

UNIVERSIDADE FEDERAL DO PARANÁ

LUCIANA ALEXANDRE BARBOSA

POLÍTICAS AFIRMATIVAS - DESEMPENHO DO COTISTA E NÃO-COTISTA
NO CURSO DE CÁLCULO DIFERENCIAL E INTEGRAL I E II NAS
ENGENHARIAS CIVIL, ELÉTRICA E MECÂNICA COM ANÁLISE
UNIVARIADA E MULTIVARIADA DOS DADOS

CURITIBA

2011

LUCIANA ALEXANDRE BARBOSA

POLÍTICAS AFIRMATIVAS - DESEMPENHO DO COTISTA E NÃO-COTISTA
NO CURSO DE CÁLCULO DIFERENCIAL E INTEGRAL I E II NAS
ENGENHARIAS CIVIL, ELÉTRICA E MECÂNICA COM ANÁLISE
UNIVARIADA E MULTIVARIADA DOS DADOS

Dissertação apresentada ao Curso de Pós-Graduação em Métodos Numéricos em Engenharia, Área de Concentração em Programação Matemática do Departamento de Matemática, Setor de Ciências Exatas e do Departamento de Construção Civil, Setor de Tecnologia da Universidade Federal do Paraná, como parte das exigências para a obtenção do título de Mestre em Ciências.

Orientadora: Prof^a. Dr^a. Liliana Madalena Gramani

Co-Orientador: Prof. Dr. Anselmo Chaves Neto

CURITIBA

2011


TERMO DE APROVAÇÃO

LUCIANA ALEXANDRE BARBOSA


POLÍTICAS AFIRMATIVAS - DESEMPENHO DO COTISTA E NÃO-COTISTA
NO CURSO DE CÁLCULO DIFERENCIAL E INTEGRAL I E II NAS
ENGENHARIAS CIVIL, ELÉTRICA E MECÂNICA COM ANÁLISE
UNIVARIADA E MULTIVARIADA DOS DADOS

Dissertação aprovada como requisito parcial para a obtenção do grau de Mestre no Curso de Pós-Graduação em Métodos Numéricos em Engenharia, do Departamento de Matemática, Setor de Ciências Exatas e do Departamento de Construção Civil, Setor de Tecnologia, Universidade Federal do Paraná, pela seguinte banca examinadora:

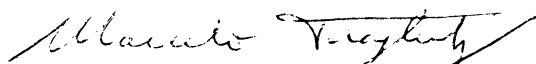
Orientadora:


Prof.^a. Dr.^a. Liliانا Madalena Gramani
Departamento de Matemática, UFPR

Co-Orientador:


Prof. Dr. Anselmo Chaves Neto

Departamento de Estatística, UFPR


Prof. Dr. Marcelo Henrique Romano Tragtenberg

Departamento de Física, UFSC


Prof. Dr. Jair Mendes Marques

Programa de Pós-Graduação em Métodos Numéricos
em Engenharia, UFPR

Curitiba, 30 de maio de 2011.

Uma pessoa só muda quando a dor de permanecer for maior que a dor de mudar.

AGRADECIMENTOS

Agradeço a algumas pessoas que foram fundamentais em minha trajetória não apenas acadêmica mas também em meu crescimento pessoal. Entre elas está João Paulo que foi como um pai para mim, ensinando uma das mais importantes lições de vida que aprendi: acreditar em mim.

Além deste homem que é muito importante em minha história tive uma mão amiga de outros como: Luis Carlos, meu compadre, que ao me incentivar a prestar vestibular me fez ampliar meus horizontes e perceber que eles eram maiores do que eu imaginava. Nilton, meu colega de graduação por ter me acompanhado em minha trajetória acadêmica e profissional.

As professoras Tamia e Elizabete que não se limitaram a transmitir conhecimento, mas doaram seu tempo, seu saber, sua presença atenta e generosa, me ajudando a preencher alguns déficit de conhecimento.

Alguns foram mais que colegas de trajetória, Vania foi para mim uma irmã mais velha, não pela idade, mas por me conduzir pela mão no penoso exercício de decifrar os códigos acadêmicos. Marina com sua habilidade em decifrar as esfinges tecnológicas, Fabio por sua generosidade e coração enorme, Rodrigo que simplesmente como eu entendeu e partilhou as lutas cotidianas que nos levaram ao mestrado e Simone que me conduziu na descoberta de um mundo além da universidade.

Um agradecimento especial deve ser feito a professora Liliana e ao professor Anselmo que aceitaram a árdua tarefa de me orientar, cada um a seu modo, mas sempre cuidadosos e atentos.

Agradeço ainda a meu filho Luiz Augusto que junto comigo trilhou esse caminho nem sempre fácil da busca por conhecimento, abrindo mão em muitos momentos da preciosa convivência familiar.

A Maristela, por sua recepção calorosa e à Capes, pelo auxílio financeiro.

Enfim agradeço a todos os anjos que passaram por mim, inclusive os que eu não pude aqui citar nominalmente. Passando em minha vida compartilhando luz e sabedoria, que cedo ou tarde nos deixam, quando percebem que sua missão de nos fazer caminhar sozinhos está cumprida.

RESUMO

Na Universidade Federal do Paraná, os cursos dos setores de exatas e de tecnologia apresentam em sua grade curricular básica algumas disciplinas, dentre elas, Cálculo I e II. Estas disciplinas por conter um grande número de reprovações, acarretam um atraso para o aluno quanto à conclusão de seu curso de graduação. A fim de auxiliar a Universidade Federal do Paraná a aumentar o rendimento acadêmico dos alunos ingressantes nos cursos de Engenharia Civil, Engenharia Elétrica e Engenharia Mecânica dos anos de 2005, 2006, 2007 e 2008, foi aplicado estatística descritiva e indutiva para descrição das variáveis, técnica da análise multivariada e redes neurais com o objetivo de traçar o perfil do aluno calouro e prever o seu desenvolvimento universitário nos primeiros anos de seu curso de graduação. Para este fim, utilizou-se o questionário sócioeducacional do aluno calouro fornecido pelo Núcleo de Concurso contendo diversas informações, além de seu histórico universitário contendo informações referentes às disciplinas de Cálculo I e II, fornecido pelas coordenações de seu respectivo curso de graduação. O software Minitab foi utilizado para fazer a análise estatística descritiva e indutiva dos dados para comparação entre alunos Cotistas e não-Cotistas. As técnicas de regressão logística e redes neurais foram utilizadas para traçar o perfil do aluno e comparação estatística entre elas.

Palavras-chave: Rendimento acadêmico. Análise multivariada. Regressão logística. Rede neural.

ABSTRACT

At the University Federal of Paraná, courses of exact and technology sector have in their curricula some basic disciplines, including Calculus I and II. These disciplines have a large number of failure, while it causes to the students a delay at course conclusion. In order to helping Federal University of Paraná increasing the academic yield of students that are admitted at in courses of Civil Engineering, Electric Engineering and Mechanical Engineering in 2005, 2006, 2007 e 2008 years, it was applied statistical analysis descriptive and inductive, multivariate analysis techniques and neural networks in order to trace the profile of freshman student and to predict their academic development during the first years of their undergraduate course. For this, it was used the freshman student's socio-educational survey provided by the Center for Tenders containing several information about them, and academic record by the undergraduate course departments, containing information about Calculus I and II disciplines. Minitab software was used to do data statistical analysis descriptive and inductive in order to compare shareholders and non-shareholders students. Logistic regression and neural networks techniques software were used in order to finding student profile, and their results were compared. Analysis results were represent graphically. Moreover, results found by logistic regression and neural networks techniques were similar.

Key-words: Academic yield. Multivariate analysis. Logistic Regression. Neural network.

LISTA DE FIGURAS

FIGURA 2.1	– POSSIBILIDADE DE ENTRADA	30
FIGURA 2.2	– FUNÇÃO SIGMÓIDE FONTE: O AUTOR (2011)	41
FIGURA 2.3	– REPRESENTAÇÃO BÁSICA DO NEURÔNIO BIOLÓGICO ...	45
FIGURA 2.4	– EXEMPLO DE NEURÔNIO ARTIFICIAL	46
FIGURA 2.5	– FUNÇÕES DE ATIVAÇÃO	47
FIGURA 2.6	– ALGORITMO GENÉRICO DE UMA RNA SUPERVISIONADA	48
FIGURA 3.1	– DISTRIBUIÇÃO DO CANDIDATO MEDIANTE ENTRADA	55
FIGURA 4.1	– APROVEITAMENTO MASCULINO	65
FIGURA 4.2	– APROVEITAMENTO FEMININO	65
FIGURA 4.3	– APROVEITAMENTO COTA AFRO	65
FIGURA 4.4	– APROVEITAMENTO COTA ESCOLA PÚBLICA	65
FIGURA 4.5	– GÊNERO EM RELAÇÃO A ENTRADA	66
FIGURA 4.6	– FREQUÊNCIA COM QUE OS ALUNOS CURSARAM A DISCIPLINA DE CÁLCULO DIFERENCIAL E INTEGRAL I	69
FIGURA 4.7	– FREQUÊNCIA COM QUE OS ALUNOS CURSARAM A DISCIPLINA DE CÁLCULO DIFERENCIAL E INTEGRAL II	70
FIGURA 4.8	– COTISTA - 2005	73
FIGURA 4.9	– NÃO-COTISTA - 2005	73

FIGURA 4.10	– COTISTA - 2005	74
FIGURA 4.11	– NÃO-COTISTA - 2005	74
FIGURA 4.12	– COTISTA - 2005	75
FIGURA 4.13	– NÃO-COTISTA - 2005	75
FIGURA 4.14	– COTISTA - 2005	76
FIGURA 4.15	– NÃO-COTISTA - 2005	76
FIGURA 4.16	– COTISTA - 2005	77
FIGURA 4.17	– NÃO-COTISTA - 2005	77
FIGURA 4.18	– COTISTA - 2005	79
FIGURA 4.19	– NÃO-COTISTA - 2005	79
FIGURA 4.20	– COTISTA - 2005	80
FIGURA 4.21	– NÃO-COTISTA - 2005	80
FIGURA 4.22	– COTISTA - 2005	81
FIGURA 4.23	– NÃO-COTISTA - 2005	81
FIGURA 4.24	– LINHA DE CORTE	85
FIGURA 4.25	– LINHA DE CORTE	88
FIGURA 4.26	– REDE NEURAL GERADA PELO SOFTWARE STATGRAPH- ICS	89
FIGURA 4.27	– REDE NEURAL GERADA PELO SOFTWARE STATGRAPH- ICS	90

LISTA DE TABELAS

TABELA 2.1	– VALORES CRÍTICOS PARA O TESTE DE NORMALIDADE ..	36
TABELA 4.1	– CONCLUSÃO NAS DISCIPLINAS CÁLCULO I E II	65
TABELA 4.3	– DISTRIBUIÇÃO DA ENTRADA	70
TABELA 4.5	– CONDIÇÕES DE MORADIA	71
TABELA 4.7	– GRAU DE ESCOLARIDADE DO PAI	72
TABELA 4.9	– GRAU DE ESCOLARIDADE DA MÃE	74
TABELA 4.11	– RENDA MENSAL FAMILIAR	75
TABELA 4.13	– NECESSIDADE DE TRABALHAR	76
TABELA 4.15	– PROCEDÊNCIA EDUCACIONAL	78
TABELA 4.17	– ESCOLHA DO CURSO	78
TABELA 4.19	– MOTIVO DA ESCOLHA DO CURSO	80
TABELA 4.21	– O QUE ESPERA DO CURSO	81
TABELA 4.23	– DESEMPENHO DE PREDIÇÃO	84
TABELA 4.25	– MODELO DE REGRESSÃO ESTIMADA	86
TABELA 4.27	– DESEMPENHO DE PREDIÇÃO “SUCESSO”	87
TABELA 2.1	– VARIÁVEIS DESMEMBRADAS	99

LISTA DE QUADROS

QUADRO 3.1	- ESTATÍSTICA DO TESTE PARA DUAS PROPORÇÕES	42
QUADRO 3.2	- RESULTADO MINITAB	46
QUADRO 3.3	- RESULTADO CONCLUSÃO - MINITAB	55
QUADRO 3.4	- RESULTADO IDADE - MINITAB	58
QUADRO 3.5	- IGUALDADE DE VARIÂNCIA - MINITAB	58

SUMÁRIO

1 INTRODUÇÃO	17
1.1 OBJETIVOS	18
1.1.1 Objetivo geral	18
1.1.2 Objetivos específicos	19
1.2 JUSTIFICATIVA	19
1.3 ESTRUTURA DA DISSERTAÇÃO	20
2 REVISÃO DE LITERATURA	21
2.1 TRABALHOS REALIZADOS	21
2.2 HISTÓRICO DA UNIVERSIDADE FEDERAL DO PARANÁ - UFPR	23
2.2.1 Vestibular	24
2.2.2 Fases do vestibular	26
2.2.3 Cotas	27
2.3 ESTATÍSTICA	30
2.3.1 Teste de Proporções	30
2.3.2 Estatística descritiva	32
2.3.3 Nível de significância	32
2.3.4 Valor p	32
2.3.5 Média aritmética	33

2.3.6 Mediana	33
2.3.7 Quartis	34
2.3.8 Desvio Padrão	34
2.3.9 Escore Padronizado	35
2.3.10 Teste de normalidade (Gaussianidade)	35
2.3.11 Igualdade de Variâncias	38
2.3.12 Teste de Levene	39
2.3.13 Teste de Mann-Whitney	40
2.4 REGRESSÃO LOGÍSTICA	41
2.5 REDES NEURAIS	42
2.5.1 Características Básicas das Redes Neurais Artificiais	43
2.5.2 Função de Ativação de uma Rede Neural	46
2.5.3 Aprendizado ou Treinamento de uma Rede Neural	47
2.5.4 Fluxo de Dados em uma Rede Neural	48
2.5.5 Rede Neural Bayesiana	49
3 MATERIAL E MÉTODO	50
3.1 OBTENÇÃO DOS DADOS	50
3.1.1 Dados fornecidos pelo núcleo de concurso	50
3.1.2 Dados fornecidos pelas coordenações	51
3.1.3 Organização dos dados	52
3.2 DISCRIMINAÇÃO DAS VARIÁVEIS	53

3.2.1 Variável Conclusão	53
3.2.2 Variável Sucesso	53
3.2.3 Variável Gênero	54
3.2.4 Variável Idade	54
3.2.5 Variável Nota Primeira Fase	54
3.2.6 Variável Nota Segunda Fase	54
3.2.7 Variável Frequência que Kursou a Disciplina Cálculo I	54
3.2.8 Variável Frequência que Kursou a Disciplina Cálculo II	55
3.2.9 Variável Tipo de Entrada	55
3.2.10 Variável Estado Civil	56
3.2.11 Variável Moradia	56
3.2.12 Variável Grau de Instrução do Pai	56
3.2.13 Variável Grau de Instrução da Mãe	57
3.2.14 Variável Renda Mensal Familiar	57
3.2.15 Variável Necessidade de Trabalhar	58
3.2.16 Variável Procedência Educacional	58
3.2.17 Variável Escolha do Curso	59
3.2.18 Variável Motivo da Escolha do Curso	59
3.2.19 Variável O que Espera do Curso	60
3.3 MÉTODOS	60
4 RESULTADOS E DISCUSSÃO	63

4.1 APRESENTAÇÃO E DESCRIÇÃO DOS DADOS	63
4.1.1 Variável Conclusão	63
4.1.2 Variável Gênero	66
4.1.3 Variável Idade	66
4.1.4 Variável Nota Primeira Fase	67
4.1.5 Variável Nota Segunda Fase	67
4.1.6 Variável Frequência que cursou a disciplina Cálculo I	68
4.1.7 Variável Frequência que cursou a disciplina Cálculo II	69
4.1.8 Variável Tipo de Entrada	70
4.1.9 Variável Estado civil	71
4.1.10 Variável Moradia	71
4.1.11 Variável Grau de instrução do Pai	72
4.1.12 Variável Grau de instrução da mãe	73
4.1.13 Variável Renda mensal familiar	74
4.1.14 Variável Necessidade de trabalhar	76
4.1.15 Variável Procedência educacional	77
4.1.16 Variável Escolha do curso	78
4.1.17 Variável Motivo da escolha do curso	79
4.1.18 Variável O que espera do curso	81
4.2 ANÁLISE DA TÉCNICA DE REGRESSÃO LOGÍSTICA	82
4.2.1 Variável dependente “Conclusão”	83

4.2.2 Variável dependente “Sucesso”	85
4.3 ANÁLISE DA TÉCNICA DE REDE NEURAL	88
4.3.1 Resultados das RNAs	88
5 CONCLUSÃO E SUGESTÕES DE TRABALHOS FUTUROS	91
REFERÊNCIAS	94
APÊNDICE A – QUESTÕES DO QUESTIONÁRIO SÓCIOEDUCACIONAL	96
APÊNDICE B – VARIÁVEIS DESMEMBRADAS	99

1 INTRODUÇÃO

Pesquisas vem sendo realizadas acerca do perfil dos graduandos de diferentes áreas do conhecimento e campos de formação. Nesta perspectiva, o presente trabalho se propõe a traçar o perfil do aluno ingressante no curso de Engenharia Civil, Engenharia Elétrica e Engenharia Mecânica da Universidade Federal do Paraná (UFPR), avaliando o seu rendimento acadêmico, quanto às disciplinas de Cálculo I e II ¹, considerando-se dois grupos: Cotista e não-Cotista.

É comum existir nos vários cursos superiores, disciplinas que se tornam símbolos do curso quer positivamente ou negativamente, em parte devido às suas dificuldades, em parte por exigirem formas de conhecimento ou raciocínio diferentes daqueles que os alunos estão acostumados. Devido essa mitificação, essas disciplinas (Cálculo I e II) representam um desafio para os alunos, e os relatos das dificuldades encontradas passam de turma em turma. Assim, os alunos acabam por considerar natural um insucesso, e os professores estabelecem altos padrões de reprovação como “normais”. Esses padrões tornam aparentemente desnecessária qualquer reflexão sobre os problemas enfrentados, já que estão “dentro da normalidade”.

Diante deste contexto e visando a alteração desse quadro de reprovação e evasão nas disciplinas (Cálculo I e Cálculo II), o departamento de Matemática da UFPR em 2000 criou uma comissão formada por professores do próprio departamento com o propósito de analisar as causas do grande número de reprovação dos alunos nas

¹ no texto do presente trabalho atribui-se a disciplina envolvendo Cálculo I para Cálculo Diferencial e Integral de uma variável e Cálculo II para a disciplina envolvendo Cálculo Diferencial e Integral de duas ou mais variáveis.

disciplinas. A análise foi feita de maneira empírica e o resultado foi insuficiente para alcançar o propósito esperado pelo departamento. Deste fato surge a necessidade de se fazer um estudo estatístico e traçar novas políticas de desempenho, para prevenir o rendimento do aluno traçando o seu perfil e descrevendo todas as variáveis. Este estudo procura alcançar esse objetivo, através da estatística descritiva, indutiva, multivariada e rede neural.

Nesta perspectiva a UFPR, nas diversas áreas de atuação, busca constantemente aumentar o índice de aprovação mantendo a qualidade da educação, porém, nos últimos anos vem encontrando dificuldades em ministrar conteúdos específicos da grade curricular, particularmente nas disciplinas de Cálculo I e II, devido à falta de pré-requisito escolar básico para acompanhar as disciplinas, aguçado pelo processo de adaptação do calouro. Isso tem como consequência um alto índice de reprovação e evasão do curso de graduação, significando prejuízo financeiro para a sociedade. Essas são disciplinas básicas dos cursos existentes nos setores de Ciências Exatas e de Tecnologia.

1.1 OBJETIVOS

1.1.1 Objetivo geral

Identificar alunos com propensão à reprovação ou desistência nas disciplinas básicas de Cálculo I e II, do setor de Tecnologia da UFPR, nos cursos de Engenharia Civil, Engenharia Elétrica e Engenharia Mecânica com entrada nos anos de 2005 à 2008. Proporcionar às coordenações envolvidas, elementos auxiliares nas eventuais tomadas de decisão e novas políticas para que o aluno aproprie-se dos conhecimentos necessários para a sua permanência e ascensão dentro do curso de graduação por ele escolhido.

1.1.2 Objetivos específicos

- Descrever estatisticamente todas as variáveis;
- Aplicar técnicas de Estatística Multivariada;
- Traçar o perfil do aluno calouro para prever o seu desenvolvimento universitário;
- Comparação entre Cotistas e não-Cotistas.

1.2 JUSTIFICATIVA

A importância desse trabalho é contribuir com o conhecimento prévio da possível potencialidade do aluno através de seu “perfil” e evidenciar sua potencialidade, objetivando uma melhora significativa na qualidade da aprovação, mudando o paradigma de normalidade na reprovação e poder traçar novas políticas para ministrar as disciplinas de Cálculo I e II. Trata-se de um estudo exploratório descritivo com abordagem quantitativa. Nesse enfoque pretende-se:

Aumentar o índice de aprovação nas disciplinas de Cálculo I e II. Para isto tem-se a hipótese de que se os professores souberem a priori qual o nível de conhecimento matemático de sua turma, poderão articular os conteúdos ensinados, visando com que o aluno seja aprovado nestas disciplinas, com o objetivo de terminar o seu curso de graduação no tempo mínimo previsto, representando uma economia para a sociedade.

Auxiliar os coordenadores da graduação a identificar alunos propensos a reprovação e desistência, mediante o “Perfil do aluno”, de forma a diminuir a repetência e evasão.

Melhorar o processo de aprendizagem, com a estipulação de programas interrelacionados de forma criteriosa, por exemplo, a monitoria, curso de pré-cálculo, fazer outras divisões de conteúdos nas disciplinas ou alterar o período de cada disciplina.

1.3 ESTRUTURA DA DISSERTAÇÃO

Este trabalho é composto pela introdução e foi desenvolvido em mais 5 capítulos, os quais são descritos em detalhes a seguir:

Capítulo 2: apresenta de uma forma resumida a literatura já existente sobre o tema deste trabalho. Uma breve abordagem sobre a Universidade Federal do Paraná é realizada, incluindo informações sobre o vestibular, suas fases e cotas. Também faz-se uma abordagem sobre a estatística indutiva descrevendo o teste de proporções e a estatística descritiva descrevendo alguns termos como nível de significância, valor p , média aritmética, mediana, quartis, desvio padrão, escore padronizado presentes no teste de normalidade, teste de igualdade de variância, teste-F, teste de Levene e teste de Mann-Whitney. Ainda é apresentada a Regressão Logística com enfoque no modelo da regressão logística múltipla. Por fim, resolve-se a Redes neurais bayesiana com o objetivo de realizar comparações e tornar os resultados obtidos cada vez mais confiáveis.

Capítulo 3: neste capítulo se descreve o banco de dados usado e o método aplicado, ou seja, todo o processo de obtenção dos dados e organização dos mesmos e a maneira como os testes e técnicas foram aplicados.

O capítulo 4 discute os resultados e finalmente no capítulo 5 apresenta-se a conclusão e as referências.

2 REVISÃO DE LITERATURA

Neste capítulo é relatado a revisão de alguns trabalhos realizados quanto ao estudo do perfil do aluno escolar, universitário e de especialização. Um breve histórico do processo de ingresso no ensino superior é apresentado, considerando suas ações afirmativas, tendo como foco a Universidade Federal do Paraná. Também é apresentado termos da matemática indutiva e descritiva e respectivos testes. Por fim descreve-se as técnicas utilizadas neste trabalho, regressão logística e rede neural.

2.1 TRABALHOS REALIZADOS

Diversos autores vem focando o perfil do aluno da educação básica, ensino médio e universitário, assim como trabalhos sob a perspectiva de políticas afirmativas, dentre os trabalhos, destaca-se Allerup (2003) que analisou e comparou a estrutura de respostas para as perguntas relacionadas à atitude cívica (percepções dos valores democráticos) e investiga ainda a relação entre o conhecimento e as atitudes vinculadas a uma perspectiva de gênero, entre estudantes de catorze anos de idade na Dinamarca. Shankland *et al.* (2010), fizeram estudos longitudinais medindo a adaptação ao ensino superior, comparando 50 participantes de escolas alternativas com 80 estudantes do sistema escolar tradicional; Dodds *et al.* (2010) compararam o desempenho acadêmico dos graduandos e pós-graduandos em um curso de medicina de uma grande universidade estrangeira (Universidade de Melbourne); Smith, Curbow e Stillman (2008) avaliaram os perfis de risco psicossocial dos alunos fumantes e seu

desempenho acadêmico no decorrer dos cursos de graduação e de pós-graduação, examinando os aspectos psicossociais do tabagismo; Teixeira (2006) apresentou um trabalho que contribuiu com os dados que mensuram a evasão escolar investigando variáveis que possam discriminar dois grupos: alunos que possuem maiores chances de sair das Instituições de Ensino Superior com habilitação e alunos que podem sair sem habilitação; Yogev e Yogev (2006) mostram um estudo que compara o perfil dos professores pesquisadores das faculdades israelenses com os departamentos universitários de educação; Ribeiro e Menezes Filho (2009) segundo a pesquisa, concluiu-se que o grau de escolaridade da mãe é fator mais importante na influência das notas dos filhos, explicando que mães com maior escolaridade estão mais atentas. Ainda segundo a pesquisa, filhos de mãe universitária tem notas quase 20% acima da média. E quanto mais se avança na escola, mais evidente fica o impacto da escolaridade da mãe; Attewell e Lavin (2007) segundo os autores baseando-se em uma análise predominante estatística afirmam que os estudantes de ação afirmativa em instituições altamente seletivas foram muito bem sucedidos em termos de graus. Em termos de cor, os alunos brancos têm um desempenho significativamente superior aos negros, mas não com relação aos que se declaram pardos (mulatos). Isto provavelmente está relacionado a características familiares não observáveis nos dados, que fazem com que o aluno negro tenha um aprendizado menor, tais como a qualidade da educação dos pais, pois os alunos negros tendem a ser mais pobres do que os brancos. Alternativamente, pode refletir um menor esforço dos alunos negros, por acharem que o estudo terá um impacto menor na sua vida, devido à discriminação e estigma, os americanos prezam a crença de que os jovens podem sair da pobreza ou desvantagem, se perseverar na escola e trabalhar seu caminho até um grau de faculdade. Espera-se também que uma vez que a primeira geração de uma família chegar a completar uma educação universitária, as gerações vão sustentar essa vantagem; Bowen e Bok (1998) o livro é uma tentativa de mapear o que significaram as normas de admissão sensíveis à raça no correr de um longo do tempo, tanto para os indivíduos que

são admitidos quanto para a sociedade que investe em sua educação, e que conta maciçamente com uma liderança futura. Trata-se de um estudo sumamente quantitativo que utiliza técnicas simples, havendo uma correlação entre ter padrões mais elevados de admissão e de graduação. Tendo como descoberta central “Aqueles que frequentaram escolas mais seletivas, formaram-se com taxas superiores que aqueles que participaram em escolas menos seletivas”. Conclui-se que “a faculdade ou universidade que um estudante frequênta é muito melhor preditor das chances de se formar do que o próprio placar do aluno”

2.2 HISTÓRICO DA UNIVERSIDADE FEDERAL DO PARANÁ - UFPR

A UFPR foi fundada em 19 de dezembro de 1912 com os cursos de Engenharia Civil, Medicina e Direito, sendo a mais antiga instituição de ensino, concebida nos moldes de uma universidade no Brasil. Atualmente as instalações da universidade estão distribuídas em vários *campi* na cidade de Curitiba e em outras cidades do Paraná. Atualmente a universidade possui 77 cursos de graduação, 124 de especialização, 41 de mestrado e 26 de doutorado, para atender às expectativas da sociedade paranaense.

Dentre os diversos cursos ofertados pela UFPR atualmente, os cursos de: Bacharelado em Ciência da Computação, Biologia, Engenharia Ambiental, Engenharia Cartográfica, Engenharia Civil, Engenharia da Produção, Engenharia de Bioprocessos e Biotecnologia, Engenharia Elétrica (Eletron., Eletrotec., Telecom.), Engenharia Elétrica (Ênfase em Sistemas Eletrônicos Embarcados), Engenharia Industrial Madeireira, Engenharia Mecânica, Engenharia Química, Estatística, Física (Bacharelado e licenciatura), Geologia, Informática Biomédica, Matemática (Bacharelado e Licenciatura), Matemática Industrial, apresentam em sua grade curricular as disciplinas de Cálculo I e Cálculo II.

Esta pesquisa está focada nessas disciplinas e dentre os cursos citados foram

analisados os de: Engenharia Civil, Engenharia Elétrica e Engenharia Mecânica. A forma de ingresso nos cursos ofertados na UFPR é através do vestibular de acordo com o Decreto 8.659, de 05/04/1911.

2.2.1 Vestibular

O processo seletivo, vestibular, é um concurso de habilitação para o ingresso no ensino superior de acordo com a Lei 4024/61 (primeira Lei de Diretrizes e Bases da Educação Nacional), que está fundamentado na lógica da meritocracia, ou seja, responsabiliza individualmente as pessoas por seu sucesso ou fracasso, independente de sua trajetória histórica ou da trajetória coletiva de um grupo, sociedade ou país. O princípio da meritocracia é a disputa justa entre pessoas iguais, em que o resultado é definido unicamente pelo mérito, sem paternalismos ou privilégios. Neste contexto, em 1968 a estrutura universitária, utiliza o vestibular como única forma de acesso ao ensino superior, tornando-o mais seletivo. A partir de 1996, com a nova lei de Diretrizes e Bases da Educação Nacional, num contexto sócio-político de grande demanda por democratização do ensino superior, ocorreu uma flexibilização das regras de abertura e funcionamento de cursos superiores e paulatinamente uma grande modificação nos mecanismos de acesso às universidades. Apesar destas modificações, o processo continua fundamentado na meritocracia, o que resulta numa representação elitista, ou seja, as camadas populares permaneciam significativamente excluídas do espaço acadêmico. No âmbito da globalização, o século XXI trouxe para os brasileiros uma demanda ainda maior de continuidade da escolarização, desencadeando a formulação de políticas públicas, apoiadas nos preceitos constitucionais, na LDBEN 9394/96 e no plano Nacional de Educação (Lei 10172/2001), que recomendavam a supressão do analfabetismo e ampliação da porcentagem da população com acesso à escolaridade básica e superior.

Uma das metas desse plano, que tem provocado alterações nos mecanismos de

ingresso ao ensino superior, principalmente nas universidades públicas, diz respeito à inclusão de alunos por meio da criação de políticas que facilitem às minorias ¹, vítimas de discriminação ao acesso à educação superior, através de programas ² de compensação de deficiências de sua formação escolar inicial, permitindo-lhes, desta forma, competir em igualdade de condições nos processos de seleção.

Estas políticas públicas, denominadas especificamente de políticas afirmativas, de acordo com Piovesan (2005), representam um poderoso instrumento de inclusão social, pois se constituem como “medidas especiais e temporárias que, buscando remediar um passado discriminatório, objetivam acelerar o processo de igualdade, com o alcance da igualdade substantiva por parte de grupos vulneráveis, como as minorias étnicas e raciais, as mulheres, dentre outros grupos”.

Historicamente, conforme Wedderburn (2005), o conceito de ação afirmativa originou-se na Índia imediatamente após a Primeira Guerra Mundial, ou seja, bem antes da própria independência deste país. Ainda segundo Wedderburn (2005) em 1919, Bhim-rao Ramji (1891 - 1956), jurista, economista e historiador, propôs pela primeira vez na história, em pleno período colonial britânico, a “representação diferenciada” dos segmentos populacionais designados e considerados inferiores.

No Brasil, os grupos considerados politicamente minoritários, prejudicados pelo processo de colonização e escravidão, e atualmente pela discriminação racial, são os afro-descendentes ³ e indígenas, que continuam socialmente, economicamente e

¹O termo minoria diz respeito a determinado grupo humano ou social que esteja em inferioridade numérica ou em situação de subordinação sócio-econômica, política ou cultural, em relação a outro grupo, que é majoritário ou dominante em uma dada sociedade. Uma minoria pode ser étnica, religiosa, linguística, de gênero, idade, condição física ou psíquica.

²A democratização do ensino superior com qualidade vem sendo discutida pela sociedade, exigindo o estabelecimento de políticas públicas que ampliem, não apenas o acesso, mas que garantam condições de permanência destes estudantes neste nível de ensino. A educação que, segundo a Constituição, é um dos meios de se alcançar a cidadania e a reversão das desigualdades sociais, começa agora a ser um dos principais temas para discussão entre a sociedade civil e o Estado. Na última década foram adotadas, no Brasil, políticas inclusivas de ações afirmativas. Dentre os programas e ações, nas universidades públicas foi instituído o sistema de cotas sociais e raciais para pessoas egressas de escolas públicas e, para o ensino superior privado o acesso dos estudantes oriundos de escolas públicas ocorre através do Programa Universidade para Todos - PROUNI.

³Termo polêmico uns entendem como simples eufemismo ou “moda”, outros são contra pois en-

culturalmente sofrendo as conseqüências destes movimentos históricos.

Neste contexto, a UFPR aprovou em abril de 2007 o Plano de Metas de Inclusão Racial e Social da instituição. A resolução valerá por dez anos, contados a partir de 11/05/04, ano em que começou a discussão sobre cotas na UFPR.

2.2.2 Fases do vestibular

O processo seletivo adotado de 1996 até 2004 pela UFPR, estava focado em avaliar habilidade de leitura e escrita do candidato. Em 2005 o vestibular foi totalmente remodelado, com seleção diferenciada considerando as especificidades de cada curso, possibilitando duas formas de avaliação: Múltipla escolha (1ª fase) e discursiva (2ª fase), além da inclusão do Plano de Metas de Inclusão Racial e Social. Nesse contexto, desde 2003, o questionário Sócioeducacional da UFPR incorporou uma questão sobre a “cor ou raça” do candidato que, seguindo a classificação do IBGE, oferece cinco opções de resposta: branca, preta, amarela, parda e indígena. O candidato responde ao questionário Sócioeducacional no momento da inscrição, o qual não influenciará na aprovação do processo seletivo. Esse questionário tem como objetivo conhecer a realidade socioeconômica e cultural dos candidatos a fim de dimensionar os níveis de carência e afirmar a necessidade de implantação de uma política de assistência estudantil, inserida no campo dos direitos sociais e cidadania.

A primeira fase do vestibular na UFPR, é constituída de uma prova com 80 questões objetivas, no formato de múltipla escolha, sobre os conteúdos do Ensino Médio, havendo nove questões sobre cada uma das seguintes matérias: matemática, física, química, biologia, geografia e história. As questões de língua estrangeira moderna são em número de oito. Quanto ao português, são ao todo 18 questões, sendo nove de com-

xergam a questão como simples fuga da estigmatização vinculada ao termo negro, termo que muitos acreditam que deva ser mantido e valorizado, outros não concordam pois entendem que toda a humanidade é afro-descendente e sendo assim ,não seria correto utilizar o termo apenas para os negros. De qualquer forma, há uma tendência cada vez maior da utilização do termo em questão (tanto nacional quanto internacionalmente, quer pela militância ou seja pela sociedade em geral).

preensão de texto e nove relacionadas ao conteúdo de literatura brasileira e de uso da língua. Cada questão vale um ponto.

Para a prova de 2ª fase, são convocados os candidatos melhor classificados na 1ª fase em cada curso ⁴ para fazerem (i), uma prova comum, cuja compreensão e produção de textos são avaliados, (ii) uma prova de questões discursivas, com matérias específicas de acordo com o curso escolhido. Para os cursos de Engenharia Civil, Engenharia Elétrica e Engenharia Mecânica as matérias específicas são matemática e física, tendo cada uma delas 10 questões discursivas, valendo 4 pontos cada, totalizando-se 40 pontos.

A prova de compreensão e produção de textos, da 2ª fase, valerá 60 pontos. Cada uma das provas específicas valerá 40 pontos. Para os cursos com duas provas específicas, a pontuação máxima é $80 + 60 + 80 = 220$. A nota do candidato nas provas do processo seletivo será obtida mediante a soma dos pontos auferidos em cada prova, dividindo-se esse valor pelo valor máximo de pontos possíveis nas provas previstas para seu curso e multiplicando-se o resultado por mil. Para composição do desempenho final do candidato, será efetuado a média ponderada entre a nota obtida pelo candidato no processo seletivo e a nota do ENEM ⁵, as quais assumirão respectivamente, os pesos de 90% e 10%.

2.2.3 Cotas

O sistema de cotas da UFPR foi adotado no vestibular de 2004/ 2005, sendo reservado 20% das vagas para candidatos que tenham cursado todo o período escolar fundamental e médio em escolas públicas e outros 20% das vagas para pessoas que se declarem negras. Fica registrado que esse tratamento diferenciado ocorre somente na 2ª fase do processo de seleção e que esta Resolução é válida pelo período de dez anos a contar da data de 11/05/2004.

⁴independente de terem optado ou não pelas vagas de inclusão racial e social

⁵Exame Nacional do Ensino Médio

Para ingressar como cotista de escola pública o estudante precisa comprovar que toda sua trajetória escolar foi desenvolvida dentro da escola pública. Já o cotista afrodescendente, é avaliado por uma Banca de validação e orientação da auto-declaração designada pelo Reitor da UFPR que busca identificar se aquela pessoa é socialmente considerada negra: Conceito de negro pelo IBGE ⁶, pessoa resultante da união de pretos e pardos, ou seja, preto é aquela pessoa que apresenta todas as características fenotípicas que a caracterizam como negra (cor da pele, cabelo crespo, formato do corpo, etc.); a pessoa parda provem do resultado da miscigenação preto com outra etnia. A depender do espaço, a pessoa classificada como parda pode negociar posições sociais, sendo considerada como “morena”.

Na conferência mundial contra racismo, discriminação racial, xenofobia e intolerância correlata realizado em Durban(2001) popularizou a expressão afrodescendente contemplando toda pessoa com ascendência africana, abarcando as classificações dos pardos e morenos. Este conceito está de acordo com as pesquisas de Nogueira (1998), sociólogo paulista, professor da USP - Universidade de São Paulo, que em meados da década de 1980, realizou uma pesquisa comparativa entre o Brasil e os Estados Unidos, demonstrando as particularidades do racismo brasileiro e norte americano. Seu trabalho indicou que a manifestação do racismo norte americano pode ser denominada de preconceito de origem. Independentemente das características fenotípicas, um norte americano é socialmente considerado negro por aproximadamente oito gerações, enquanto que no Brasil, o racismo fundamenta-se no preconceito de marca, ou seja, quanto mais próximo do fenótipo negro uma pessoa estiver, mais discriminação este indivíduo sofrerá.

Diante da conceituação exposta, a Banca da UFPR formada por professores da instituição, busca identificar se aquele indivíduo que ingressou no processo de seleção do vestibular por meio das Cotas Afrodescendente, é socialmente considerado negro, ou ainda, se esta pessoa sofre as desvantagens sócio-raciais que atingem a

⁶Instituto Brasileiro de Geografia e Estatística

população negra.

O candidato que desejar concorrer às vagas de inclusão racial deverá assinalar essa opção no ato de inscrição, estando ciente de que, se aprovado, deverá fazer a auto-declaração de que pertence ao grupo racial negro e de que possui os traços fenotípicos que o caracteriza como de cor preta ou parda, e deverá ser entrevistado pela Banca. O candidato oriundo de escola pública que desejar concorrer às vagas de inclusão social deverá assinalar essa opção no ato de inscrição, estando ciente de que, se aprovado, deverá comprovar que fez os estudos do ensino fundamental (1ª a 8ª séries) e médio, com aprovação, exclusivamente em escola pública no Brasil. O candidato poderá assinalar apenas uma das opções mencionadas, sendo elas mutuamente excludentes.

O segmento Cotista não se apresenta de forma homogênea. Existe nesta categoria uma subdivisão que pode ser explicada pelo debate etnicorracial acerca da discriminação social, inevitavelmente culminando em desigualdade social e racial. Neste sentido, as políticas afirmativas, que expressam-se no formato das Cotas, podem se manifestar por meio das Cotas Universais, aquelas destinadas a resolver desigualdades fundamentadas no pertencimento social, que atingem negros e brancos. Neste caso o exemplo é a reserva de vagas para estudantes de escola pública, ou nas Cotas Diversalistas, política utilizada no trato de desigualdades que ultrapassam a esfera social e articulam-se com a discriminação racial.

Nesta perspectiva, os dados discriminados nas seções 3.1.1 e 3.1.2 que possibilitaram esta pesquisa foram obtidos por meio da conferência dos alunos devidamente matriculados nos cursos em foco. Esses alunos foram posteriormente selecionados e divididos em dois segmentos: Cotistas e não-Cotistas. No segmento Cotista existem duas possibilidades de participação, a Cota de Escola Pública e a Cota Afrodescendente, no segmento não-Cotista concorrência geral e chamada complementar⁷, não

⁷Nas chamadas complementares ocorre reclassificação geral, na qual há uma única listagem de pontuação, sendo incorporados os remanescentes de cotas, seguindo a classificação.

sendo concomitante ⁸. A FIGURA 2.1 representa essa discriminação.

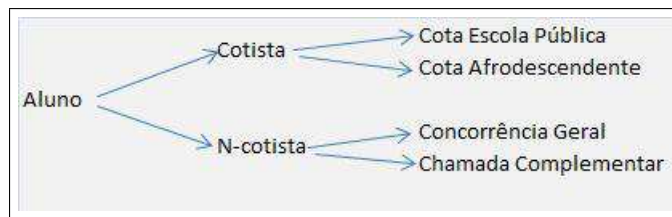


Figura 2.1: Possibilidade de entrada

Esses dados foram considerados como observações em toda a pesquisa para se fazer as análises de aproveitamento do aluno de uma maneira geral e o aproveitamento desses alunos em cada segmento.

2.3 ESTATÍSTICA

2.3.1 Teste de Proporções

Nesta seção são abordados métodos de aplicação a duas proporções amostrais para fazer inferências (teste de hipóteses e construção de intervalos de confiança) sobre duas proporções populacionais. Este teste é utilizado sempre que se deseja determinar a existência de uma diferença significativa entre dois tratamentos.

Ao testar uma hipótese sobre duas proporções populacionais, tem-se como:

- *Finalidade*: testar se a diferença $p_1 - p_2$ entre duas proporções populacionais de um determinado evento é significativa estatisticamente.
- *Condições*: as duas amostras, de tamanhos n_1 e n_2 , devem ser independentes, satisfazendo às condições $n_1 \cdot p_1 \geq 5$ e $n_1 \cdot (1 - p_1) \geq 5$, $n_2 \cdot p_2 \geq 5$ e $n_2 \cdot (1 - p_2) \geq 5$.
- *Estimativa Combinada de p_1 e p_2* : A estimativa combinada de p é denotada por \bar{p} , é dada por $\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$ com o complemento por \bar{q} , $\bar{q} = 1 - \bar{p}$, onde x_1 e x_2 representam o número de sucessos em cada amostra, respectivamente.

⁸Ao mesmo tempo

- *Estimativa Intervalar*: É uma estimativa de um intervalo de valores possíveis, no qual se admite estar o parâmetro populacional. Neste tipo de estimativa tem-se um intervalo de valores em torno do parâmetro amostral, no qual julga-se, com um risco conhecido de erro, estar o parâmetro da população. A esse intervalo chama-se intervalo de confiança. Pode-se construir um intervalo de confiança de nível $1 - \alpha$ para a diferença entre as proporções populacionais $p_1 - p_2$ usando:

$$(\hat{p}_1 - \hat{p}_2) \pm z \sqrt{\frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2}}, \quad (2.1)$$

onde \hat{p}_1 e \hat{p}_2 são respectivamente as proporções amostrais, n_1 e n_2 os tamanhos das amostras e z o escore padronizado correspondente ao nível de confiança de $1 - \alpha$ do intervalo. Então, tem-se:

$$P\left((\hat{p}_1 - \hat{p}_2) - z\sqrt{A} \leq p_1 - p_2 \leq (\hat{p}_1 - \hat{p}_2) + z\sqrt{A}\right) = 1 - \alpha,$$

$$A = \frac{\hat{p}_1(1 - \hat{p}_1)}{n_1} + \frac{\hat{p}_2(1 - \hat{p}_2)}{n_2} \quad (2.2)$$

- *Estatística de Teste para Duas Proporções*: o seguinte teste se aplica a hipótese nula (H_0) e que tem hipótese alternativa (H_1) que se enquadra em um dos três formatos:

Formato 1	Formato 2	Formato 3
$H_0 : p_1 = p_2 = p$	$H_0 : p_1 \geq p_2$	$H_0 : p_1 \leq p_2$
$H_1 : p_1 \neq p_2$	$H_1 : p_1 < p_2$	$H_1 : p_1 > p_2$
$z = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{\bar{p}\bar{q}}{n_1} + \frac{\bar{p}\bar{q}}{n_2}}}$ <p>Hipótese nula: $p_1 - p_2 = 0$</p> $\hat{p}_1 = \frac{x_1}{n_1} \text{ e } \hat{p}_2 = \frac{x_2}{n_2}$ $\bar{p} = \frac{x_1 + x_2}{n_1 + n_2}$ $\bar{q} = 1 - \bar{p}$		

QUADRO 2.1 - Estatística do Teste para Duas Proporções

FONTE: (TRIOLA, 1999)

onde z é a estatística do teste, p_1 e p_2 são as proporções populacionais, \hat{p}_1 e \hat{p}_2 proporções amostrais, \bar{p} estimativa combinada de proporção p na hipótese de igualdade e \bar{q} o complemento de \bar{p} .

2.3.2 Estatística descritiva

Nesta seção, apresenta-se os termos estatísticos mais utilizados neste trabalho com uma breve definição, sendo que todas as variáveis foram submetidas ao teste de normalidade e descritas. Esse teste é determinante para verificação da igualdade de variância, tendo como consequência a utilização dos testes paramétricos (Teste-F) ou não-paramétricos (Teste de Levene e Teste Mann-Whitney).

2.3.3 Nível de significância

É definido como a probabilidade máxima de se cometer erro quando se rejeita a hipótese nula H_0 . Essa probabilidade é simbolizada por α . O valor de α deve ser fixado pelo pesquisador, levando-se em conta a importância prática dos resultados da pesquisa. Na prática é usual a utilização de $\alpha = 0,05$ (5%) ou $\alpha = 0,01$ (1%). Neste trabalho foi considerado o nível de significância de 5%.

2.3.4 Valor p

O p -valor ou valor p representa a probabilidade calculada com base nos dados amostrais que, comparada com o nível de significância α , permite a tomada de decisão, ou seja, rejeitar ou não a hipótese nula H_0 , correspondente ao risco que se corre ao rejeitar-se a hipótese nula.

Para os testes realizados neste trabalho usou-se o software Minitab e o software Statgraphics. Os resultados baseados no p -valor, tem a seguinte interpretação:

- se $p < 0,01$, significativo a 1% e a 5%;

- se $0,01 < p < 0,05$ significativo a 5%;
- se $p > 0,05$, não significativo.

Quanto menor o valor de p , menos provável é o resultado e rejeita-se a hipótese nula H_0 . Geralmente, rejeita-se a hipótese nula se o valor de p é inferior a 0,05 ou a 0,01, correspondendo a 5% ou 1% de chance de um erro tipo I⁹. Para todos os testes descritos foram considerados o nível de significância estatística de $\alpha = 0,05$.

2.3.5 Média aritmética

A média aritmética \bar{x} de um conjunto de n valores x_1, x_2, \dots, x_n , é definida por

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (2.3)$$

2.3.6 Mediana

A mediana Me de um conjunto de n valores ordenados x_1, x_2, \dots, x_n , é representada pelo valor central do conjunto (elemento de ordem $n/2$) para n ímpar ou pela média aritmética dos dois valores de ordem $n/2$ e $(n/2) + 1$ para n par. Portanto, para n ímpar, a mediana do conjunto

5, 7, 9, 13, 17, 19, 20

já ordenado, é igual ao valor central 13. O conjunto para n par, também ordenado

3, 7, 8, 10, 12, 15,

tem a mediana igual a 9, ou seja, a média aritmética entre os valores 8 e 10.

⁹É o erro cometido quando se rejeita a hipótese H_0 , sendo que a mesma é verdadeira. A probabilidade de ocorrência desse erro é representada por α . O valor de α foi fixado em 0,05.

A mediana é útil principalmente quando o conjunto de dados é muito influenciado nos valores extremos, refletindo aqui com mais fidelidade que a média aritmética a medida de tendência central correspondente.

2.3.7 Quartis

Um quartil é qualquer um dos três valores que divide o conjunto ordenado de dados em quatro partes com número igual de termos, e assim cada parte representa $1/4$ da amostra ou estudantes.

Assim, tem-se,

- primeiro quartil (designado por Q_1) ou quartil inferior é o valor abaixo do qual situa-se 25% da amostra ordenada;
- segundo quartil (designado por Q_2) ou mediana é o valor abaixo do qual situam-se 50% da amostra ordenada;
- terceiro quartil (designado por Q_3) ou quartil superior é o valor a partir do qual se encontram 25% dos valores mais elevados. Então, abaixo de Q_3 situam-se 75% da amostra ordenada;
- à diferença entre os quartis superior e inferior chama-se de amplitude inter-quartil.

2.3.8 Desvio Padrão

É a raiz quadrada da soma dos quadrados dos desvios dividida pelo número de graus de liberdade, ou dito de outra forma, é raiz quadrada da média aritmética dos quadrados dos desvios.

2.3.9 Escore Padronizado

Sejam x_1, x_2, \dots, x_n os dados observados em uma amostra de tamanho n de uma população normal, com média μ e desvio padrão σ . Então,

$$Z_i = \frac{X_i - \mu}{\sigma}, \quad i = 1, \dots, n \quad Z_i \sim \mathcal{N}(0, 1) \quad (2.4)$$

é denominado escore padronizado, sendo que quando σ é desconhecido tem-se o escore t , ou seja,

$$t = \frac{X_i - \mu}{s} \sim t_{n-1}. \quad (2.5)$$

2.3.10 Teste de normalidade (Gaussianidade)

Dado um conjunto de n observações x_1, x_2, \dots, x_n pode-se testar a normalidade das mesmas utilizando um método que envolve cálculo de correlação. As etapas são as seguintes:

- Ordenar as observações originais (ordem crescente) obtendo $x_{(1)}, x_{(2)}, \dots, x_{(i)}$ e fixar valores de probabilidades $\frac{1 - \frac{1}{2}}{n}, \frac{2 - \frac{1}{2}}{n}, \dots, \frac{i - \frac{1}{2}}{n}$, com $(i = 1, 2, \dots, n)$.
- Determinar os valores de Z (variável normal padronizada), correspondente a cada uma dessas probabilidades, tal que $Z_i = G^{-1} \left[\frac{i - \frac{1}{2}}{n} \right]$, com $(i = 1, 2, \dots, n)$, sendo que G^{-1} é a função distributiva acumulada.
- Plotar os pares $(z_{(1)}, x_{(1)}), (z_{(2)}, x_{(2)}), \dots, (z_{(n)}, x_{(n)})$, examinando se os pontos estão nas proximidades de uma reta, o que indica qualitativamente a tendência de normalidade. Quantitativamente, pode-se testar a normalidade calculando a correlação entre $x(i)$ e $z(i)$ através da expressão:

$$r = \frac{\sum_{i=1}^n (x_{(i)} - \bar{x})(z_{(i)} - \bar{z})}{\sqrt{\sum_{i=1}^n (x_{(i)} - \bar{x})^2 \sum_{i=1}^n (z_{(i)} - \bar{z})^2}} \quad (2.6)$$

onde x'_i 's são os valores amostrais ordenado e z'_i 's são as variáveis normais padronizadas.

Se $r \geq r_\alpha$, o valor crítico no nível de significância α , veja a TABELA 2.1 não deve-se rejeitar a hipótese de normalidade das observações.

Tabela 2.1: Valores críticos para o teste de normalidade

n	$\alpha = 0.01$	$\alpha = 0.05$	$\alpha = 0.10$
10	0.880	0.918	0.935
15	0.911	0.938	0.951
20	0.929	0.950	0.960
25	0.941	0.958	0.966
30	0.949	0.964	0.971
40	0.960	0.972	0.977
50	0.966	0.976	0.981
60	0.971	0.980	0.984
75	0.976	0.984	0.987
100	0.981	0.986	0.989
150	0.987	0.991	0.992
200	0.990	0.993	0.994

FONTE: Johnson e Wichern (1988)

A justificativa teórica do teste de normalidade ocorre da seguinte forma, se a amostra $[x_1, x_2, \dots, x_n]$ vem de uma distribuição Gaussiana, $\mathcal{N}(\mu, \sigma^2)$ (Normal) com média μ e variância σ^2 , então o escore padronizado z correspondente ao valor x , tem

a forma:

$$Z_i = \frac{X_i - \mu}{\sigma}, \quad (2.7)$$

e portanto:

$X_i = \mu + \sigma Z_i$ representa uma reta de correlação entre X_i e Z , sendo que esta correlação tem que ser alta, este teste é conhecido como Teste de Filliben. Assim, calcula-se a correlação entre X_i e Z_i conforme expressão $r(X_i, Z_i)$ e decide-se com base em uma tabela.

Neste trabalho aplicou-se este teste, com o auxílio do software Minitab baseado no teste de Kolmogorov-Smirnov (KS). O teste de Kolmogorov-Smirnov diz respeito ao grau de concordância entre uma distribuição de frequências absolutas observadas e uma distribuição de frequências absolutas esperadas ou teórica. (MARQUES, 2004).

10 11

Quando o valor de α (nível de significância) é maior do que o p -valor deve-se rejeitar a hipótese de normalidade.

Ao testar a normalidade de uma amostra, ao nível de significância de $\alpha = 10\%$, o seu gráfico tem que ter uma tendência grosseiramente linear para os pontos. Também pode ser confirmado a normalidade através do cálculo do coeficiente de correlação r e compará-lo com o valor crítico, ocorrendo r menor que o valor tabelado rejeita-se a hipótese de normalidade dos dados.

No teste KS essa comparação é feita entre o p -valor e o nível de significância α . Portanto se $p < \alpha$ rejeita-se a hipótese nula de Gaussianidade, caso contrário aceita-se, por exemplo quando usado o nível de significância de 0,10 e p -valor apresentado na janela de gráfico é 0,07, $p > \alpha$ então deve-se rejeitar a hipótese de normalidade ao

¹⁰Distribuição de frequências absolutas observadas - Número de vezes que o valor de uma variável se repete.

¹¹Distribuição de frequências absolutas esperadas - Número de vezes que um valor da variável deve se repetir se a hipótese nula for verdadeira ou ainda número de vezes que um valor da variável deve se repetir de acordo com um determinado modelo.

nível de significância correspondente.

2.3.11 Igualdade de Variâncias

A análise da variância é um método para testar a hipótese de igualdade entre três ou mais médias populacionais. Uma das premissas básicas para aplicação desta técnica, além da normalidade dos dados, é a de igualdade de variâncias. Inicialmente testa-se a hipótese H_0 da igualdade de variâncias:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 \text{ contra a hipótese alternativa}$$

$$H_1 : \text{pelo menos um dos } \sigma_i^2 \text{ é diferente do outro } i = 1, \dots, k$$

quando se vai aplicar a ANOVA ¹² em k grupos.

Para se testar essa hipótese nula H_0 usa-se o teste ($k > 2$) de Levene e o teste-F ($k = 2$). O procedimento estatístico, quando se tem duas populações com distribuição normal e deseja-se testar a hipótese nula de igualdade destas variâncias, ou seja:

$$H_0 : \sigma_1^2 = \sigma_2^2,$$

usa-se a estatística de teste,

$$F_0 = \frac{S_1^2}{S_2^2} \sim F_{v_1, v_2}. \quad (2.8)$$

onde S_1^2 e S_2^2 são as variâncias amostrais, se n_1 e n_2 forem os tamanhos das amostras aleatórias retiradas das populações, a distribuição F terá $v_1 = n_1 - 1$ graus de liberdade e $v_2 = n_2 - 1$ graus de liberdade. Com o nível de significância α e os graus de liberdade, do numerador e do denominador, será obtido o F crítico F_c na distribuição teórica.

¹²A principal aplicação da ANOVA (análise de variância) é a comparação de médias oriundas de grupos diferentes, também chamados tratamentos.

Se $F_0 > F_c$ a hipótese nula deverá ser rejeitada, se $F_0 \leq F_c$ a hipótese nula deverá ser aceita. Assim, comparando o p -valor = $P(F \geq F_0)$ e o nível de significância α . Se p -valor $< \alpha$ a hipótese nula é rejeitada, caso contrário, se p -valor $> \alpha$ a hipótese nula deverá ser aceita.

O Teste-F é muitas vezes usado para determinar se as variações dos dois grupos são iguais. Como o Teste-F é limitado a apenas dois grupos, o Minitab usa os testes de Bartlett e Levene, pois esses testes não têm essa limitação.

2.3.12 Teste de Levene

O teste de Levene é uma estatística inferencial para avaliar a igualdade das variâncias em diferentes amostras baseado na hipótese nula de que as variâncias populacionais são iguais.

Suponha que sejam tomadas $k \geq 2$ amostras aleatórias independentes entre si, com x_{i1}, \dots, x_{in} , $i = 1, \dots, k$. A amostra i representa uma coleção de n_i variáveis aleatórias independentes e identicamente distribuídas com distribuição G_i , com média μ_i e variância σ_i^2 desconhecidos. Sendo assim, a hipótese nula representa a igualdade de variâncias,

$$\begin{cases} H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 \\ H_1 : \text{pelo menos um dos } \sigma_i^2 \text{ é diferente } i = 1, \dots, k. \end{cases}$$

Denota-se os desvios absolutos das variáveis X_{ij} com relação as médias amostrais dos grupos por $\bar{X}_i = n_i^{-1} \sum_{j=1}^{n_i} x_{ij}$ em que $z_{ij} = |x_{ij} - \bar{x}_i|$, com $j = 1, \dots, n_i, i = 1, \dots, k$ e define-se a estatística,

$$W_0 = \left(\frac{n-k}{k-1} \right) \frac{\sum_{i=1}^k n_i (\bar{Z}_i - \bar{Z})^2}{\sum_{i=1}^k \sum_{j=1}^{n_i} (Z_{ij} - \bar{Z}_i)^2} \quad (2.9)$$

em que $\bar{Z}_i = n_i^{-1} \sum_{j=1}^{n_i} Z_{ij}$, $\bar{Z} = n_i^{-1} \sum_{i=1}^k n_i Z_i$ e $n = \sum_{i=1}^k n_i$.

A vantagem do teste de Levene é a não exigência da normalidade dos dados subjacentes. Este teste considera as distâncias das observações da mediana da amostra em vez de sua média amostral.

O resultado apresentado no software Minitab é baseado no p -valor do teste de Levene. Se os resultados forem inferiores a um valor crítico (normalmente $\alpha = 0,05$), a hipótese nula de variâncias iguais é rejeitada e conclui-se que há uma diferença entre as variações na população.

2.3.13 Teste de Mann-Whitney

O teste de Mann-Whitney é o teste não-paramétrico usado para testar se duas amostras independentes foram retiradas de populações com medianas iguais. É uma alternativa ao teste paramétrico para igualdade de médias (Teste-t), pois o teste de Mann-Whitney não exige nenhuma hipótese sobre as distribuições populacionais e suas variâncias. Ele testa a hipótese nula:

H_0 : não há diferença entre as medianas,

contra a hipótese alternativa

H_1 : há diferença entre as medianas.

2.4 REGRESSÃO LOGÍSTICA

A regressão logística é uma técnica estatística que tem como objetivo produzir, a partir de um conjunto de observações, um modelo que permita a predição de valores tomados por uma variável categórica, formulada para prever e explicar uma variável.

A ideia básica do método logístico é simples, ou seja, é o estudo de relações entre variáveis, buscando as variáveis que podem influenciar de alguma forma uma variável dependente, tal como estimar a chance de ocorrência de um determinado evento, ou seja, a chance de evasão de um aluno. Esta estimativa é feita com base em variáveis explicativas, tais como: variáveis sócio-econômicas, notas do vestibular, entre outras.

Para usar-se a regressão logística não é necessário a suposição de normalidade, por ela ser uma técnica robusta. Sua aplicação é apropriada em grande variedade de situações. Mas, a variável resposta tem que ser sempre dicotômica, ou seja, só assume 1 ou 0.

O modelo logístico usa a função matemática sigmóide cuja expressão é:

$$Y = \frac{1}{e^{-x} + 1} = \frac{e^x}{1 + e^x}, \quad x \in \mathbb{R}. \quad (2.10)$$

O gráfico dessa função é:

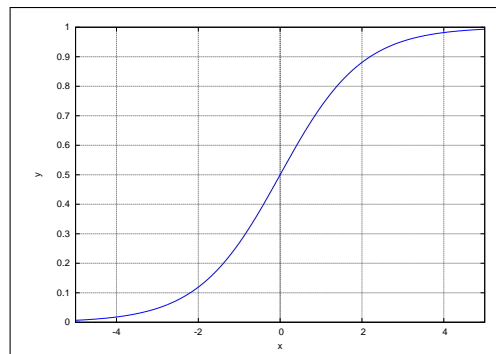


Figura 2.2: Função Sigmóide
FONTE: O autor (2011)

Como se sabe esta função varia de 0 a 1, e pode ser usada para modelar variáveis dicotômicas, ou seja, no lugar de x coloca-se a função $\mu = h(x_1, x_2, \dots, x_p)$ que depende das variáveis explicativas x_1, x_2, \dots, x_p . Então, com $h(x_1, x_2, \dots, x_p)$, uma função das variáveis explicativas x_1, x_2, \dots, x_p que carregam as informações sobre o fenômeno que se quer estimar a probabilidade de ocorrência tem a forma:

$$y = \frac{e^{h(\underline{x})}}{1 + e^{h(\underline{x})}} \quad (2.11)$$

com $y = 0, 1$; $\underline{x}' = [x_1, x_2, \dots, x_p]$ e $h(\underline{x}) \in \Re$.

A regressão logística é uma função de reconhecimento de padrões quando se tem dois grupos. Suponha que se quer prever a probabilidade de um icterico ¹³ sofra de câncer ou de colédoco ¹⁴ entupido. Então, um conjunto de exames clínicos gerou as informações que compõe o vetor $\underline{x}' = [x_1, x_2, \dots, x_{22}]$ e tem este vetor observado para dois grupos: cancerosos e colédoco entupido. Assumindo 1 para canceroso e zero para colédoco entupido tem-se para n_1 indivíduos de um grupo e n_2 indivíduos de outro grupo. Um vetor de zeros e uns. O ajuste logístico de:

$$y = \frac{e^{\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}{1 + e^{\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}} \quad (2.12)$$

fornecerá a chance estimada de um indivíduo ser canceroso.

2.5 REDES NEURAIIS

Uma Rede Neural Artificial (RNA) é uma técnica da computação que visa trabalhar no processamento de dados de maneira semelhante ao cérebro humano. Para isso, é necessário que muitas unidades de processamento simples sejam utilizadas, cada uma delas tendo, possivelmente, uma pequena quantidade de memória local. Essas

¹³A icterícia é a coloração amarelada da pele e das escleras (branco dos olhos) causada por concentrações anormalmente elevadas da bilirrubina (pigmento biliar) no sangue.

¹⁴é um ducto que transporta a bile, a bile é sintetizada no fígado.

unidades precisam ser conectadas por meio de canais de comunicação (conexões), aos quais, usualmente, estão associados valores (pesos) numéricos (Palma Neto; NICOLETTI, 2005).

O cérebro é conhecido por ser um “processador” altamente complexo e que realiza processamentos de maneira paralela. Para isso, ele organiza sua estrutura, ou seja, os neurônios, de forma que eles realizem o processamento necessário.

Nas redes neurais artificiais, a ideia é realizar o processamento de informações tendo como princípio a organização de neurônios do cérebro. Como o cérebro humano é capaz de aprender e tomar decisões baseadas na aprendizagem, as redes neurais artificiais devem fazer o mesmo. Assim, as unidades de processamento que realizam operações, são capazes de armazenar conhecimento baseado em aprendizagem (experiência) e disponibilizam este conhecimento através das conexões para a aplicação em questão. O comportamento inteligente de uma Rede Neural Artificial (RNA) vem das interações entre as unidades de processamento da rede.

Assim, uma RNA pode ser caracterizada por:

- unidades de processamento (neurônios);
- função de ativação;
- padrão de conexão existente entre os neurônios;
- algoritmo de aprendizagem.

2.5.1 Características Básicas das Redes Neurais Artificiais

As Redes Neurais Artificiais (RNA's) são criadas a partir de algoritmos projetados para uma determinada finalidade. É impossível criar um algoritmo desse sem ter conhecimento de modelos matemáticos que simulem o processo de aprendizado do cérebro humano. Basicamente, uma rede neural se assemelha ao cérebro humano

pois:

- o conhecimento é obtido através de etapas de aprendizagem;
- pesos sinápticos são usados para armazenar o conhecimento.

Uma sinapse é o nome dado à conexão existente entre neurônios. Nas conexões são atribuídos valores, que são chamados de pesos sinápticos, ou seja, as redes neurais artificiais têm em sua constituição uma série de neurônios artificiais (ou virtuais) que serão conectados entre si, formando uma rede de elementos de processamento.

Além de uma RNA ser altamente interconectada, também é necessário dizer que:

- Apresenta paralelismo maciço, ou seja, muitos neurônios operando ao mesmo tempo;
- O processamento é distribuído de modo que a informação é não localizada, significando que um fato pode corresponder à atividade de certo número de neurônios;
- Admite tolerância a falhas, assim o prejuízo a poucos neurônios não afeta a operação do cérebro significativamente;
- A aprendizagem é exibida pelo ajustamento do efeito de acoplamento de 2 neurônios.

Assim, para que se possa compreender uma rede neural artificial é necessário ter conhecimento dos movimentos e funções dos neurônios no organismo humano.

O cérebro humano é constituído por cerca de 200 bilhões de neurônios. Cada neurônio processa e se comunica com outros milhares de neurônios continuamente e em paralelo.

A estrutura de um neurônio biológico está apresentada na FIGURA 2.3. A unidade básica de nosso cérebro, o neurônio, apresenta uma região onde informações são

processadas (corpo celular, soma), algumas entradas para a recepção de estímulos (os dendritos) e um prolongamento mais longo chamado de axônio (que estimula os neurônios subsequentes).

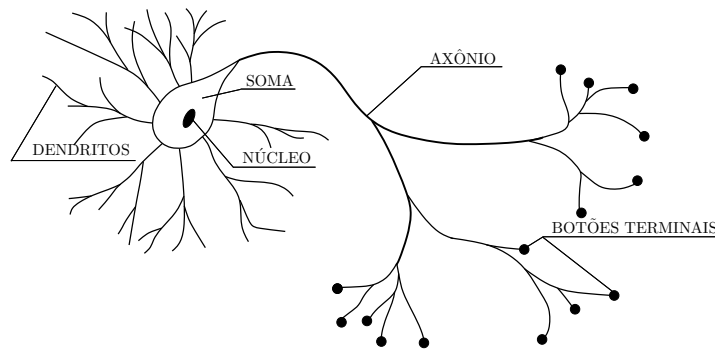


Figura 2.3: Representação Básica do Neurônio Biológico
 FONTE: Medeiros (2006)

Os dendritos ligam-se ao redor da célula a outras células e o axônio faz uma conexão mais longa. A estas conexões dá-se o nome de sinapses (MEDEIROS, 2006). É através das sinapses que os estímulos nervosos se propagam. Cada neurônio pode ser estimulado ou inibido por grande quantidade de neurônios, através de suas sinapses. O somatório dos estímulos e inibições determinará se o neurônio transmitirá o impulso ou não.

Os neurônios de uma RNA, como pode ser observado na FIGURA 2.4 devem estar conectados entre si. Os neurônios são dispostos em camadas, onde os neurônios de uma mesma camada normalmente se comportam da mesma maneira.

Um neurônio artificial (FIGURA 2.4) possui um número n de entradas x_1, x_2, \dots, x_n (dendritos). Cada sinal de entrada é multiplicado por um peso w_1, w_2, \dots, w_n , podendo ser negativo ou positivo. A seguir, a soma ponderada dos sinais de entrada aplicado ao neurônio, é executada. É aplicado o *bias* (θ), cuja função é aumentar o número de graus de liberdade do modelo, aumentando a capacidade da rede de ajustar-se ao conhecimento a ela fornecido. A soma é processada pela função de ativação para produzir a saída Y do neurônio (axônio).

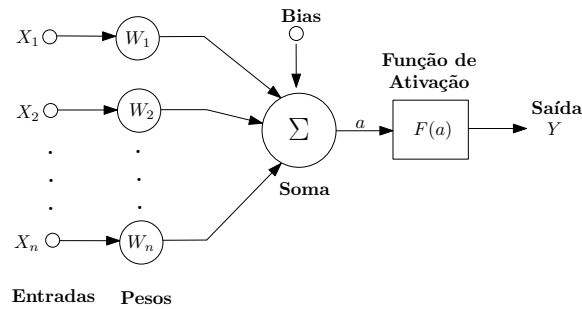


Figura 2.4: Exemplo de Neurônio Artificial

FONTE: Medeiros (2006)

A disposição dos neurônios nas camadas e o padrão de conexão entre estas definem a arquitetura da RNA. As redes sem realimentação (*feedforward*) têm neurônios agrupados em camadas onde o sinal percorre a rede em uma única direção, da entrada para a saída. Os neurônios da mesma camada não são conectados. Nas redes com realimentação ou recorrentes (*recurrent*), a saída de alguns neurônios alimentam neurônios da mesma camada (inclusive o próprio) ou de camadas anteriores, o sinal percorre a rede em duas direções, tem memória dinâmica e capacidade de representar estados em sistemas dinâmicos, um exemplo é a rede de Hopfield (1982).

2.5.2 Função de Ativação de uma Rede Neural

A função de ativação é muito importante para o comportamento de uma rede neural porque é ela que define a saída do neurônio artificial e, portanto, o caminho pelo qual a informação é conduzida (STEINER, 2006).

É desejável que uma função de ativação seja contínua e diferenciável, em virtude da necessidade da utilização de sua derivada.

As funções de ativação podem ser de vários tipos, sendo que as mais utilizadas são as funções gaussianas, sigmóides, senoidais, lineares, tangentes hiperbólicas, logarítmicas, degrau, entre outras como pode ser observado na FIGURA 2.5.

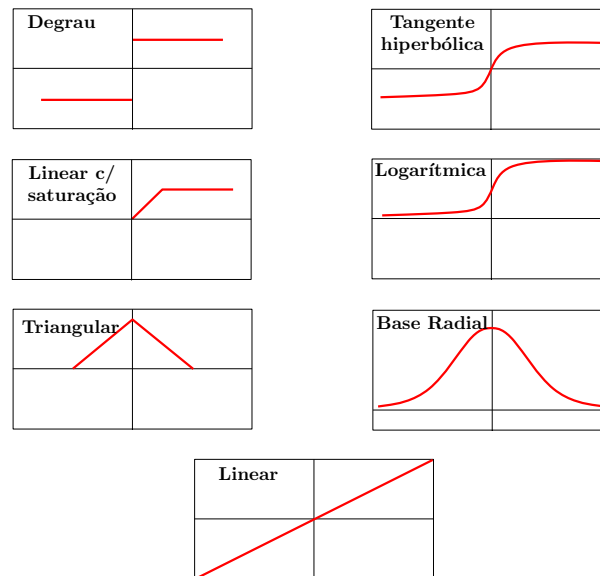


Figura 2.5: Funções de Ativação

FONTE: Lima (2006)

Normalmente, as funções não-lineares são as mais utilizadas. Outra prática comum é a de utilizar uma mesma função de ativação para todos os elementos de uma mesma camada em uma rede e, até mesmo, para toda a rede, entretanto, não é uma regra.

2.5.3 Aprendizado ou Treinamento de uma Rede Neural

O aprendizado consiste na modificação dos pesos das conexões entre os neurônios, os pesos iniciais (sinapses) são modificados de forma iterativa, por um algoritmo que segue um dos seguintes paradigmas de aprendizado:

- **Aprendizado Supervisionado:** a descrição de cada exemplo de treinamento traz também a classe que o exemplo representa. Durante a fase de treinamento, o algoritmo tenta ajustar os pesos das conexões de maneira que a saída da rede coincida com a classe associada ao exemplo, para cada exemplo do conjunto de treinamento.

- **Aprendizado por Reforço:** para cada entrada apresentada, é produzida uma indicação (reforço) sobre a adequação das saídas correspondentes produzidas pela rede, ou seja um crítico externo avalia a resposta fornecida pela rede.
- **Aprendizado Não-supervisionado:** a rede é treinada sem considerar a informação da classe associada a cada exemplo de treinamento. O aprendizado não-supervisionado funciona por meio da modificação de pesos de conexões, de maneira que exemplos semelhantes entre si sejam associados à mesma unidade de saída.

O algoritmo padrão de uma RNA com aprendizado supervisionado segue os passos do fluxograma da FIGURA 2.6.

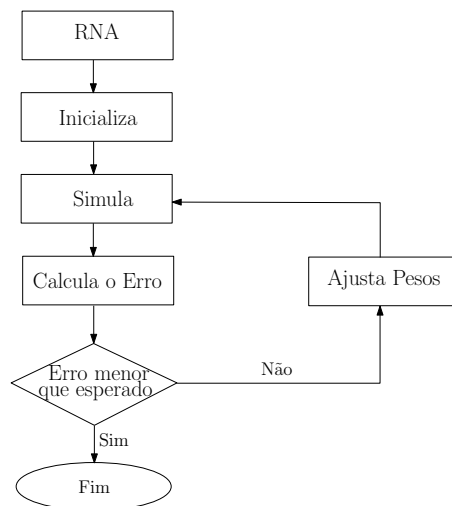


Figura 2.6: Algoritmo genérico de uma RNA supervisionada

FONTE: Lima (2006)

2.5.4 Fluxo de Dados em uma Rede Neural

A direção em que o fluxo de ativações acontece caracteriza as RNA's em:

1. *feedforward* (acíclica): as conexões não formam ciclos. São rápidas para fornecer resposta a uma entrada. O *back-propagation* é um método numérico eficiente para o treinamento das redes *feedforward*.

2. feedback (cíclica): as conexões podem formar ciclos. Cada vez que uma entrada é apresentada à rede, esta executa um processo iterativo antes de produzir a resposta. São redes mais difíceis de serem treinadas.

2.5.5 Rede Neural Bayesiana

Neste trabalho, o software “Statgraphics” (que faz a análise de dados) define que a rede neural que melhor se encaixa ao tipo dos dados presentes é a Rede Neural Bayesiana (RNB) (JENSEN, 1996).

As RNB's são constituídas da fusão das redes neurais convencionais e da estatística Bayesiana, fazendo com que esta abordagem (neural e estatística) produza um modelo robusto com a potencialidade das redes neurais e a solidez dos métodos estatísticos (HUARSAYA, 1999). O uso da RNB permite uma modelagem de dados em dois níveis de inferência:

- Cálculo de parâmetros (pesos e bias) e hiperparâmetros (distribuição de probabilidade e nível de ruído) da rede neural (para ajustar o modelo aos dados de treinamento);
- Comparação entre modelos (fornecer preferência por determinado modelo em função da complexidade).

Uma das vantagens dessa rede neural é poder resolver o problema de convergência para poucos dados de treinamento; fácil entendimento do modelo, cujas relações causais entre as variáveis são em grande parte intuitivas; eficiência computacional, visto que a maioria dos algoritmos de redes bayesianas são polinomiais; facilita a medida da incerteza associada aos processos; prevê informações sobre o efeito de possíveis intervenções.

3 MATERIAL E MÉTODO

Neste capítulo, apresenta-se a organização dos dados utilizados e os procedimentos realizados no desenvolvimento da pesquisa. Os dados são discriminados adiante.

3.1 OBTENÇÃO DOS DADOS

Os dados utilizados para a análise foram coletados através do questionário Sócio-educacional, dados do vestibular aplicado na ficha de inscrição do vestibular da UFPR 2005 - 2008 e histórico direcionado somente a disciplina de Cálculo I e II, sendo utilizado somente os dados dos alunos que se matricularam nos cursos de Engenharia Civil, Engenharia Elétrica e Engenharia Mecânica. As informações foram levantadas (coletadas) entrando em contato com o Núcleo de Concurso (NC) e as Coordenações (Co) dos cursos referidos da UFPR.

3.1.1 Dados fornecidos pelo núcleo de concurso

O núcleo de concurso é um departamento da UFPR, responsável por inúmeros concursos públicos, principalmente no estado do Paraná, inclusive o vestibular, então através do NC foi possível obter todos os dados dos vestibulando até a sua matrícula em seus referidos cursos, os dados obtidos foram através de planilhas excel contendo as seguintes informações: matrícula, nome, curso, nome do curso, sexo, nascimento, nota da 1ª fase incluindo as disciplinas de matemática, biologia, química, geografia, física, português, história, língua estrangeira, nota da 2ª fase incluindo as disciplinas

matemática da 2ª fase e física da 2ª fase, status e descrição da forma de entrada. Esses dados também foram compostos pelo questionário Sócioeducacional apresentado no Apêndice A.

Dentre as informações obtidas pelo NC, questões consideradas relevantes para o trabalho foram as seguintes: nome, curso, sexo, nascimento, nota da 1ª fase, nota da 2ª fase e descrição da forma de entrada, quanto ao questionário sócioeducacional referente ao Apêndice A foram consideradas: Qual o seu estado civil? Qual a sua situação quanto à moradia? Qual o nível de instrução do seu pai? Qual o nível de instrução da sua mãe? A renda total mensal de sua família se situa na faixa? Durante o curso, você terá obrigatoriamente que trabalhar? Como fez seus estudos do Ensino Fundamental e médio? Qual o motivo que o levou a escolher o curso para o qual está se candidatando? Quanto à sua escolha pelo curso, você se considera: O que você espera, em primeiro lugar de um curso universitário?

3.1.2 Dados fornecidos pelas coordenações

Os cursos de Engenharia Civil, Engenharia Elétrica e Engenharia Mecânica pertencem ao departamento de Setor de tecnologia, suas respectivas coordenações são responsáveis pela orientação, coordenação, supervisão e certificação das ações extensionistas realizadas pelas diferentes unidades administrativas que compõem a UFPR. Com os respectivos coordenadores obteve-se as seguintes informações: matrícula, nome, situação da matrícula, código da disciplina, nome da disciplina, referência de semestre, situação referente a disciplina, média e número de faltas. Para o trabalho foi relevante a matrícula, nome, código da disciplina, nome da disciplina, situação referente a disciplina.

3.1.3 Organização dos dados

Os dados relevantes para a pesquisa foram organizados e reestruturados formando 25 dados chamados de variáveis disponibilizadas da seguinte forma: conclusão, sucesso, gênero, idade, nota_1fase, nota_2fase, cardinalidade ¹ da disciplina de cálculo I, cardinalidade da disciplina de cálculo II, cota afrodescende (Cota-afro), cota escola pública (Cota EscPub), chamada complementar(ChamComp), concorrência geral (ConcGeral), engenharia civil (EngCivil), engenharia elétrica (EngEletrica), engenharia mecânica (EngMec), estado civil, moradia, grau de instrução do pai, grau de instrução da mãe, renda familiar, trabalhar durante o curso, procedência educacional, escolha do curso, motivo da escolha do curso, o que espera do curso.

No entanto as onze variáveis do questionário Sócioeducacional como: estado civil, moradia, grau de instrução do pai, grau de instrução da mãe, renda familiar, trabalhar durante o curso, procedência educacional, escolha do curso, motivo da escolha do curso, o que espera do curso são desmembradas em outras variáveis, por exemplo a variável estado civil gera no mínimo duas outras variáveis dadas por solteiro(a), casado(a) e outros, ou ainda como a variável moradia que ficou da seguinte forma: mora casa dos pais (quitada ou financiada), mora casa dos pais (alugada), mora casa (própria, quitada ou financiada), mora casa alugada (paga pelo aluno), mora casa (estudante, pensão ou pensionato), mora em casa (parentes ou amigos), mora em casa alugada (paga por seus pais). Desta forma o conjunto inicial com 24 variáveis transforma-se em um conjunto de 82 variáveis como descritas no Apêndice B.

Estas variáveis foram utilizadas para o ano de 2005 totalizando 229 alunos sendo considerados como observações, separados em dois grupos Cotista e não-Cotista, esse agrupamento é válido para estatística descritiva e indutiva. Para a análise de Regressão Logística e Rede Neural os dados referem-se aos alunos que ingressaram nos anos de 2005, 2006, 2007 e 2008 totalizando 1213 alunos.

¹ número de vezes que cursou a disciplina

O quadro de tabulação, referente aos valores relevantes da pesquisa apresenta em sua maioria dados dicotômicos ² representados com (0 ou 1) e que naturalmente tem distribuição Bernoulli, $b(1, \theta)$ com $0 < \theta < 1$.

3.2 DISCRIMINAÇÃO DAS VARIÁVEIS

Segue a discriminação das variáveis Conclusão, Sucesso, Gênero, Idade, nota_1-fase, nota_2fase, frequência que cursou a disciplina Cálculo I, frequência que cursou a disciplina Cálculo II, tipo de entrada, estado civil, moradia, grau de instrução do pai, grau de instrução da mãe, renda mensal familiar, necessidade de trabalhar, procedência educacional, escolha do curso, motivo da escolha do curso e o que espera do curso³.

3.2.1 Variável Conclusão

Representa a conclusão das disciplinas de Cálculo I e II entre os anos de 2005 a 2008.

Esta variável é representada por uma distribuição de Bernoulli, sendo discreta com contradomínio $\{0, 1\}$.

3.2.2 Variável Sucesso

Representa a conclusão das disciplinas de Cálculo I e Cálculo II dentro de um determinado período, portanto além de concluinte é necessário ter cursado apenas três vezes estas disciplinas, no total.

²que são originados de questões do tipo verdadeiro o falso; sim ou não; concordo ou não concordo

³Variáveis com distribuição de Bernoulli foi considerado 1 como sucesso e 0 insucesso.

3.2.3 Variável Gênero

Representa o gênero dos alunos classificado como feminino e masculino com o objetivo de reconhecer estatisticamente a proporção entre os gêneros.

3.2.4 Variável Idade

Representa a idade com que os alunos ingressaram nos cursos citados, e ter conhecimento da idade média do aluno ingressante na UFPR dos cursos em foco.

3.2.5 Variável Nota Primeira Fase

Representa a nota que o aluno tirou na 1ª fase do vestibular.

A primeira fase do vestibular na UFPR, é constituída de uma prova com questões objetivas, no formato de múltipla escolha, sobre os conteúdos do Ensino Médio.

3.2.6 Variável Nota Segunda Fase

Representa a nota que o aluno tirou na 2ª fase do vestibular.

Para a 2ª fase são convocados os candidatos melhor classificados na 1ª fase, independente dele ter optado ou não pelas vagas de inclusão racial e social, já na classificação da 2ª fase a opção das vagas de inclusão são levadas em consideração.

3.2.7 Variável Frequência que Kursou a Disciplina Cálculo I

Representa a quantidade de vezes que os alunos fizeram a disciplina de Cálculo I sem considerar a informação da aprovação na mesma. Com o objetivo de constatar quantas vezes o aluno kursou a disciplina de Cálculo I desde o ano de 2005 até o ano de 2008.

3.2.8 Variável Frequência que Coursou a Disciplina Cálculo II

Representa a quantidade de vezes que os alunos fizeram a disciplina de Cálculo II, sem considerar a aprovação na mesma. Com o objetivo de constatar quantas vezes o aluno cursou a disciplina de Cálculo II desde o ano de 2005 até o ano de 2008.

3.2.9 Variável Tipo de Entrada

Representa a forma de entrada na UFPR. No vestibular referente ao ano de 2005 existiam quatro formas de ingresso na UFPR, FIGURA 3.1 sendo classificados nas seguintes categorias: cota afro e cota escola pública ⁴, concorrência geral e chamada complementar ⁵.



Figura 3.1: Distribuição do candidato mediante entrada

Para as próximas variáveis as informações foram retiradas do questionário Sócioeducacional, as questões com suas respectivas alternativas e suas siglas estão discriminadas, onde “C” e “Nc” no final da sigla indica o grupo Cotista e não-Cotista respectivamente.

⁴Alunos Cotistas

⁵Alunos que não ingressaram na 1ª chamada são reclassificados na listagem de concorrência geral, ou seja, considerados não-Cotista

3.2.10 Variável Estado Civil

Representa o estado civil, com a seguinte discriminação solteiro(a), casado(a) e outros.

3.2.11 Variável Moradia

Representa a condição de Moradia.

É a variável que explica se a moradia dos Universitários/Pais é própria, alugada, financiada ou cedida. No questionário Socioeducacional é apresentada com as seguintes alternativas:

10.1 Mora em casa dos pais, quitada ou financiada - csPqC e csPqNc;

10.2 Mora em casa dos pais, alugada - csPaC e csPaN;

10.3 Mora em casa própria, quitada ou financiada - csEpC e csEpNc;

10.4 Mora em casa alugada, paga por você - csEaC e csEaNc;

10.5 Mora em república, casa de estudante, pensão ou pensionato - csEstPenC e csEstPenNc;

10.6 Mora em casa de parentes ou amigos - csParenAmigC e csParenAmigNc;

10.7 Mora em casa alugada para você, paga por seus pais - csAPC e csAPNc.

3.2.12 Variável Grau de Instrução do Pai

Representa o nível de instrução do pai. Esta variável diz respeito às relações familiares de acordo com a situação de dependência/independência econômica do candidato em relação aos pais e se ele já está inserido no mercado de trabalho ou tem essa necessidade para continuar o estudo universitário, apresentado com os seguintes itens:

11.1 Sem escolaridade - seC e seNc;

11.2 Ensino fundamental incompleto - fiC e fiNc;

- 11.3 Ensino fundamental completo - fcC e fcNc;
- 11.4 Ensino médio incompleto - emiC e emiNc;
- 11.5 Ensino médio completo - emcC e emcNc;
- 11.6 Superior incompleto - siC e siNc;
- 11.7 Superior Completo - scC e scNc
- 11.8 Não sabe informar - nsC e nsNc.

3.2.13 Variável Grau de Instrução da Mãe

Representa o nível de instrução da mãe. Com a variável “Grau de Instrução da Mãe” estuda-se o indicativo que a escolaridade da mãe incide no desempenho dos filhos no vestibular, entre os grupos, sendo discriminada com as seguintes alternativas:

- 12.1 Sem escolaridade - seC e seNc;
- 12.2 Ensino fundamental incompleto - fiC e fiNc;
- 12.3 Ensino fundamental completo - fcC e fcNc;
- 12.4 Ensino médio incompleto - emiC e emiNc;
- 12.5 Ensino médio completo - emcC e emcNc;
- 12.6 Superior incompleto - siC e siNc;
- 12.7 Superior Completo - scC e scNc;
- 12.8 Não sabe informar - nsC e nsNc.

3.2.14 Variável Renda Mensal Familiar

Indica condição econômica familiar. A “Renda Mensal Familiar” foi utilizada para analisar a categoria da “situação Sócioeconômica” dos universitários. No questionário Sócioeducacional a questão é apresentada com as seguintes alternativas:

- 13.1 Até R\$ 260,00 - ate260C e ate260Nc;
- 13.2 de R\$ 261,00 a R\$ 500,00 - 261-500C e 261-500Nc;
- 13.3 de R\$ 501,00 a R\$ 1.000,00 - 501-1000C e 501-1000Nc;

- 13.4 de R\$ 1.001,00 a R\$1.500,00 - 1001-1500C e 1001-1500Nc;
- 13.5 de R\$ 1.501,00 a R\$ 2.000,00 - 1501-2000C e 1501-2000Nc;
- 13.6 de R\$ 2.001,00 a R\$ 3.000,00 - 2001-3000C e 2001-3000Nc;
- 13.7 de R\$ 3.001,00 a R\$ 4.000,00 - 3001-4000C e 3001-4000Nc;
- 13.8 de R\$ 4.001,00 a R\$ 5.000,00 - 4001-5000C e 4001-5000Nc;
- 13.9 Acima de R\$ 5.001,00 - ac5000C e ac5000Nc.

3.2.15 Variável Necessidade de Trabalhar

Representa se o aluno precisa sustentar-se. A variável “Necessidade de Trabalhar” verifica se os estudantes de graduação terão chances de concluir o curso se dedicando apenas ao estudo. Para analisar esta variável no questionário sócio educacional tem-se os seguintes itens:

- 14.1 Trabalhar - apenas nos últimos anos - TuIC e TuINc;
- 14.2 Trabalhar - desde 1º ano, em tempo parcial - T1ºpC e T1ºpNc;
- 14.3 Trabalhar desde 1º ano, em tempo integral - T1ºiC e T1ºiNc;
- 14.4 não sabe se vai trabalhar - nsC e nsNc;
- 14.5 não vai trabalhar - nvC e 1501-2000Nc.

3.2.16 Variável Procedência Educacional

Representa a origem escolar e base educacional. A seleção atual para se ingressar na UFPR, baseia-se em uma avaliação dos conhecimentos adquiridos pelos candidatos ao longo de sua formação no ensino fundamental e médio. Em relação à origem dos estudantes, os alunos foram distribuídos em cinco grupos:

- 15.1 Estudou Integralmente em Escola Pública (fundamental e médio) - esclpuC e esclpuNc;
- 15.2 Estudou Integralmente em Escola Particular (fundamental e médio) - esclpaC e esclpaNc;

15.3 Estudou Maior parte em Escola Pública (fundamental e médio) - escPpuC e escPpuNc;

15.4 Estudou Maior parte em Escola Particular (fundamental e médio) - escPpaC e escPpaNc;

15.5 Estudou escolas comunitárias ou outros (fundamental e médio) - escComC e escComNc.

3.2.17 Variável Escolha do Curso

Representa a segurança com que o aluno escolheu o curso. Uma das características marcantes do profissional da área de Engenharia é ter inclinação para se aprofundar na matemática, portanto é interessante que o aluno já tenha tomado uma decisão quanto a sua escolha, para que durante o período da graduação, tenha condições de explorar as áreas desejadas e aproveitar ao máximo as oportunidades que a UFPR proporciona.

Para avaliar esta variável optou-se como possível resposta as seguintes alternativas:

16.1 Absolutamente decidido - adC e adNc;

16.2 Muito decidido - mdC e mdNc;

16.3 Decido - dC e dNc;

16.4 Indeciso - iC e iNc;

16.5 Muito indeciso - miC e miNc.

3.2.18 Variável Motivo da Escolha do Curso

Representa a motivação quanto a escolha do curso. Esta seção, foi pesquisada para se identificar e analisar a motivação na escolha dos cursos de Engenharia Civil, Engenharia Elétrica e Engenharia Mecânica da UFPR, no ano de 2005, e assim, compreender os motivos e fatores que interferem na escolha do curso superior de gradua-

ção, sendo discriminada nas seguintes opções:

- 17.1 Mercado de trabalho e salário - mtC e mtNc;
- 17.2 Possibilidade de contribuir para a sociedade - csC e csNc;
- 17.3 realização pessoal - rpC e rpNc;
- 17.4 Gosto pela profissão - gpC e gpNc;
- 17.5 Gosto pelas matérias do curso - gmC e gmNc;
- 17.6 Baixa concorrência - bcC e bcNc;
- 17.7 Permite conciliar aula e trabalho - pcC e pcNc;
- 17.8 Outro motivo - omC e omNc.

3.2.19 Variável O que Espera do Curso

O objetivo dos Cursos de Engenharia Civil, Engenharia Elétrica e Engenharia Mecânica proposto pela UFPR é formar engenheiros capacitados a atender diferentes solicitações profissionais com visão crítica, criativa e inovadora. Analisando a relação entre o objetivo do aluno e o da universidade, na variável “O que Espera do Curso”, pretende-se estudar a expectativa do estudante ao ingressar no curso de Engenharia da UFPR. O estudo foi feito nas seguintes categorias, para esta variável:

- 18.1 Formação profissional, voltada para mercado de trabalho - acC e acNc;
- 18.2 Formação teórica, voltada a pesquisa - fpC e fpNc;
- 18.3 Formação acadêmica para melhorar a prática - ipC e ipNc;
- 18.4 Compreender melhor o mundo - fpC e fpNc;
- 18.5 Melhorar o meu nível de instrução - niC e niNc;
- 18.6 Diploma de nível superior - dsC e dsNc.

3.3 MÉTODOS

A análise do perfil foi realizada com base nos dados fornecidos pelo NC e Co re-organizados. Estes dados contém informações tais como dados pessoais, nota da 1ª

e 2ª fase, número de vezes que cursou a disciplina de Cálculo I e II, alunos que ingressaram por cota ou não, opção de curso, condições de moradia, grau de instrução dos pais, renda familiar, necessidade de trabalhar durante o curso, procedência educacional, motivação pela escolha do curso e a expectativa com relação a graduação, foram tabulados em planilhas excel, representando uma matriz com 82 variáveis x 1213 observações, referente ao ano base de 2005.

A pesquisa inicialmente projetada pretendia investigar a separação entre grupos tais como: (i) possibilidade de desistência das disciplinas em questão, (ii) possibilidade do término do curso no tempo previsto, (iii) possibilidade de término do curso em atraso e (iv) possibilidade de evasão do curso, utilizando como técnica análise fatorial e componentes principais. Entretanto com a verificação da não normalidade da matriz 82 x 1213, foi necessário alterar o projeto inicial, passando a trabalhar variável a variável com estatística indutiva e descritiva separando em dois segmentos Cotista e não-Cotista. Para isto aplicou-se o teste de proporções, constando a igualdade ou a diferença entre as proporções baseado no p -valor superior a 0,05. Também foi possível verificar o mínimo, o máximo, a média e os quartis dos dados. Mesmo trabalhando variável a variável constatou-se a não normalidade das mesmas, sendo assim, aplicou-se alguns testes para uma melhor descrição. Dentre estes testes destaca-se o teste de Levene utilizado para verificar a igualdade de variância em dados não normalizados, o teste-F para as variáveis com distribuição normal e o teste de Mann-Whitney testando se as amostras possui que médias iguais sem a necessidade de constatar a normalidade e igualdade de variância.

Além de utilizar a estatística básica, também foram utilizadas as técnicas de regressão logística e de redes neurais em uma comparação entre elas. Com o objetivo de selecionar as variáveis para a aplicação das técnicas de regressão logística e rede neural foram utilizadas variáveis com p -valores maiores que 0,05. Desta forma, utilizou-se dez variáveis (Nota_2fase, Cálculo I, Cálculo II, sexo, Pai_Ensino fundamental incompleto, Pai_Ensino fundamental completo, Pai_Ensino médio com-

pleto, Mãe_Superior completo, Escolha_Indeciso, Espera_Melhorar o meu nível de instrução) para escrever o modelo de regressão logística tendo como variável dependente “Conclusão”. Outras dez variáveis (nota_2fase, Cálculo I, Pai_Ensino fundamental incompleto, Pai_Ensino fundamental completo, Pai_Ensino médio completo, Mãe_Superior completo, trabalhar desde do 1ºano_em tempo integral, motivo_Mercado de trabalho e salário, Espera_compreender melhor o mundo, Espera_Melhorar o meu nível de instrução) deixa como variável dependente a variável “Sucesso”. Para escrever a rede neural foram utilizadas as mesmas variáveis.

Os cálculos realizados para estatística indutiva e descritiva foram executados no software Minitab e a técnica de regressão logística e rede neural foram executados no software Statgraphics.

No capítulo 4 são discutidos os resultados obtidos, utilizando a metodologia e os dados que foram descritos neste capítulo.

4 RESULTADOS E DISCUSSÃO

Este capítulo refere-se a descrição e discussão dos resultados obtidos.

4.1 APRESENTAÇÃO E DESCRIÇÃO DOS DADOS

Segue a análise estatística descritiva das variáveis discriminadas na seção 3.2.

4.1.1 Variável Conclusão

Para esta variável testou-se a hipótese de igualdade na proporção de concluinte entre Cotista e não-Cotista, ou seja, a hipótese nula:

Hipótese nula $\rightarrow H_0 : P_c = P_{nc}$

Hipótese alternativa $\rightarrow H_1 : P_c \neq P_{nc}$

Nível de decisão alfa bilateral = 0,05

aplicou-se o teste z de estatística conforme apresentada no QUADRO 4.1 e usando o software Minitab, para testar a hipótese nula H_0 obtendo-se:

Resultados obtidos no QUADRO 4.1:

Variável	Sucesso	Total	Proporção
Conclusão_C	32	59	0.542373
Conclusão_Nc	99	170	0.582353
Diferença = p (Conclusão_C) - p (Conclusão_Nc)			
Estimativa para a diferença : -0.0399801			
Intervalo de confiança de nível 95% para a diferença: (-0.187141; 0.107181)			
Teste para diferença = 0 (diferente de 0): $z = -0.53$ p -Valor = 0.594			

QUADRO 4.1 - Resultado Conclusão

FONTE: O autor

As proporções observadas no QUADRO 4.1 mostram que 54,23% dentre 59 Cotistas são concluintes e 58,23% dentre os 170 não Cotistas são concluintes. O valor $p = 0,594$ fornecido pelo teste z indica que a diferença na proporção entre as duas amostras não é significativa pois $p > 0,05$, não existindo influência no fato do aluno ser ou não ser Cotista para conclusão nos cursos referidos neste trabalho.

Portanto para a variável Conclusão deve-se aceitar a hipótese nula H_0 , ou seja, não existe diferença entre os Cotistas e não-Cotistas na conclusão das disciplinas.

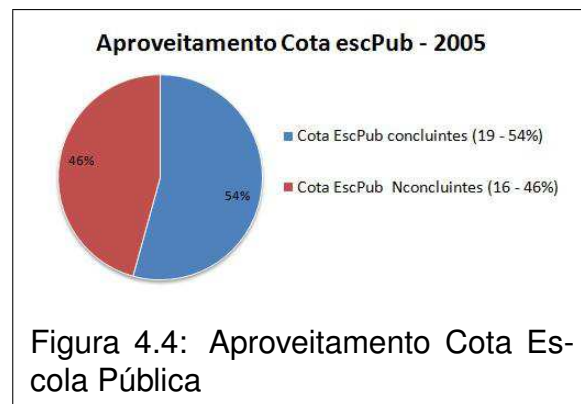
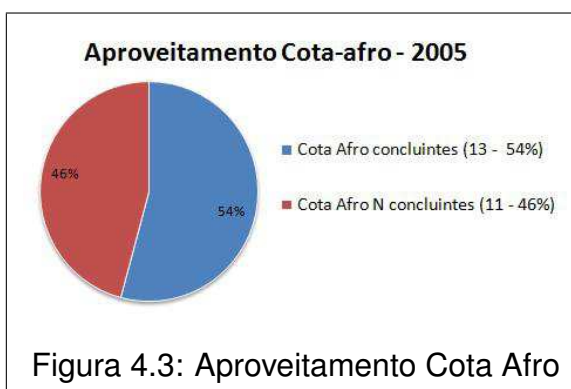
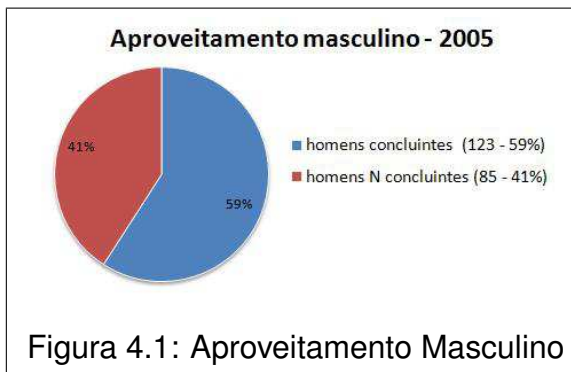
Esse procedimento de teste foi aplicado nas outras seguintes variáveis: Conclusão Masculino, Conclusão feminino, Conclusão cota-afro, conclusão cota escPub. Os resultados estão na TABELA 4.1, sendo que P_{HC} representa proporção de homens Cotista; P_{HNc} representa proporção de homens não-Cotista; P_{MC} representa mulheres cotista; P_{MNc} representa proporção de mulheres não-Cotista; P_{CAfro} representa proporção de Cotista afrodescendente e P_{Nc} proporção não-Cotista.

Tabela 4.1: Conclusão nas disciplinas Cálculo I e II

Hipótese	Est. do teste Z	valor p
$H_{01} : P_{HC} = P_{HNc}$	-0,35	0,727
$H_{02} : P_{MC} = P_{MNc}$	-0,58	0,565
$H_{03} : P_{CAfro} = P_{Nc}$	-0,37	0,708
$H_{04} : P_{CEscPub} = P_{Nc}$	-0,43	0,669

FONTE: O autor(2010)

Como pode-se observar não houve diferença estatística significativa em nenhum dos subgrupos, pois os valores p apresentados foram todos maiores que 0,05. Também é possível visualizar estas informações através dos gráficos de setores representado, nas FIGURAS (4.1 - 4.4).



4.1.2 Variável Gênero

No total de 59 alunos Cotistas, 6 alunos são do sexo feminino correspondendo aproximadamente a 10,17% e 53 do gênero masculino correspondendo 89,83%. Para o grupo não-Cotistas composto por 170 estudantes, 15 são do sexo feminino representando 8,82% e 155 masculino correspondendo aproximadamente a 91,18%, ver FIGURA 4.5.

Utilizando o teste para comparação entre duas proporções comparando o grupo de Cotistas e não-Cotistas obteve-se o valor $p = 0,765$ superior a 0,05. Sendo assim, rejeita-se a hipótese de igualdade entre as proporções de Cotistas e não-Cotista, não havendo diferença estatisticamente significativa entre o número de homens e mulheres aprovados em cada grupo. Considerando apenas o gênero, as mulheres tem índice de aprovação de 76% e os homens 62%.

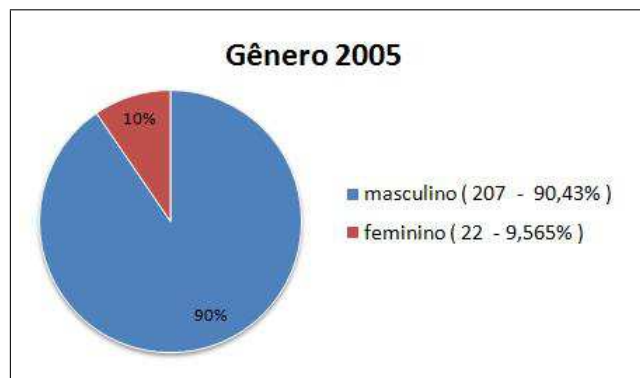


Figura 4.5: Gênero em relação a entrada

4.1.3 Variável Idade

A idade média no grupo dos Cotistas foi de 19 anos, sendo que 25% dos alunos entraram com idade mínima de até 18 anos e 75% tem idade de 19 a 20 anos chegando a idade máxima de 25 anos. O grupo dos alunos não-Cotista difere na idade média sendo esta de 18 anos atingindo a idade máxima de 23 anos. Ocorre uma maior variabilidade entre as observações ocorridas no grupo dos Cotistas com desvio padrão de

1,937 e desvio padrão de 0,951 para os não-Cotistas.

As variáveis “idade_C” e “idade_Nc” não são normalmente distribuídos com $p = 0,048$ e $p = 0,150$ respectivamente, portanto foi utilizado o teste não-paramétrico Mann-Whitney destinada a comparar duas amostras independentes, baseando-se nesse teste, aceita-se a hipótese de nulidade ($p = 0,001$), ou seja, não há diferença nas idades entre os Cotistas e não-Cotistas.

4.1.4 Variável Nota Primeira Fase

As variáveis envolvidas são Nota_1fase_C e Nota_1fase_NC com apenas dois grupos: Cotista e não-Cotista. Testando a normalidade das variáveis citadas, pode-se afirmar que a primeira variável tem distribuição normal ($p = 0,019$) e a segunda variável nega a hipótese de normalidade ($p = 0,15$). Utilizando o teste não-paramétrico Mann-Whitney, cujo resultado aponta para a igualdade de proporções entre os grupos, mostra-se que existe diferença significativa entre os grupos $p = 0,0001$, mostrando que o fato de aluno não ser Cotista influencia na nota da 1ª fase. O pressuposto de independência entre os grupos é automático, pois são grupos vindos de populações mutuamente excludentes: Cotista e não-Cotista.

A nota média geral para a classificação para a segunda fase do vestibular da UFPR para os cursos citados são: grupo dos Cotistas, aproximadamente de 51,35 e grupo dos não-Cotistas 56,77, sendo que 25% entrou com nota 49 e 52 e 75% com nota 56 e 62, respectivamente. As duas variáveis apresentam a mesma variabilidade, coeficiente de variabilidade 13,51% e 13,74%.

4.1.5 Variável Nota Segunda Fase

As variáveis consideradas são nota_2fase_C para alunos Cotistas e nota_2fase_NC para alunos não-Cotistas. A nota geral média para a classificação nesta fase foi de 561,3 para Cotistas e 653,58 para não-Cotistas, sendo a nota mínima 260,2

e a máxima 753,8 entre os Cotistas, e mínima 487,88 e a máxima de 896,21 entre os não-Cotista, existe diferença entre a nota mínima e máxima no grupos, porém a variável nota_2fase_C apresenta uma maior variabilidade entre suas observações com coeficiente de variabilidade de 16,74% em comparação com a segundo grupo 13,86%, havendo uma maior discrepância entre as variáveis. A primeira variável é menos homogênea que a segunda, apresentando desvio padrão 94 enquanto que a segundo grupo tem desvio padrão de 90,57 ambas com distribuição não-normal, com $p = 0,01$. Dessa forma utilizou-se o teste não-paramétrico Mann-Whitney apresentando $p = 0,0001$, portanto, deve rejeitar-se a hipótese nula de que as duas amostras tem a medianas iguais, ou seja, deve concluir-se que o fato do aluno ser Cotista interfere na nota da segunda fase.

4.1.6 Variável Frequência que cursou a disciplina Cálculo I

As variáveis consideradas são CálculoI-C dos alunos Cotistas e CálculoI-Nc dos alunos não-Cotistas. O gráfico da FIGURA 4.6 mostra a quantidade de vezes que o aluno cursou a disciplina nos grupos.

Como resultado obteve-se que a média da quantidade de vezes que o aluno cursou esta disciplina foi de 1,678 para Cotistas e 1,5294 para não-Cotista, com a quantidade mínima de 1 e a máxima de 6 em ambas as amostras, sendo que até 25% cursaram a disciplina apenas um vez, 75% cursou até 2 vezes e apenas 25% cursou a disciplina mais de duas vezes.

A variável “Frequência que cursou a disciplina Cálculo I” não segue uma distribuição normal, com $p = 0,140$ e $p = 0,01$ respectivamente, portanto utilizou-se o teste Mann-Whitney verificando a igualdade das medianas. Com $p = 0,3634$ pode-se afirmar que possui medianas iguais, indicando que o fato de aluno ser ou não Cotista não interfere na quantidade de vezes que aluno cursará a disciplina de Cálculo I.

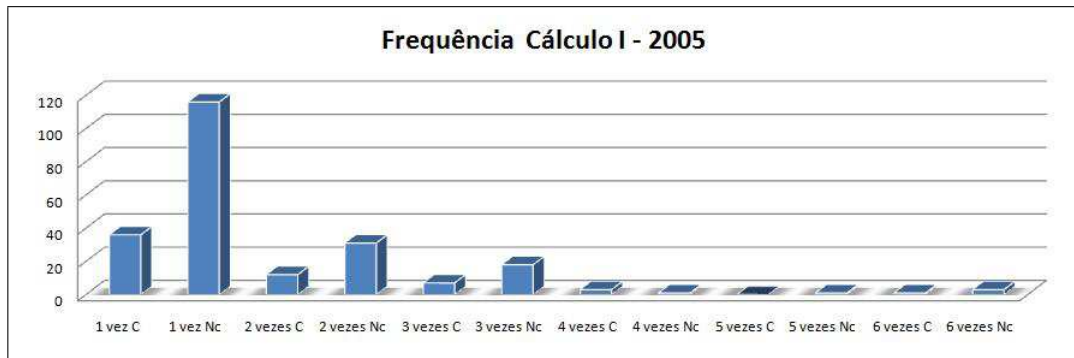


Figura 4.6: Frequência com que os alunos cursaram a disciplina de Cálculo Diferencial e Integral I

4.1.7 Variável Frequência que cursou a disciplina Cálculo II

Para a análise foram consideradas CálculoII_C dos alunos Cotistas e CálculoII_Nc dos alunos não-Cotistas. Como descrição estatística pode-se informar que a média da quantidade de vezes que os alunos cursaram esta disciplina são de 1,203 para Cotistas e 1,2529 para não-Cotista, ocorrendo maior dispersão entre as observações no grupo dos alunos não-Cotista com coeficiente de variância de 74,72 enquanto que no grupo dos Cotistas apresenta coeficiente de variância igual a 66,87. Para esta variável também somente 25% dos alunos cursaram mais de duas vezes a disciplina de Cálculo II, podendo chegar a máxima de 4 vezes para Cotista e 5 vezes para não-Cotista.

A variável “Variável Frequência que cursou a disciplina Cálculo II” possui distribuição normal, ambas com $p = 0,15$. Testando a variância entre as amostras, o teste-F é significativo entre os grupos ($F = 44,564, p = 0.0001$). A comparação entre as médias dos grupos exibe diferenças significativas entre os grupos de Cotistas e não-Cotistas. Pode-se concluir, assim, que o fato de o aluno ser Cotista ou não, interfere na quantidade de vezes que o aluno cursou a disciplina de Calculo II.

O gráfico da FIGURA 4.7 pode-se observar a frequência com que os alunos cursaram a disciplina de Cálculo II, por grupo, ou seja, Cotista ou não-Cotista.

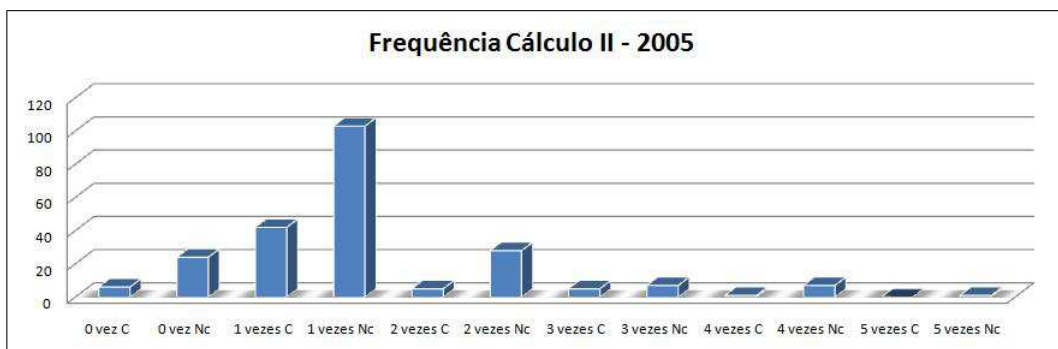


Figura 4.7: Frequência com que os alunos cursaram a disciplina de Cálculo Diferencial e Integral II

4.1.8 Variável Tipo de Entrada

No concurso 2004-2005, 14 candidatos optaram pelo curso de Engenharia Civil, 22 candidatos optaram pelo curso de Engenharia Elétrica e 23 candidatos optaram pelo curso de Engenharia Mecânica no grupo dos Cotistas. Para os não-Cotistas 51 candidatos optaram por Engenharia Civil, 60 por Engenharia Elétrica e 59 por Engenharia Mecânica, a distribuição dos candidatos para os cursos ficou da seguinte forma:

Tabela 4.3: Distribuição da Entrada

Tipo de entrada	Quantidade	Porcentagem(%)
Cota Afro	24	10.48
Cota Escola Pública	35	15.28
Concorrência Geral	79	34.5
Chamada Complementar	91	39.74
Total	229	100

FONTE: O autor(2010)

As próximas variáveis apresentam subitens, foram analisadas em grupos, e somente os resultados dos testes, serão mostrados a seguir.

4.1.9 Variável Estado civil

A variável Estado civil, esta sendo representada com as seguinte opções: solteiro(a), casado(a) e outros, entre as opções marcadas, ficou-se quase na totalidade o item “solteiro(a)”, ou seja, dos 229 candidatos, 227 são solteiros.

Aplicando o Teste de igualdade de proporções na variável Estado Civil com os itens solteiro(a), no grupo Cotista e não-Cotista, conclui-se que nas proporções observadas de Cotistas, 100% dos Cotistas e 98% dos não-Cotistas são solteiros, com estimativa de diferença de aproximadamente 1,1%. Considerando 95% para o índice de confiabilidade e p -valor = 0,155, pode-se afirmar que a diferença entre as duas amostras não é significativa.

4.1.10 Variável Moradia

Está apresentada na TABELA 4.5 a investigação da existência da igualdade de proporção entre Cotista e não-Cotista em todos os itens da seção 3.2.11.

Tabela 4.5: Condições de Moradia

Hipótese	Est. do teste z	valor p
$H_{01} : P_{csPqC} = P_{csPqNc}$	0,03	0,976
$H_{02} : P_{csPaC} = P_{csPaNc}$	-0,57	0,569
$H_{03} : P_{csEpC} = P_{csEpNc}$	-2,27	0,023
$H_{04} : P_{csEaC} = P_{csEaNc}$	1,78	0,075
$H_{05} : P_{csEstPenC} = P_{csEstPenNc}$	0,62	0,534
$H_{06} : P_{csParenAmigC} = P_{csParenAmigNc}$	-0,06	0,950
$H_{07} : P_{csAPC} = P_{csAPNc}$	-0,59	0,557

FONTE: O autor(2010)

É possível observar que não houve diferença estatística significativa nas hipóteses

$H_{01}, H_{02}, H_{04}, H_{05}, H_{06}$ e H_{07} pois os valores de p apresentados foram todos maiores que 0,05. Houve diferença significativa no item H_{03} apresentando p menor que 0,05.

4.1.11 Variável Grau de instrução do Pai

Fazendo o teste de igualdade de 2-proporções sobre os itens da seção 3.2.12 entre os grupos Cotista e não-Cotista, com as seguintes hipóteses:

H_0 : ser Cotista interfere no grau de instrução do pai : $p_1 = p_0$;

H_1 : ser não-Cotista não interfere no grau de instrução do pai: $p_1 \neq p_0$;

Nível de decisão: $\alpha = 0,05$, bilateral.

Tabela 4.7: Grau de Escolaridade do Pai

Hipótese	Est. do teste z	valor p
$H_{01} : P_{seC} = P_{seNc}$	1,15	0,249
$H_{02} : P_{fiC} = P_{fiNc}$	3,53	0
$H_{03} : P_{fcC} = P_{fcNc}$	2,15	0,032
$H_{04} : P_{emiC} = P_{emiNc}$	2,43	0,015
$H_{05} : P_{emcC} = P_{emcNc}$	1,16	0,245
$H_{06} : P_{siC} = P_{siNc}$	-7,35	0,708
$H_{07} : P_{scC} = P_{scNc}$	-0,59	0
$H_{08} : P_{nsC} = P_{nsNc}$	-0,04	0,978

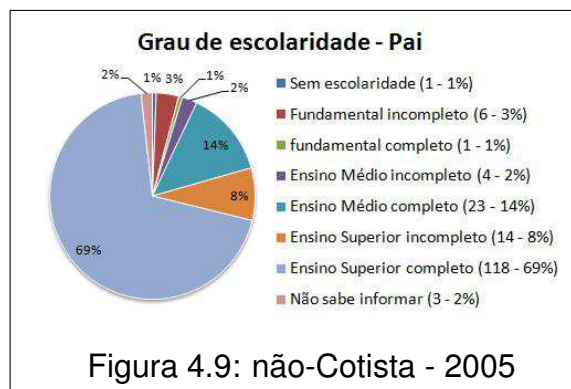
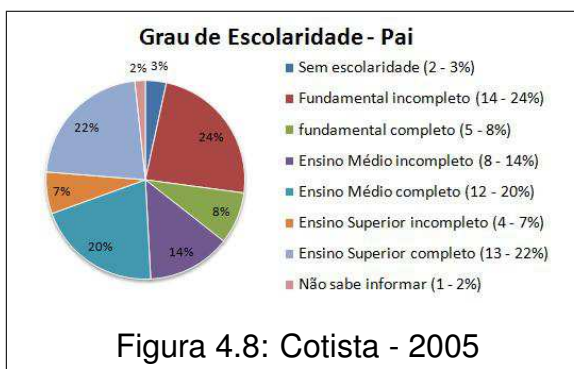
FONTE: O autor(2010)

Com base nos dados da TABELA 4.7 pode-se afirmar que existe diferença significativa entre as hipóteses H_{02}, H_{03}, H_{04} e H_{07} , baseado nos p valores 0¹; 0,032; 0,015 e 0 respectivamente, desta forma rejeita-se a hipótese de nulidade e aceita-se a alternativa, no sentido de que o pai com fundamental incompleto, fundamental completo, ensino médio incompleto e ensino superior completo existe diferença significativa en-

¹os p-valores com resultados muito próximo de zero, o próprio programa arredonda para zero.

tre os Cotistas e não-Cotistas

É possível visualizar a distribuição dos itens para cada grupo através dos gráficos de setores (FIGURA 4.8 e FIGURA 4.9)



O gráfico da FIGURA 4.8 indica que a escolaridade do pai tem maior incidência sobre dois itens, tendo-se 24% dos pais com fundamental incompleto e 22% com ensino superior completo. O gráfico da FIGURA 4.9 para os alunos não-Cotista é representado por 69% dos pais com ensino superior completo.

4.1.12 Variável Grau de instrução da mãe

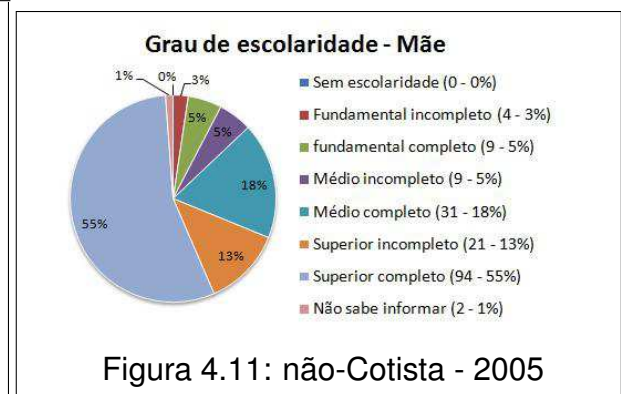
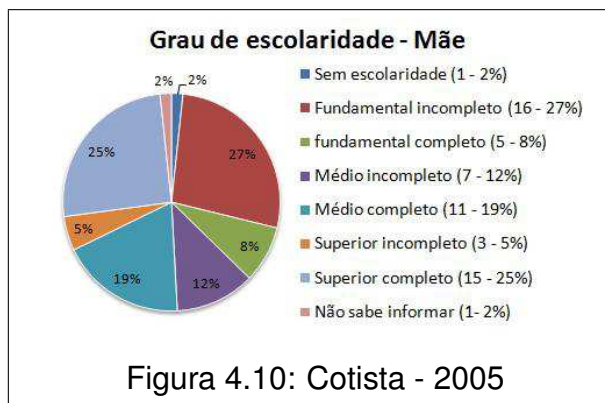
Os resultados foram dispostos graficamente na FIGURA 4.10 e FIGURA 4.11, sendo que os resultados do teste de 2-proporções com itens da seção 3.2.13 foram agrupados na TABELA 4.9 nas categorias Cotista e não-Cotista. Também é possível visualizar através do p valores as variáveis que apresentam diferenças significativas entre os grupos. Para as hipóteses H_{02} e H_{07} tendo "0" como p valor, dando a informação que existe diferença significativa entre os grupos na variável ensino fundamental incompleto e ensino superior completo.

Tabela 4.9: Grau de Escolaridade da Mãe

Hipótese	Est. do teste z	valor p
$H_{01} : P_{seC} = P_{seNc}$	1,01	0,313
$H_{02} : P_{fiC} = P_{fiNc}$	4,20	0
$H_{03} : P_{fcC} = P_{fcNc}$	0,79	0,428
$H_{04} : P_{emiC} = P_{emiNc}$	1,43	0,148
$H_{05} : P_{emcC} = P_{emcNc}$	0,07	0,944
$H_{06} : P_{siC} = P_{siNc}$	-1,91	0,0578
$H_{07} : P_{scC} = P_{scNc}$	-4,37	0
$H_{08} : P_{nsC} = P_{nsNc}$	0,28	0,782

FONTE: O autor(2010)

O gráfico da FIGURA 4.10 indica que a escolaridade da mãe Cotista tem maior incidência sobre dois itens, tendo-se 27% das mães com fundamental incompleto e 25% com ensino superior completo para não-Cotista e o gráfico da FIGURA 4.11 é representado por 55% das mães com ensino superior completo.



4.1.13 Variável Renda mensal familiar

Investigando-se a existência de igualdade na proporção entre Cotista e não-Cotista através da Renda Mensal Familiar (seção 3.2.14),

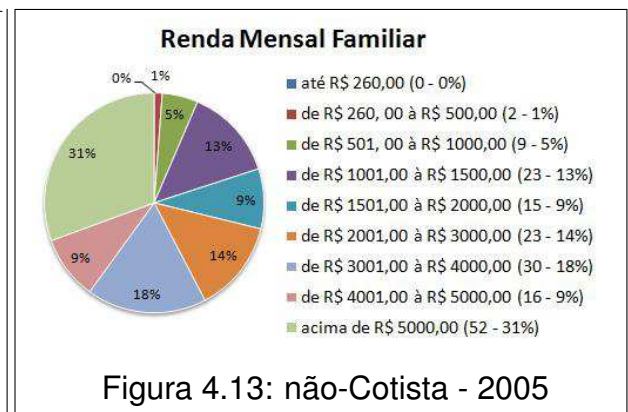
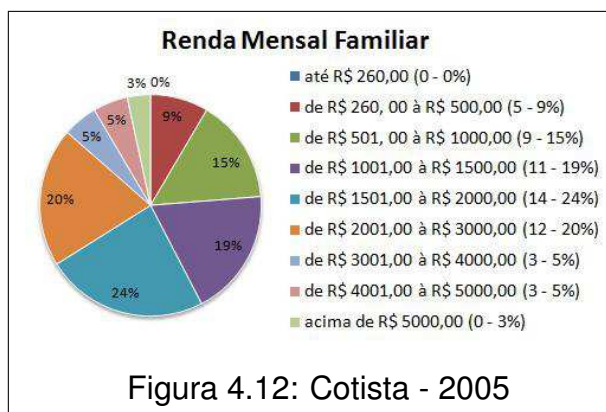
Tabela 4.11: Renda Mensal Familiar

Hipótese	Est. do teste z	valor p
$H_{01} : P_{ate260C} = P_{ate260Nc}$	*	*
$H_{02} : P_{261-500C} = P_{261-500Nc}$	1,96	0,050
$H_{03} : P_{501-1000C} = P_{501-1000Nc}$	2	0,046
$H_{04} : P_{1001-1500C} = P_{1001-1500Nc}$	0,90	0,370
$H_{05} : P_{1501-2000C} = P_{1501-2000Nc}$	2,5	0,012
$H_{06} : P_{2001-3000C} = P_{2001-3000Nc}$	1,16	0,245
$H_{07} : P_{3001-4000C} = P_{3001-4000Nc}$	-3,07	0,002
$H_{08} : P_{4001-5000C} = P_{4001-5000Nc}$	-1,19	0,234
$H_{09} : P_{ac5000C} = P_{ac5000Nc}$	-6,40	0

FONTE: O autor(2010)

os resultados são apresentados na TABELA 4.11.

Dos resultados acima pode-se perceber que as hipóteses $H_{01}, H_{02}, H_{04}, H_{06}$ e H_{08} não apresentam diferenças significativas na Renda Mensal Familiar, com a observação que para a renda familiar até R\$260,00 não houve nenhuma citação, quanto as hipóteses H_{03}, H_{05}, H_{07} e H_{09} com p valores 0,046; 0,012; 0,002 e 0 respectivamente, apresentam diferenças significativas. A seguir são apresentados os gráficos de setores ilustrando as FIGURAS 4.12 e 4.13.



O gráfico da FIGURA 4.12 exibe uma maior concentração no grupo dos Cotistas referente aos salários de R\$500,00 a R\$3000,00 concentrando 87% das opções. No gráfico da FIGURA 4.13 18% da concentração ocorre na faixa de R\$3000,00 a R\$4000,00, no grupo dos não-Cotista e 31% tem renda superior a R\$5000,00, enquanto que para os Cotistas somente 3% tem esta renda.

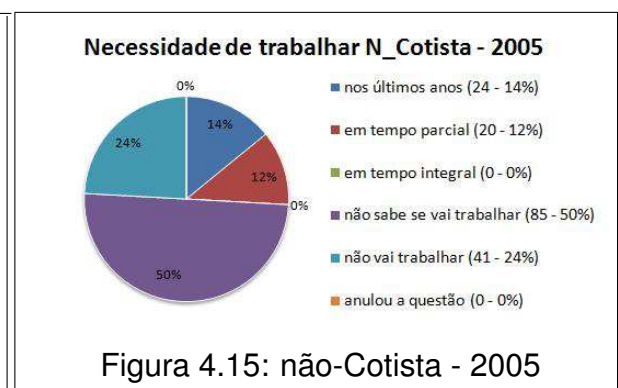
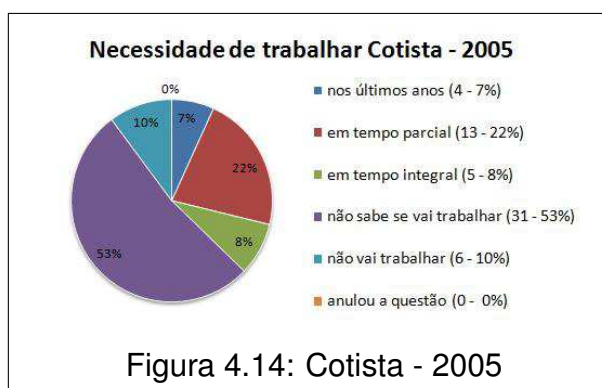
4.1.14 Variável Necessidade de trabalhar

Verificando a existência de igualdade entre as proporções Cotista e não-Cotista em todos os itens da seção 3.2.15, analisando a TABELA 4.13 pode-se afirmar que as hipóteses H_{03} e H_{05} apresentam diferença significativa com p valores 0,0019 e 0,006 entre Cotista e não-Cotista.

Tabela 4.13: Necessidade de Trabalho

Hipótese	Est. do teste z	valor p
$H_{01} : P_{TulC} = P_{TulNc}$	-1,74	0,082
$H_{02} : P_{T1o_pC} = P_{T1o_pNc}$	1,73	0,084
$H_{03} : P_{T1o_iC} = P_{T1o_iNc}$	2,34	0,0019
$H_{04} : P_{nsC} = P_{nsNc}$	0,34	0,736
$H_{05} : P_{mvC} = P_{1501-2000Nc}$	-2,72	0,006

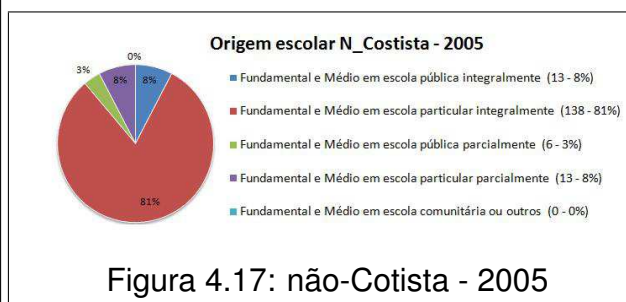
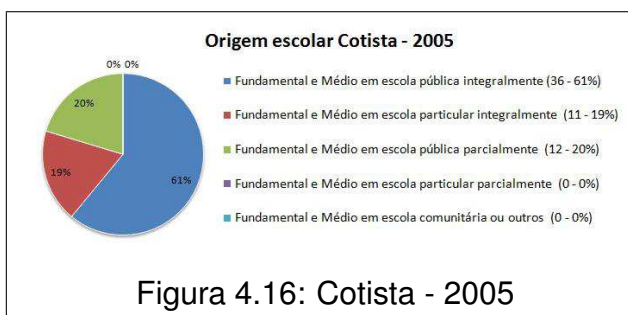
FONTE: O autor(2010)



Quase metade dos alunos que entraram na Universidade Federal do Paraná (UFPR) no ano de 2005 disseram que não sabiam se teriam que trabalhar durante o curso. No grupo dos Cotistas 22% afirmaram que para se manter na faculdade precisariam ter um emprego desde o primeiro ano em tempo parcial, enquanto que 12% dos não-Cotistas marcaram essa opção, no grupo dos não-Cotista 24% afirmam que não irão trabalhar, 10% dos Cotistas marcaram essa opção.

4.1.15 Variável Procedência educacional

Na sequência os itens da seção 3.2.16 estão tratados graficamente.



Os dados correspondentes ao teste de igualdade de proporções entre Cotista e não-Cotista estão reproduzidos no gráfico da TABELA 4.15. Os itens apresentados mostraram diferenças significativas em todos os itens da variável “Procedência Educacional” nos grupos Cotistas e não-Cotistas, sendo que nenhum aluno teve sua formação em escola comunitária, ou seja, é predominante a informação que se o aluno é Cotista tem sua formação em escola pública e aluno não-Cotista tem sua formação em escola privada.

Tabela 4.15: Procedência Educacional

Hipótese	Est. do teste z	valor p
$H_{01} : P_{escIpuC} = P_{escIpuNc}$	8	0
$H_{02} : P_{escIpaC} = P_{escIpaNc}$	-10,48	0
$H_{03} : P_{escPpuC} = P_{escPpuNc}$	3,10	0,002
$H_{04} : P_{escPpaC} = P_{escPpaNc}$	-3,75	0
$H_{05} : P_{escComC} = P_{escComNc}$	*	*

FONTE: O autor(2010)

4.1.16 Variável Escolha do curso

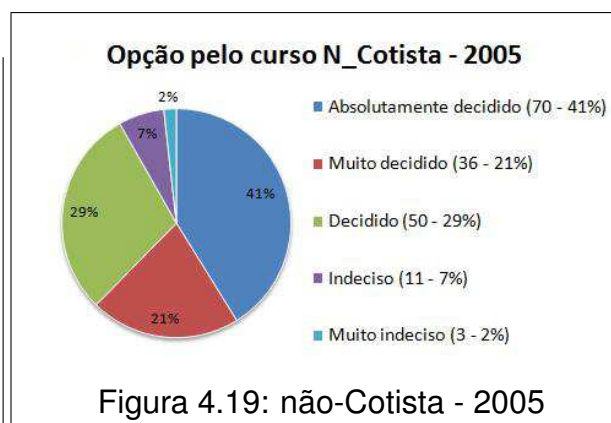
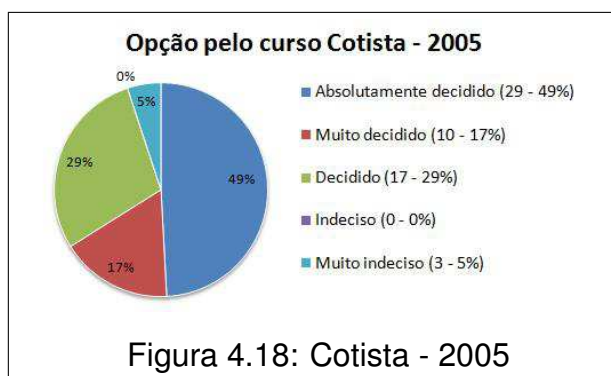
As respostas obtidas referente as alternativas da seção 3.2.17 foram organizadas na TABELA 4.17,

Tabela 4.17: Escolha do curso

Hipótese	Est. do teste z	valor p
$H_{01} : P_{adC} = P_{adNc}$	1,06	0,289
$H_{02} : P_{mdC} = P_{mdNc}$	-0,73	0,466
$H_{03} : P_{dC} = P_{dNc}$	-0,09	0,930
$H_{04} : P_{iC} = P_{iNc}$	-3,43	0,0001
$H_{05} : P_{miC} = P_{miNc}$	1,09	0,274

FONTE: O autor(2010)

analisadas de maneira a caracterizar qualitativamente esta pesquisa, de forma apresentar a convicção do aluno quanto a escolha do curso. Nesta tabela é possível visualizar o teste de proporções dos itens no grupo dos Cotistas e dos não-Cotistas. Apresentando diferença significativa somente na variável indeciso ($p = 0,0001$), o fato de aluno ser Cotista mostra uma maior convicção quanto a sua escolha.



Analisando graficamente as respostas do questionário no gráfico das FIGURAS 4.18 e 4.19, pode-se verificar que 49% dos universitários estão absolutamente decididos, 17% muito decididos, 29% estão decididos, 0% indecisos e 5% muito indecisos. Estas porcentagens representam os Cotistas. Para os não-Cotistas, 41% dos universitários estão absolutamente decididos, 21% muito decididos, 29% estão decididos, 7% indecisos e 2% muito indecisos. Tendo a mesma porcentagem o item “decidido” com 29%.

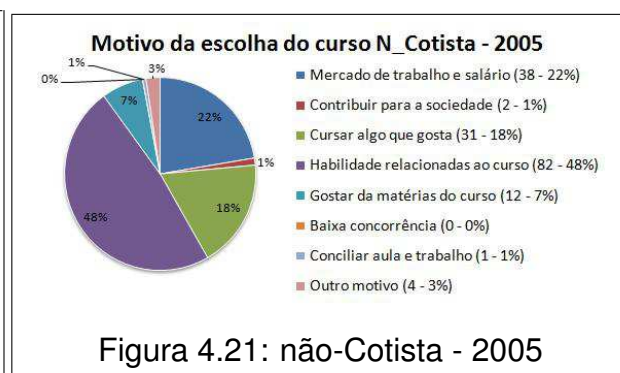
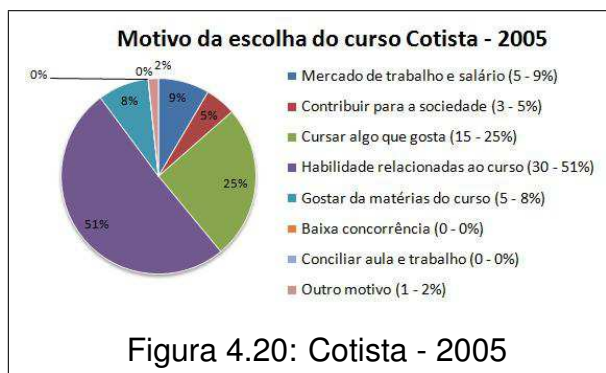
4.1.17 Variável Motivo da escolha do curso

Analisando o teste de 2-proporções da seção 3.2.18, sendo que os resultados estão na TABELA 4.19, constatando no conjunto de dados descritos pelo teste de 2-proporções, que somente a hipótese H_{01} ($p = 0,004$) apresenta diferença significativa, com a opção “Mercado de trabalho e salário” representando 9% para Cotista e 22% para não-Cotista, onde 0% dos candidatos marcaram a opção “baixa concorrência”, conforme é ilustrado nos gráficos das FIGURAS 4.20 e 4.21.

Tabela 4.19: Motivo da Escolha do Curso

Hipótese	Est. do teste z	valor p
$H_{01} : P_{mtC} = P_{mtNc}$	-2,87	0,004
$H_{02} : P_{csC} = P_{csNc}$	1,31	0,189
$H_{03} : P_{rpC} = P_{rpNc}$	1,12	0,261
$H_{04} : P_{gpC} = P_{gpNc}$	0,35	0,729
$H_{05} : P_{gmC} = P_{gmNc}$	0,34	0,731
$H_{06} : P_{bcC} = P_{bcNc}$	*	*
$H_{07} : P_{pcC} = P_{pcNc}$	-1	0,316
$H_{08} : P_{omC} = P_{omNc}$	-0,32	0,747

FONTE: O autor(2010)



Dentre estas categorias de motivos de escolha do cursos citados, observou-se que a categoria “Habilidade relacionada com o curso” apresentou aproximadamente metade das categorias escolhidas, 51% para Cotistas e 48% para não-Cotistas. Verificou-se também que no ano 2005 houve uma maior incidência das categorias de motivos voltados a fazer o curso que gosta e ao mercado de trabalho, havendo uma diferença de 7% e 13% entre os grupos. As outras categorias, que envolvem os motivos se mostraram menos atuantes nas decisões de escolha dos cursos.

4.1.18 Variável O que espera do curso

Aplicando o teste de 2-proporções na seção 3.2.19 e apresentando os resultados na TABELA 4.21,

Tabela 4.21: O que Espera do Curso

Hipótese	Est. do teste Z	valor p
$H_{01} : P_{acC} = P_{acNc}$	1,17	0,242
$H_{02} : P_{fpC} = P_{fpNc}$	-0,24	0,808
$H_{03} : P_{ipC} = P_{ipNc}$	0,34	0,731
$H_{04} : P_{fpC} = P_{fpNc}$	0,89	0,375
$H_{05} : P_{niC} = P_{niNc}$	-1,04	0,298
$H_{06} : P_{dsC} = P_{dsNc}$	-1,75	0,081

FONTE: O autor(2010)

os resultados mostram que, nos conjuntos descritos não existe nenhuma diferença significativa, porém nos gráficos das FIGURAS 4.22 e 4.23 pode-se visualizar que “Formação profissional, voltada para mercado de trabalho” foi o item mais citado pelos alunos, nos sete grupos, representando 70% e 71% respectivamente. Isto parece indicar a crescente procura dos cursos de Engenharia Civil, Elétrica e Mecânica da UFPR, a necessidade de respostas rápidas às exigências do mercado de trabalho cada vez mais concorrido e exigente.



A análise feita na seção 4.2, para os alunos serem considerados “Sucesso”, além de ser concluinte é necessário ter cursado apenas três vezes as disciplinas de Cálculo I e Cálculo II, no total. Assim, nessa classificação a reprovação é possível somente uma vez em uma delas, pois essas disciplinas são ofertadas no 1º ano nos dois semestres.

4.2 ANÁLISE DA TÉCNICA DE REGRESSÃO LOGÍSTICA

Para a implementação da técnica de Regressão Logística e Redes Neurais no presente trabalho foram analisados e avaliados 1213 observações, agrupando 1013 observações para aplicação do teste e 200 observações para validação do mesmo, de maneira aleatória, com os alunos dos cursos referidos no período de 2005 à 2008, tendo como variável dependente “Conclusão” e “Sucesso”, as análises ocorreram com base no software Statgraphics.

Para os alunos serem considerados “concluintes” nas disciplinas de Cálculo I e II, após o ingresso na UFPR nos cursos citados, com o ano base o ano de entrada contando a partir de 2005, é preciso serem aprovados nas disciplinas de Cálculo I e II.

A descrição e interpretação dos mesmos são apresentados neste trabalho em tabelas, gráficos e análises estatísticas.

Através das variáveis obtidas no capítulo 3, foi realizada a análise baseando-se somente nos p -valores ($p < 0,05$), analisando a relação de cada variável com a variável dependente “Conclusão” e posteriormente com a variável dependente “Sucesso” através do software Statgraphics, aplicando a técnica de regressão logística.

Cada conjunto de dados ficou constituído com 1013 observações, onde foram testadas as 82 variáveis dispostas na TABELA 2.1 aplicando a regressão logística sobre esse conjunto de variáveis, verificou-se a necessidade de excluir variáveis para atender a hipótese de $p < 0,05$, desta forma sendo analisado somente as variáveis

nota_2fase, Cálculo I, Cálculo II, Sexo, Pai_ Ensino fundamental incompleto, Pai_ Ensino fundamental completo, Pai_ Ensino médio completo, Mãe_ Superior completo, Escolha_ Indeciso e Espera_ Melhorar o meu nível de instrução, encontrado a cada inserção de variável. Um p -valor pequeno (menor de 0,05 opera no nível de significância de 5%) indica que o modelo tem reduzido significativamente o desvio e, portanto, é útil para prever a probabilidade dos resultados estudados.

4.2.1 Variável dependente “Conclusão”

Para esta variável dependente considerou-se somente se o aluno que concluiu as disciplinas de Cálculo I e II, ou não no período de 2005 à 2008. Com o modelo de regressão logística para descrever a relação entre conclusão e 10 variáveis independentes é:

Conclusão = $\exp(y) / (1 + \exp(y))$, onde,

$$y = -0,562809 + 0,00522062 * \text{nota_2fase} - 0,82661 * \text{calculo I} + 0,518919 * \text{calculo II} - 0,496423 * \text{SEXO} - 0,855729 * \text{Pai_ Ensino fundamental incompleto} - 0,966004 * \text{Pai_ Ensino fundamental completo} - 0,381893 * \text{Pai_ Ensino médio completo} - 0,270729 * \text{Mãe_ Superior completo} + 0,397124 * \text{Escolha_ Indeciso} + 0,631018 * \text{Espera_ Melhorar o meu nível de instrução}$$

Com a TABELA 4.23 de desempenho pode-se mostrar um resumo da capacidade de previsão do modelo ajustado. Primeiro, o modelo é usado para prever a resposta usando as informações em cada linha do arquivo de dados. Se o valor previsto é maior que o de corte, a resposta é prevista para ser verdade, se o valor é menor ou igual ao corte a resposta esta prevista para ser falsa. A TABELA 4.23 mostra a porcentagem de dados observados prevista corretamente a valores de corte igual a 0,5 com 78,72% de todas as respostas “1”(concluiu) prevista corretamente, enquanto que 57,59% de todas as respostas “0”(não-concluiu), foram previstas corretamente para um total de 69,37%. Usando o valor de corte que maximiza a porcentagem total correta pode-se

fornecer um bom valor a ser usado para prever indivíduos adicionais.

Tabela 4.23: Desempenho de Predição

Linha de Corte	1	0	Total
0,0	100	0,0	55,73
0,05	99,82	1,34	56,23
0,1	99,82	3,57	57,21
0,15	99,47	6,03	58,10
0,2	98,76	10,27	59,58
0,25	97,70	15,40	61,26
0,3	95,92	22,77	63,54
0,35	93,44	28,79	64,82
0,4	88,12	37,95	65,91
0,45	84,40	46,43	67,59
0,5	78,72	57,59	69,37
0,55	72,34	64,96	69,07
0,6	62,47	73,21	67,39
0,65	51,06	79,02	63,44
0,7	38,30	85,71	59,29
0,75	24,65	91,29	54,15
0,8	14,89	95,98	50,79
0,85	6,03	98,21	46,84
0,9	1,24	98,88	44,47
0,95	0,0	100	44,27
1,0	0,0	100	44,27

FONTE: O autor(2010)

A seguir a FIGURA 4.24 mostra graficamente a linha de Corte. Através do modelo foi gerada uma linha de corte, sugerida como ideal (0,5), baseado nos dados da TABELA 4.23 e FIGURA 4.24.

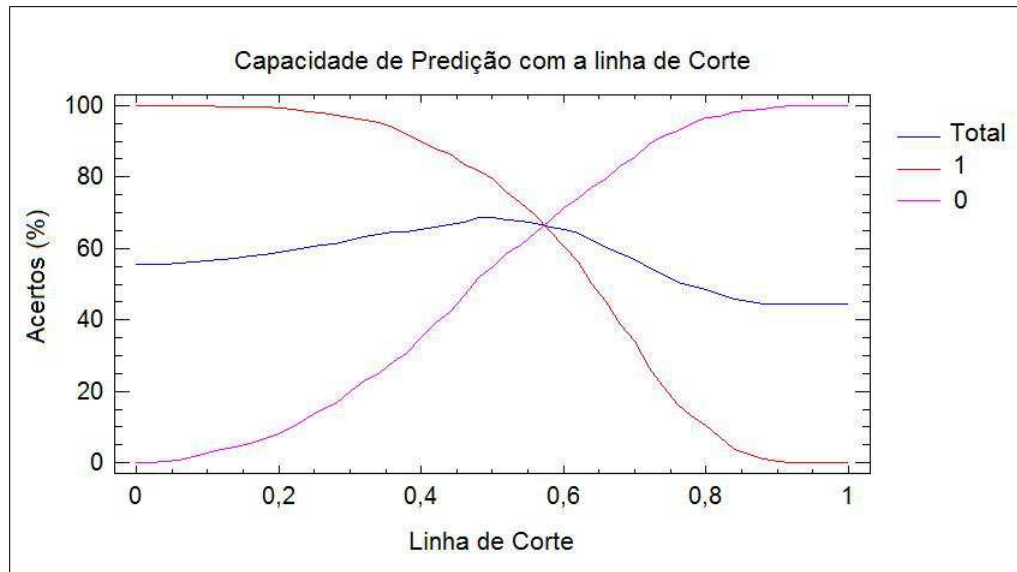


Figura 4.24: Linha de Corte

Essa linha indica que se os valores ao passarem pelo modelo ajustado ficarem abaixo da linha de corte, esses serão classificados como saída “0” e se assumirem valores superiores a linha de corte serão classificados como saída “1”.

Para esta técnica de classificação, o percentual de acertos ficou com cerca de 69% para o conjunto de treinamento e para o conjunto de validação (200 observações) o percentual de acertos foi de 60%.

4.2.2 Variável dependente “Sucesso”

Esta variável considerou além do aluno ter concluído, ter feito as disciplinas de Cálculo I e II dentro de 3 períodos. A TABELA 4.25 mostra as variáveis que entraram no modelo, gerada pelo software STATGRAPHICS de regressão logística bem como os seus coeficientes e os resultados do teste de significância das variáveis independentes. A segunda coluna (coeficientes) apresenta o modelo na sequência.

Tabela 4.25: Modelo de Regressão estimada

Parâmetro	Estimativa coeficientes	Padrão Erro
Constante	0,0130754	0,923285
nota_2fase	-0,00364209	0,000641343
Cálculo	1,25034	0,127393
Pai_Ensino fundamental incompleto	1,09398	0,302589
Pai_Ensino fundamental completo	1,17553	0,384114
Pai_Ensino médio completo	0,596369	0,215742
Pai_superior completo	0,464264	0,1877
Trabalhar desde do 1ºano_em tempo integral	-1,33893	0,461623
motivo_Mercado de trabalho e salário	73,16	62,26
Espera_compreender melhor o mundo	-0,419122	0,220135
Espera_Melhorar o meu nível de instrução	-0,623688	0,318254

FONTE: O autor(2010)

Sucesso = $\exp(y)/(1+\exp(y))$, onde:

$y = 0,0130754 - 0,00364209 * \text{nota_2fase} + 1,25034 * \text{Cálculo I} + 1,09398 * \text{Pai_Ensino fundamental incompleto} + 1,17553 * \text{Pai_Ensino fundamental completo} + 0,596369 * \text{Pai_ensino médio completo} + 0,464264 * \text{Pai_superior completo} - 0,29018 * \text{motivo_Mercado de trabalho e salário} - 1,33893 * \text{Trabalhar desde do 1º ano_em tempo integral} - 0,419122 * \text{Espera_compreender melhor o mundo} - 0,623688 * \text{Espera_Melhorar o meu nível de instrução}$

Através do modelo foi gerada uma linha de corte ilustrada na TABELA 4.27 e FIGURA 4.25.

Tabela 4.27: Desempenho de Predição “Sucesso”

Linha de Corte	1	0	Total
0,0	100	0,0	48,23
0,05	100	0,0	48,23
0,1	100	0,0	48,23
0,15	99,49	0,80	48,39
0,2	98,12	3,66	49,22
0,25	94,70	13,22	52,51
0,3	88,21	28,50	57,30
0,35	82,74	46,50	63,97
0,4	73,16	62,26	67,52
0,45	64,96	73,25	69,25
0,5	57,09	79,94	68,92
0,55	51,45	85,03	68,84
0,6	44,96	87,74	67,11
0,65	37,61	90,76	65,13
0,7	31,62	92,99	63,40
0,75	24,10	95,54	61,09
0,8	18,29	97,45	59,27
0,85	14,02	98,57	57,79
0,9	8,89	99,68	55,89
0,95	3,93	100	53,67
1,0	0,0	100	51,77

FONTE: O autor(2010)

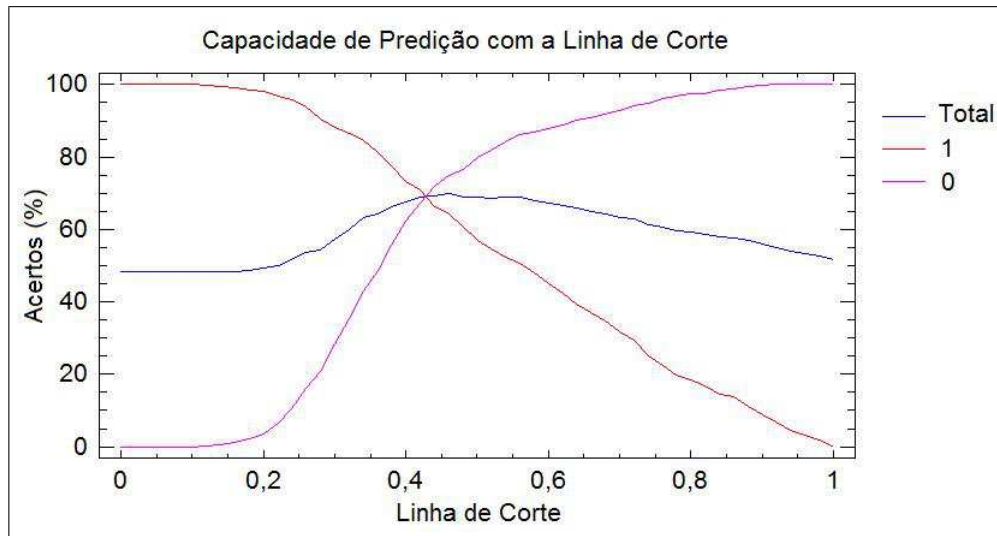


Figura 4.25: Linha de Corte

A regressão logística com inserção dos dados e uso do software Statgraphics prevê 69,25% de previsão de acertos para este conjunto de treinamento (1013 observações) e 62% para o conjunto de validação (200 observações).

4.3 ANÁLISE DA TÉCNICA DE REDE NEURAL

Os dados utilizados foram os mesmos da Regressão Logística Múltipla, a fim de obter parâmetros de comparação entre as duas técnicas estatísticas, obtendo um modelo ajustado, os resultados obtidos são apresentados a seguir:

4.3.1 Resultados das RNAs

O software STATGRAPHICS gerou automaticamente uma RNA com dez neurônios na camada de entrada, como mostra a FIGURA 4.27, apresentando cerca de 71% de acertos para o conjunto de treinamento, fazendo uma melhor previsão quando o aluno concluirá as disciplinas (87,06%) do que a possível reprovação com (12,94%) de acerto.

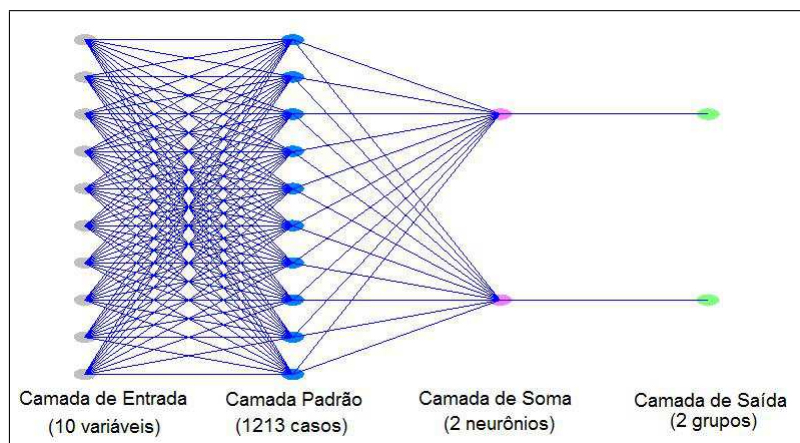


Figura 4.26: Rede Neural gerada pelo software STATGRAPHICS

O diagrama da Rede Neural ilustra a configuração básica da rede, sendo esta composta por dez neurônios na camada de entrada, correspondentes as variáveis (listados adiante), e dois neurônios na camada de saída.

No treinamento e teste da rede neural as variáveis nota_2fase, Cálculo I, Cálculo II, SEXO, Pai_Ensino fundamental incompleto, Pai_Ensino fundamental completo, Pai_ensino médio completo, Mãe_superior completo, Escolha_Indeciso e Espera_Melhorar o meu nível de instrução, foram inseridas como dados de entrada, enquanto Concluiu e não_Concluiu será obtido como saída. A camada de saída dos neurônios produz a predição da rede como resultado.

Foi feito o mesmo teste com as variáveis nota_2fase, Cálculo I, Pai_Ensino fundamental incompleto, Pai_Ensino fundamental completo, Pai_ensino médio completo, Pai_superior Completo, Trabalhar desde do 1ºano_em tempo integral, motivo_Mercado de trabalho e salário, Espera_compreender melhor o mundo e Espera_Melhorar o meu nível de instrução, inseridas como entrada e como saída Sucesso e não_Sucesso. Apresentando o seguinte diagrama com acerto de 66,94% para classificação correta de um aluno, com 53% de chance de classificar corretamente esse aluno com concluinte e 47% de chance de classificar corretamente como aluno não concluinte.

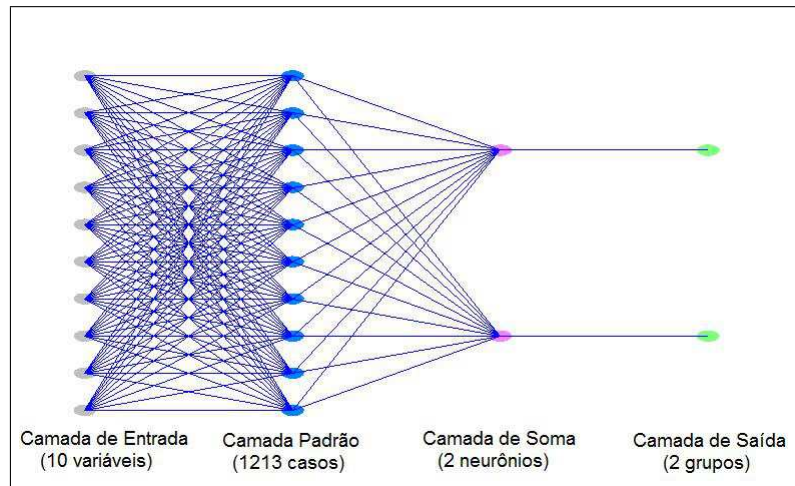


Figura 4.27: Rede Neural gerada pelo software STATGRAPHICS

5 CONCLUSÃO E SUGESTÕES DE TRABALHOS FUTUROS

O objetivo deste estudo foi analisar o perfil do aluno dos cursos de Engenharia Civil, Engenharia Elétrica e Engenharia Mecânica da UFPR baseado nas disciplinas de Cálculo I e II, por meio da estatística descritiva, indutiva e comparação das técnicas de Regressão logística e Rede neural, referente a alunos regularmente matriculados nos anos de 2005 a 2008, sendo o primeiro vestibular (2004 – 2005) a contemplar as políticas de cotas.

Com a estatística descritiva e indutiva constatou-se diferença significativa entre o grupo dos Cotistas e não-Cotistas nas variáveis nota_1fase, nota_2fase e frequência que cursou a disciplina Cálculo II contrapondo a igualdade das variáveis Conclusão, Gênero, Idade, nota_1fase, Frequência que cursou a disciplina Cálculo I e Frequência que cursou a disciplina Cálculo II. Também foi possível confirmar o predomínio do gênero masculino nos cursos em foco, sendo 90% masculino e 10% feminino com idade média de 18 anos em ambos os gêneros.

Como a maioria dos alunos são jovens, no quesito estado civil, 100% dos alunos Cotistas e 98% do alunos não-Cotistas são solteiros sendo que 84% moram com os pais. A renda familiar dos Cotistas esta concentrada nos salários de R\$ 500,00 a R\$ 3000,00 correspondente a 1,6 – 10 salários mínimos da época (2005). Os não-Cotistas estão na faixa de R\$ 3000,00 a R\$ 4000,00 correspondendo de 10 – 13 salários mínimos. De uma forma geral 31% dos não-Cotistas têm renda superior a 16 salários mínimos, enquanto que para os Cotistas somente 3%. Quase metade dos alunos

afirmaram que não sabiam se iriam trabalhar durante o curso, talvez devido ao fato de que o curso escolhido é em período integral.

O questionário Sócioeducacional também questionou aos alunos, sobre procedência educacional mostrando que 39% dos Cotistas cursou o ensino fundamental e médio em escola particular e 81% dos alunos não-Cotista cursou em escola particular. Quanto às expectativas de mercado de trabalho, a maior parte deles respondeu que o primeiro objetivo é “Formação profissional, voltada para o mercado de trabalho”, poucos alunos consideraram mais importante apenas obter o diploma. Na questão “Qual o motivo da escolha do curso?”, quase metade dos alunos optou pela resposta “Habilidade relacionada com o curso” e com a mesma porcentagem afirmam estar absolutamente decidido em relação a escolha do curso.

No aspecto familiar, o estudo mostra que apenas 22% dos pais de alunos Cotistas tem ensino superior completo, enquanto que 69% dos pais de alunos não-Cotista tem ensino superior completo, em relação as mães, 25% e 55% respectivamente, possui ensino superior completo.

Com a análise multivariada verifica-se por meio do modelo de regressão logística que as variáveis significativamente importantes para explicar o “Perfil do aluno” usando como a variável dependente “Conclusão” são: nota_2fase, Cálculo I, Cálculo II, sexo, Pai_Ensino fundamental incompleto, Pai_Ensino fundamental completo, Pai_Ensino médio completo, Mãe_superior completo, Escolha_Indeciso, Espera_Melhorar o meu nível de instrução, as demais variáveis não foram significativas. Para variável sucesso tem-se como variáveis independentes: nota_2fase, Cálculo I, Pai_Ensino fundamental incompleto, Pai_Ensino fundamental completo, Pai_Ensino médio completo, Pai_Superior completo, Trabalhar desde do 1ºano em tempo integral, motivo_Mercado de trabalho e salário, Espera_compreender melhor o mundo, Espera_Melhorar o meu nível de instrução. Existindo uma diferença de apenas 0,74% entre as duas variáveis dependentes.

De um modo geral, pode-se dizer que as motivações iniciais que possibilitam uma possível desistência do curso são as variáveis citadas no parágrafo anterior, concentrando de uma forma acentuada a opção grau de instrução do pai e outras como quanto a escolha e motivação, são as primeiras que podem influenciar para a desistência dos alunos no curso.

Comparando as técnicas de regressão logística e rede neural para previsão de acertos pode-se afirmar que existe uma pequena diferença relacionada com as duas variáveis dependentes “Conclusão” e “Sucesso”, na técnica de regressão logística e redes neurais com a diferença de 2% e 3% respectivamente para classificação correta de um novo aluno.

Lembrando que os dados colhidos foram analisados de maneira fria, ou seja, apenas escolhida pelo p -valor a cada inserção no *software* Statgraphics, pode-se dizer que outras razões podem existir e essas motivações para a possível desistência do curso possam ser caracterizadas como intrínsecas, no sentido de que o aluno chega a envolver-se nas atividades acadêmicas pelas características da própria atividade, como um fim em si mesmo, independentemente de fatores externos. (BZUNECK, 2001)

Como sugestão para novos trabalhos , podem ser citadas:

- Inclusão de novos cursos;
- Correlação das notas de ingresso e o desempenho nas disciplinas de Cálculo I e II;
- Correlação pré cálculo e cálculo I e II;
- Inclusão de mais variáveis;
- Desagregar em afrodescendente, indígena e escola pública para análise .

REFERÊNCIAS

ALLERUP, P. Multivariate analyses of student response profiles-across countries and gender. **International Journal of Educational Research**, v. 39, p. 5551–563, 2003.

ATTEWELL, P.; LAVIN, D. E. **Passing The torch: does higher education for the disadvantaged pay off across the generations**. New York: Russell Sage Foundation Publications, 2007.

BOWEN, W.; BOK, D. **The Shape of the River**. Estados Unidos: Princeton University Press, 1998.

BZUNECK, J. **Motivação do aluno**. Petrópolis: Vozes, 2001.

DODDS, A. E. *et al.* Comparing the academic performace of graduate-and undergraduate-entry medical students. **Medical Education**, v. 44, p. 197–204, 2010.

HOPFIELD, J. J. Neural networks and physical systems with emergent collective computational abilities. **Proceedings of national academy of sciences**, v. 79, p. 2554–2558, 1982.

HUARSAYA, E. A. T. **Aprendizado Bath Bayesiano para Redes Neurais**. Dissertação (Mestrado) — Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 1999.

JENSEN, F. V. **An Introduction to Baysean Networks**. London: UCL Press, 1996.

JOHNSON, R. A.; WICHERN, D. W. **Applied Multivariate Statistical Analysis**. New Jersey: Thomson, 1988.

LIMA, J. P. H. **Redes neurais artificiais aplicadas à otimização de processos de deposição de filmes finos poliméricos**. Dissertação (Mestrado) — Universidade de São Paulo, São Paulo, 2006.

MARQUES, J. M. **Testes Estatísticos para Cursos das áreas Biológica e da Saúde com uso do computador**. Curitiba - PR: UFPR, 2004.

MEDEIROS, L. F. **Redes Neurais em Delphi**. [S.l.]: Visual Books, 2006.

NOGUEIRA, O. **Preconceito de Marca: as relações raciais em Itapetininga**. São Paulo: [s.n.], 1998.

Palma Neto, L. G.; NICOLETTI, M. C. **Introdução às Redes Neurais Construtivas**. São Carlos, SP: Editora da Universidade Federal de São Carlos, 2005.

PIOVESAN, F. **Ações Afirmativas Sob a Perspectiva dos Direitos Humanos**. Brasília: Ministério da Educação, Secretaria de Educação Continuada, Alfabetização e Diversidade, 2005.

RIBEIRO, F. P.; Menezes Filho, N. A. **Os Determinantes da Melhoria do Desempenho Escolar**. Rio de Janeiro: Educação Básica no Brasil., 2009.

SHANKLAND, R. *et al.* Student adjustment to higher education: the role of alternative educational pathways in coping with the demands of student life. **High Educ**, v. 59, p. 353–204, 2010.

SMITH, S. Y. S.; CURBOW, B. A.; STILLMAN, F. A. Differing psychosocial risk profiles of college freshmen waterpipe, cigar, and cigarette smokers. **Addictive Behaviors**, v. 33, p. 1619–1624, 2008.

STEINER, M. T. A. **Redes Neurais: Notas de aula apresentadas na disciplina Aplicações de Pesquisa Operacional**. Universidade Federal do Paraná: [s.n.], 2006.

TEIXEIRA, L. L. **O uso de técnicas de estatística multivariada no prognóstico de desistência de alunos em IES privadas: Um estudo de caso na cidade de Foz do Iguaçu - PR**. Dissertação (Mestrado) — UFPR, 2006.

TRIOLA, M. F. **Introdução à Estatística**. [S.l.]: LTC, 1999.

WEDDERBURN, C. M. **Marco Histórico das Políticas de Ação Afirmativa**. Brasília: Ministério da Educação, Secretaria de Educação Continuada, Alfabetização e Diversidade, 2005.

YOGEV, S.; YOGEV, A. Teacher educators as researchers: A profile of research in Israeli teacher colleges versus university departments of education. **Teaching and Teacher Education**, v. 22, p. 32–41, 2006.

APÊNDICE A – QUESTÕES DO QUESTIONÁRIO SÓCIOEDUCACIONAL

- Qual o seu sexo?
- Quantos anos você completará em (ano de entrada)?
- Qual o seu estado civil?
- Qual o Estado em que nasceu?
- Qual o local de sua residência? (Trata-se de residência permanente e não temporária, para fins de estudo)
- Qual a sua situação quanto à moradia?
- Qual o nível de instrução do seu pai?
- Qual o nível de instrução da sua mãe?
- Qual a principal ocupação do seu pai?
- Qual a principal ocupação da sua mãe?
- Você trabalha atualmente em atividade remunerada?
- Qual é a sua participação na vida econômica de seu grupo familiar?
- A renda total mensal de sua família se situa na faixa:
- Quantas pessoas contribuem para a obtenção da renda familiar?
- Quantas pessoas são sustentadas com a renda familiar?

- Com que idade você começou a exercer atividade remunerada?
- Durante o curso, você terá obrigatoriamente que trabalhar?
- Como fez seus estudos do Ensino Fundamental e médio?
- Em que ano você concluiu (ou concluirá) o curso de Ensino Médio?
- Como fez seus estudos de Ensino Médio?
- Em que turno você fez o curso de Ensino Médio?
- Por quanto tempo você fez cursinho?
- Por que você fez cursinho?
- Você já fez vestibular em outros anos?
- Considerando sua capacidade de leitura em língua Estrangeira, em que situação você melhor se enquadra?
- Você já iniciou algum curso superior?
- Qual o principal motivo que o levou a inscrever-se no Processo Seletivo da UFPR?
- Quanto a sua escolha por este curso, você se considera:
- No caso de ter respondido indeciso ou muito indeciso, o que contribuiu para sua indecisão?
- Que recursos você utilizou para fazer sua escolha?
- Qual o motivo que o levou a escolher o curso para o qual está se candidatando?
- Quem ou o que mais o influenciou na escolha do curso?
- Quando você se decidiu pelo curso a que está se candidatando?

- Quanto à sua escolha pelo curso, você se considera: O que você espera, EM PRIMEIRO LUGAR, de um curso universitário?
- Como você se informou sobre o Processo Seletivo 2005 da UFPR?
- A sua cor ou raça é?
- Qual o curso escolhido?
- Dentre os problemas da sociedade atual, qual você considera mais importante?

Esses questionamentos podem apresentar algumas alterações com os passar dos anos esse questionário tem como ano base 2005.

APÊNDICE B – VARIÁVEIS DESMEMBRADAS

Tabela 2.1: Variáveis desmembradas

Variável Original	Variáveis Geradas
gênero	masculino feminino
estado civil	solteiro(a) casado(a) outros
habitação	mora casa dos pais (quitada ou financiada) mora casa dos pais (alugada) mora casa (própria, quitada ou financiada) mora casa alugada (paga pelo aluno) mora casa (estudante pensão ou pensionato) mora em casa (parentes ou amigos) mora em casa alugada (paga por seus pais)
grau de instrução do pai	sem escolaridade ensino fundamental incompleto ensino fundamental completo ensino médio completo ensino médio incompleto superior incompleto superior completo não sabe informar
grau de instrução da mãe	sem escolaridade ensino fundamental incompleto ensino fundamental completo ensino médio incompleto ensino médio completo superior incompleto superior completo não sabe informar
renda familiar	Renda até R\$ 830,00 R\$ 831,00 à R\$1 660,00 R\$ 1661,00 à R\$ 2490,00 R\$ 2491,00 à R\$ 3320,00 R\$ 3321,00 à R\$ 4980,00 R\$ 4981,00 à R\$ 6640,00

	R\$6641,00 à R\$ 8300,00 acima de R\$ 8301,00
trabalhar durante o curso	nos últimos anos desde do 1º ano em tempo parcial desde do 1º ano em tempo integral não sabe se vai trabalhar não vai trabalha anulou a questão
procedência educacional	Integralmente em Escola Pública (fund. e médio) integralmente em Escola Particular (fund. e médio) maior parte em Escola Pública (fund. e médio) maior parte em Escola particular (fund. e médio) escolas comunitárias ou outros(fund. e médio)
escolha do curso	absolutamente decidido muito decidido decido indeciso muito indeciso
motivo da escolha do curso	mercado de trabalho e salário possibilidade de contribuir para a sociedade possibilidade de cursar algo que custa por ter habilidades relacionadas ao curso gostar das matérias do curso baixa concorrência permite conciliar aula e trabalho outro motivo
o que espera do curso	aquisição de cultura geral formação profissional voltada para o mercado de trabalho formação teórica voltada para pesquisa formação acadêmica para melhorar a prática compreender melhor o mundo melhorar o nível de instrução melhorar o nível de instrução melhorar o nível social/financeiro diploma do nível superior outros
