-itens encontrado seja um conjunto vazio. (de Vasconcelos e de Carvalho, 2004)

3. METODOLOGIA UTILIZADA

Neste trabalho foi usado uma técnica estatística (análise de sobrevivência) em associação com a mineração de regras de associação (usando algoritmo *Apriori*). Considerou-se o processo de *KDD - Knowledge Discovery in Database* (Descoberta de Conhecimentos em Banco de Dados). Utilizou-se neste trabalho o processo de KDD em três grandes etapas de acordo com Silva e Vieira (2002 apud Pimentel e Omar, 2006) e Rezende et al. (2003 apud Pimentel e Omar, 2006): Preparação (seleção dos dados, pré-processamento e limpeza, transformação dos dados); Extração de Padrões (aplicação dos algoritmos de mineração de dados) e Pós-Processamento (avaliação dos padrões encontrados e apresentação do conhecimento).

Os estudos apresentados na seção 4 foram construídos a partir de dados fornecidos pela UnB. Antes de aplicar as técnicas de Análise de Sobrevivência e o algoritmo *Apriori* foi feita a limpeza dos dados na base de dados. As seguintes variáveis foram mantidas nesse trabalho: Número de matrícula do aluno, situação do aluno (se evadido ou não), tempo de permanência no curso (em número de períodos cursados), Gênero (masculino ou feminino), Reprovação na disciplina Cálculo 1 (sim ou não), Reprovação em disciplinas iniciais de algoritmo/lógica (sim ou não). A reprovação na disciplina de Cálculo 1 foi avaliada pois esta é uma disciplina geralmente estudada no início do curso e deseja-se analisar se isso tem impacto significativo sobre a evasão. O mesmo foi feito com as disciplinas de lógica/algoritmo, que costumam ser estudadas antes de se aprofundar o estudo sobre linguagens e paradigmas de programação. Foram considerados nesse estudo todos os alunos ingressantes no período de 2005 a 2015 nos cursos de Bacharelado em Ciência da Computação, Licenciatura em Computação, Engenharia de Computação e Engenharia de Software.

Considera-se, nas curvas de sobrevivência apresentadas, o p-valor (nível de significância) $< 0,05$ aceitável para avaliar a significância dos resultados obtidos. A abordagem realizada foi a seguinte: primeiro foi utilizado o estimador de *Kaplan-Meier* e apresentadas as curvas de sobrevivência para as variáveis anteriormente descritas. Os testes estatísticos aplicados às curvas de sobrevivência geradas pelo estimador de *Kaplan-Meier* foram o *LogRank*, *Breslow* e *Tarone-Ware*. O software utilizado para efetuar as análises de sobrevivência foi o *SPSS Statistics*. Em seguida foram apresentados os resultados da aplicação da mineração de regras de associação por meio do Algoritmo *Apriori*. Utilizou-se um suporte mínimo de 0,30 e a confiança de 0,60. Devemos considerar nesse caso, que se um suporte mínimo muito alto é definido, a busca poderia perder algumas associações interessantes que ocorrem envolvendo conjuntos menos frequentes mas ainda assim importantes. As variáveis usadas nessa etapa foram: Situação do aluno, Reprovação na disciplina Cálculo 1, Reprovação em disciplinas iniciais de algoritmo/lógica. Outras três novas variáveis foram acrescentadas nessa etapa, em decorrência dos estudos realizados com as curvas de *Kaplan-Meier*, conforme será detalhado na Seção 4: O aluno cursou de 1 a 4 períodos (sim ou não), o aluno cursou de 5 a 10 períodos (sim ou não), o aluno cursou 11 ou mais períodos. Nesse estudo, nem todas as regras obtidas via algoritmo *Apriori* foram exibidas, já que algumas não acrescentam em nada informações relevantes sobre o assunto. O software utilizado para aplicação do algoritmo *Apriori* foi o R.

Foram considerados nesse estudo um total de 2972 registros de alunos de Computação. A sobrevivência dos alunos no curso é considerada em relação ao evento evasão, sendo que nesse estudo, qualquer desligamento do curso, seja espontâneo ou não (exceto por motivo de falecimento) incluindo transferências (ainda que o aluno permaneça na instituição fazendo outro curso) é considerado evasão. O tempo de sobrevivência do aluno no curso é apresentado em períodos cursados (sendo que cada período tem a duração de um semestre).

4. ESTUDOS REALIZADOS

As figuras 1, 2 e 3 a seguir mostram a aplicação da técnica de *Kaplan-Meier* para a construção dos gráficos de sobrevivência, usando o software *SPSS Statistics*. Todos os testes estatísticos apresentados na análise de sobrevivência mostraram diferenças significativas entre as curvas (p-valor $< 0,05$).

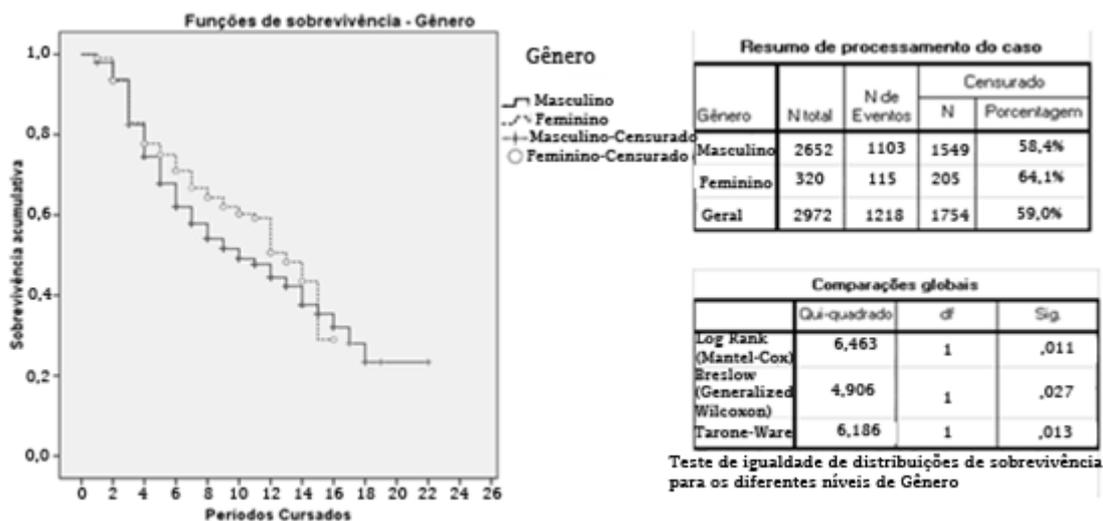


FIGURA 1 – CURVAS DE SOBREVIVÊNCIA DE KAPLAN-MEIER PARA GÊNEROS

A figura 1 mostra as curvas de sobrevivência para os gêneros masculino e feminino. A curva correspondente ao sexo feminino encontra-se acima da curva do gênero masculino para a maioria dos períodos avaliados, indicando uma sobrevivência superior quanto ao evento evasão. Os 3 testes estatísticos empregados revelam que há diferenças estatísticas significativas entre a sobrevivência apresentada entre os

gêneros. A evasão apresentada nos gêneros masculino e feminino é praticamente a mesma até o terceiro período, quando a partir do quarto a evasão começa a se mostrar maior em meio ao público masculino. A evasão no sexo feminino cessa a partir do período 16, enquanto a evasão no sexo masculino cessa a partir do período 18. A decisão pela evasão é mais rápida no sexo masculino. No caso dos homens, no quarto período há 75% de sobrevivência enquanto para as mulheres a sobrevivência cai a 75% apenas a partir do quinto período. No décimo período verifica-se uma sobrevivência masculina de 49%, enquanto a sobrevivência feminina chega a 50% apenas no décimo segundo período. O que se verifica é que a evasão acontece mais rapidamente no início do curso (os quatro primeiros períodos) e a medida que os alunos vão avançando no curso tende a ocorrer mais lentamente.

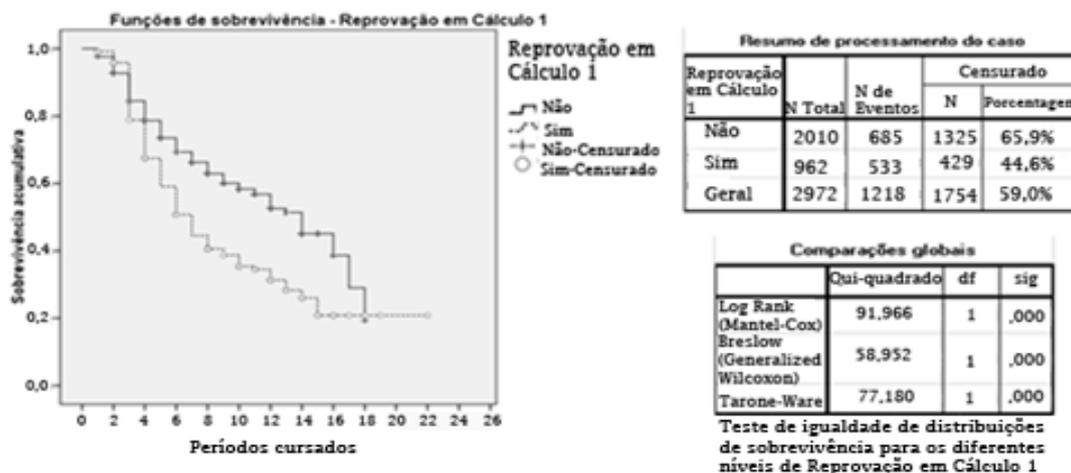


FIGURA 2 – CURVAS DE SOBREVIVÊNCIA DE KAPLAN-MEIER PARA REPROVAÇÃO EM CÁLCULO 1

As figuras 2 e 3 mostram, respectivamente, as curvas de sobrevivência para os reprovados e os não reprovados na disciplina Cálculo 1 e as curvas para os reprovados e os não reprovados em algoritmos. Conforme se percebe nos gráficos, os não reprovados em Cálculo 1 e em algoritmo apresentam uma sobrevivência ao evento evasão estatisticamente superior aos reprovados. A evasão em meio aos reprovados passa a ser maior, no caso da disciplina Cálculo 1 a partir do terceiro período e no caso das disciplinas de Algoritmo a partir do segundo.



FIGURA 3 – CURVAS DE SOBREVIVÊNCIA DE KAPLAN-MEIER PARA REPROVAÇÃO EM ALGORITMO

A tabela 1 a seguir mostra as saídas do software R ao aplicar o algoritmo *Apriori*. Conforme dito anteriormente, a evasão acontece mais rapidamente no início do curso (os quatro primeiros períodos) e tende a ocorrer mais lentamente depois, conforme verificado nas curvas de sobrevivência. Diante disso, para aplicação do algoritmo *Apriori*, 3 variáveis foram criadas, podendo assumir os valores sim ou não: Se o aluno cursou apenas de 1 a 4 períodos; se o aluno cursou de 5 a 10 períodos; se o aluno cursou 11 ou mais períodos. Seguem os significados de cada uma das variáveis usadas: EV = Evadido; CUR_1_A_4_P = Cursou de 1 a 4 períodos; CUR_5_A_10_P = Cursou de 5 a 10 períodos; CUR_11_MAIIS_P = Cursou 11 ou mais períodos; REP_ALG=Reprovado em Algoritmo; REP_CALC_1=Reprovado em Cálculo 1. As variáveis podem assumir o valor SIM ou NÃO. Significados das colunas: *Lhs* - *left hand side* (itens da esquerda); *Rhs* - *right hand side* (itens da direita); *Support* indica a proporção de transações que contém os itens; *Confidence* é a probabilidade de que uma transação que tenha um item *Lhs* também contenha o item *Rhs*. *Lift* mede o quão frequente *Lhs* e *Rhs* ocorrem juntos se comparados a se fossem estatisticamente independentes.

TABELA 1 – REGRAS DE ASSOCIAÇÃO OBTIDAS USANDO O ALGORITMO APRIORI

Num Regra	Lhs	Rhs	Support	Confidence	Lift
1	{EV=SIM}	{CUR_11_MAIIS_P=NÃO}	0,39	0,95	1,11
2	{CUR_1_A_4_P=NÃO}	{REP_CALC_1=NÃO}	0,39	0,70	1,03
3	{CUR_1_A_4_P=NÃO}	{REP_ALG=NÃO}	0,42	0,75	1,06
4	{REP_CALC_1=NÃO}	{EV=NÃO}	0,45	0,66	1,12
5	{EV=NÃO}	{REP_ALG=NÃO}	0,47	0,79	1,12
6	{REP_ALG=NÃO}	{EV=NÃO}	0,47	0,66	1,12
7	{REP_CALC_1=NÃO}	{REP_ALG=NÃO}	0,51	0,76	1,08
8	{REP_ALG=NÃO}	{REP_CALC_1=NÃO}	0,51	0,73	1,08
9	{CUR_11_MAIIS_P=NÃO}	{REP_ALG=NÃO}	0,59	0,69	0,98
10	{REP_CALC_1=NÃO, REP_ALG=NÃO}	{EV=NÃO}	0,37	0,73	1,24

Seguem os comentários sobre algumas regras encontradas: Regra 1 - 95% de quem evade não cursa 11 ou mais períodos, o que é compatível com o verificado nas curvas de sobrevivência pois a evasão ocorre mais rapidamente nos primeiros períodos; Regra 2 – 70% de quem não cursa apenas até o quarto período, também não reprova em Cálculo 1, indicando como o êxito na disciplina Cálculo 1 está associada a uma maior sobrevida no curso; Regra 3 – 75% de quem não cursa apenas até o quarto período, também não reprova em Algoritmo, indicando como o êxito nas disciplinas de Algoritmo está associado a uma maior sobrevida no curso; Regra 4 – 66% de quem não reprova em Cálculo 1 também não evade; Regra 6 – 66% de quem não reprova em Algoritmos também não evade; Regra 7 – 76% de quem não reprova em Cálculo 1 também não reprova em Algoritmo; Regra 8 – 73% de quem não reprova em Algoritmo também não reprova em Cálculo 1; Regra 10 – 73% dos não reprovados em Cálculo 1 e não reprovados em Algoritmos não evadem. Importante ressaltar que algumas regras foram omitidas na tabela 1, por não acrescentarem informação relevante ao estudo. Por exemplo: {CURSOU_1_A_4_P=SIM}=> {CURSOU_11_MAIIS_P=NÃO}. Essa regra indica que quem cursou apenas 4 períodos não cursou mais de 11 períodos. Essa é uma informação óbvia e irrelevante nesse estudo. Informações desse tipo foram descartadas.

5. CONCLUSÕES

O presente trabalho mostrou, utilizando técnicas estatísticas de análise de sobrevivência e mineração de regras de associação, fatores relacionados à evasão em cursos superiores de Computação. Encontrou-se evidências de que as mulheres demoram mais tempo para se decidirem pela evasão em relação aos homens. Deve-se destacar as evidências encontradas de que a reprovação em disciplinas iniciais do curso relacionadas à Matemática e Algoritmos aumenta a probabilidade do aluno evadir. Ressalta-se que há disciplinas nos demais períodos do curso, que dependem diretamente da disciplina Cálculo 1 e das disciplinas iniciais de

Algoritmo, seja por que estas disciplinas são pré-requisitos e, ainda que não sejam pré-requisitos, os conhecimentos de algoritmos e de Cálculo 1 são fundamentais para o sucesso de disciplinas como Cálculo 2 ou alguma disciplina relacionada a paradigmas de programação e técnicas de programação. De maneira geral, a evasão ocorre mais rapidamente nos períodos iniciais.

A evasão em cursos superiores é um tema complexo e não se explorou nesse trabalho todos os fatores relacionados a evasão nos cursos de Computação. No entanto, usando Análise de Sobrevivência associada ao algoritmo *Apriori*, obteve-se evidências de que dificuldades enfrentadas pelos alunos em disciplinas matemáticas e disciplinas de algoritmos contribuem para a evasão e de que a não reprovação em Cálculo 1 e Algoritmos aumenta a probabilidade do aluno não evadir. É necessário buscar estratégias que trabalhem as deficiências do discente já no primeiro semestre do curso.

REFERÊNCIAS

Conferências

- Abu-Oda, G. S., & El-Halees, A. M. (2015). Data Mining in Higher Education: University Student Dropout Case Study. *International Journal of Data Mining & Knowledge Management Process*, 5(1), 15.
- Botelho, F., Silva, C., & Cruz, F. (2009). Epidemiologia explicada – análise de sobrevivência. *Acta Urológica*, 26(4):33–38.
- De Amo, S. “Técnicas de mineração de dados”, XXIV Congresso da Sociedade Brasileira de Computação, vol. 1–1, Jul/Ago 2004, pp. 43
- De Vasconcelos, L. M. R., & De Carvalho, C. L. (2004). Aplicação de Regras de Associação para Mineração de Dados na Web. Instituto de Informática da Universidade Federal de Goiás.
- Gruner, S (2015). On the Future of Computer Science in South Africa: A Survey amongst Students at University. 44th Annual Southern African Computer Lecturers Association 2015 (SACLA 2015). Johannesburg - South Africa.
- Kinnunen, P., & Malmi, L. (2006, September). Why students drop out CS1 course?. In *Proceedings of the second international workshop on Computing education research* (pp. 97-108). ACM.
- Lima Júnior, P., Silveira, F. L. D., & Ostermann, F. (2012). Análise de sobrevivência aplicada ao estudo do fluxo escolar nos cursos de graduação em física: um exemplo de uma universidade brasileira. *Revista brasileira de ensino de física*. São Paulo. Vol. 34, n. 1 (jan./mar. 2012), 1403, 10 p.
- Liston, M., Frawley, D. & Patterson, V. (2016). A study of progression in Irish higher education. Higher Education Authority, Dublin.
- Ministério da Educação (MEC). Secretaria de Educação Superior., (1997). *Diplomação, retenção e evasão nos cursos de graduação em instituições de ensino superior públicas*. Brasília - DF: Comissão Especial de Estudos Sobre a Evasão nas Universidades Públicas Brasileiras., p.20.
- Mooney, O., Patterson, V., O’Connor, M., & Chantler, A. (2010). A study of progression in Irish higher education. Higher Education Authority, Dublin.
- Oliveira, C. d. S., & Lins, L. N. (2011). Identificação das causas da evasão e retenção prolongada de estudantes do ensino superior a partir de dados censurados. XLIII Simpósio Brasileiro de Pesquisa Operacional. Ubatuba - São Paulo.
- Pierrakeas, C., Xenos, M., Panagiotakopoulos, C., & Vergidis, D. (2004). A comparative study of dropout rates and causes for two different distance education courses. *The International Review of Research in Open and Distributed Learning*, 5(2).
- Pimentel, E. P., & Omar, N. (2006, January). Descobrimos conhecimentos em dados de avaliação da aprendizagem com técnicas de mineração de dados. In *Anais do Workshop de Informática na Escola (Vol. 1, No. 1)*.
- Romão, W., Niederauer, C. A., Martins, A., Tcholakian, A., Pacheco, R. C., & Barcia, R. M. (1999). Extração de regras de associação em C&T: O algoritmo Apriori. XIX Encontro Nacional em Engenharia de Produção.
- Silva Filho, R. L. L., Motejunas, P. R., Hipólito, O., & Lobo, M. B. C. M. (2007). A evasão no ensino superior brasileiro. *Cadernos de pesquisa*, 37(132), 641-659
- Souza, C., Petró, C., & Gessinger, R. (2012). Um estudo sobre evasão no ensino superior do Brasil nos últimos dez anos: as possíveis causas e fatores que influenciam no abandono. *Prevenção do risco do abandono*. In *Conferência Latinoamericana Sobre El Abandono Em La Educacion Superior—Clabes (Vol. 2)*.
- Stamouli, I., Doyle, E., & Huggard, M. (2004, October). Establishing structured support for programming students. In *Frontiers in Education, 2004. FIE 2004. 34th Annual* (pp. F2G-5). IEEE.
- Xenos, M., Pierrakeas, C., & Pintelas, P. (2002). A survey on student dropout rates and dropout causes concerning the students in the Course of Informatics of the Hellenic Open University. *Computers & Education*, 39(4), 361-377.