



UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE MATEMÁTICA
DEPARTAMENTO DE ESTATÍSTICA



**Modelo preditivo de efetivação de matrículas com utilização da técnica de
Regressão Logística**

Autor: Leonardo de Miranda Pinheiro

Orientadora: Professora Dr.^a. Lisiane Priscila Roldão Selau

Porto Alegre, dezembro de 2014.

Universidade Federal do Rio Grande do Sul
Instituto de Matemática
Departamento de Estatística

**Modelo preditivo de efetivação de matrículas com utilização da técnica de
Regressão Logística**

Autor: Leonardo de Miranda Pinheiro

Monografia apresentada para obtenção
do grau de Bacharel em Estatística.

Banca Examinadora:
Professora Dra. Lisiane Priscila Roldão Selau
Professora Dra. Suzi Alves Camey
Professora Dra. Ingrid Finger

Porto Alegre, dezembro de 2014.

AGRADECIMENTOS

Primeiramente gostaria de agradecer a minha orientadora Lisiane, pela paciência e dedicação. Ao longo do trabalho sempre se disponibilizou a me ajudar com empenho e sabedoria, provando mais uma vez a excelente profissional que é.

Também gostaria de agradecer aos meus amigos próximos, que estiveram ao meu lado durante esta jornada, pela força e parceria. Em especial, Mayara e Michele, que me acompanharam desde o início do curso, e Alessandra Analu, que sempre me ajudou nos momentos mais complicados da faculdade.

Aproveito este momento para agradecer a instituição UFRGS como um todo por proporcionar um ensino de excelente qualidade. Assim como não poderia deixar de agradecer o CPD, pelo excelente trabalho em fornecer o banco de dados para este estudo.

E por fim, gostaria de agradecer a minha família, pelo carinho, confiança e apoio que sempre me foi dado ao longo desta etapa da minha vida. A todos, eu serei eternamente grato.

RESUMO

As universidades federais estão em um processo contínuo de desenvolvimento e busca de técnicas, a fim de otimizar os recursos utilizados e disponibilizados, garantindo, dessa forma, a prosperidade acadêmica e financeira. Os excessivos gastos de recursos por tais instituições, no intuito da oferta de vagas (matrículas) para seus alunos, todavia, são pouco aproveitados, gerando por consequência o seu uso desnecessário, bem como a criação de custos. Com isso, a utilização de modelos de gerenciamento de matrículas tem se tornado uma ferramenta importante para as instituições de educação superior (IES). Aplicação deste estudo piloto, utilizando a técnica de regressão logística tem o intuito de melhorar o sistema de matrículas da Universidade Federal do Rio Grande do Sul (UFRGS). Para construção do modelo preditivo (efetivação) de matrículas da UFRGS, se escolheu a disciplina de MAT01354 - Cálculo e Geometria Analítica II como piloto. Essa disciplina foi escolhida devido a sua dimensão, por ser ofertada a diversos cursos e também por possuir pré-requisitos (disciplina de Cálculo I) e diversos perfis de alunos. Foi utilizado uma base de dados fornecida pela instituição, na qual se aplicou a modelagem estatística, construindo um modelo preditivo através da técnica de Regressão Logística. O modelo construído obteve índices satisfatórios de predição para o teste KS, curva ROC, e % de acerto total, sendo assertivo para a classificação da efetivação da matrícula. Dessa forma o uso do modelo preditivo para a disciplina de Cálculo e Geometria Analítica II se mostrou eficiente na predição de matrículas, o que contribui para uma melhor gestão dos recursos da universidade.

PALAVRAS-CHAVE: Gestão de matrículas; Planejamento estratégico; Modelo logístico.

ABSTRACT

The federal universities are in a continuous process of development and search of techniques, in order to optimize the available resources- thereby ensuring, the academic and financial prosperity. The overspending resources by these institutions, with the purpose of offering vacancies (enrollment) for its students, however, are barely taken, making as a consequence, its unnecessary use, as excessive costs. Thus, the use of enrollment management models, have become a crucial tool for higher education institutions (HEIs). Application of this test study using logistic regression model in order to improve the enrollment system of the Federal University of Rio Grande do Sul. For the building of the predictive modelling (effectuation) of enrollments of UFRGS, the discipline MAT01354- Calculus and analytic geometry II- was chosen as a test; this choice was made due its highly embracing scope, as its offered in many formations, has high requirements (such as 'Calculus I') and gathers a variety student profiles. It was used an database provided by the institution, in which it was applied the statistical modeling, making a predictive model through the logistical regression technique. The model has obtained satisfactory levels of prediction for KS test, ROC curve and % accuracy, being assertive for the classification effectuation of the enrollment. DISCUSSION: The use of the model for the discipline Calculus and Analytic Geometry II, has shown to be effective at the prediction of enrollments , which enable a better management of resources at UFRGS. Therefore the use of the model for the discipline Calculus and Analytic Geometry II, has shown to be effective at the prediction of enrollments , which contributes to a better management of resources at UFRGS.

Key Word's : Enrollment management; Strategic planning; Logistic model.

ÍNDICE TABELA

Tabela 1 - Nível de discriminação do teste KS.....	17
Tabela 2 - Nível de discriminação da curva ROC.....	18
Tabela 3 - Distribuição de <i>missing</i> ao longo dos semestres.....	19
Tabela 4 - Particionamento das amostras.....	23
Tabela 5 - Estimativas dos coeficientes para o modelo final.....	27
Tabela 6 - Pontos de cortes simulados.....	28
Tabela 7 - Percentual de acerto.....	28

ÍNDICE FIGURA

Figura 1 - Pirâmide de matrículas.....	12
Figura 2 - Processo do modelo preditivo de matrículas.....	14
Figura 3 - Definição das variáveis da base estudada.....	15
Figura 4 - Classes de níveis de odds.....	16
Figura 5 - Odds das variáveis do modelo final.....	20
Figura 6 - Agrupamento de cursos pelo valor do odds.....	22
Figura 7 - Série histórica de efetivação de matrículas.....	24
Figura 8 - Variáveis de Referência.....	26
Figura 9 - Curva ROC.....	29
Figura 10 - Nível de discriminação do teste KS.....	30

Sumário

1. INTRODUÇÃO	8
2. FUNDAMENTAÇÃO TEÓRICA	10
2.1. Gestão de Matrículas	10
2.2. Regressão Logística	12
3. MATERIAIS E MÉTODOS.....	13
4. RESULTADOS.....	18
4.1. Análise da qualidade do banco	18
4.2. Análise Preliminar	19
4.3. Amostra.....	22
4.4. O modelo	24
4.5 Resultados Finais.....	25
4.6. Discussão	30
5. CONSIDERAÇÕES FINAIS	31
5.1. Conclusões	31
5.2. Limitações	31
5.3. Trabalhos Futuros.....	32
REFERÊNCIAS.....	33

1. INTRODUÇÃO

As universidades federais estão em um processo contínuo de desenvolvimento e busca de técnicas, a fim de otimizar os recursos disponibilizados, garantindo, dessa forma, a prosperidade acadêmica e financeira. Nesse contexto, Kalsbeek (2003) nota que, no século XXI, a gestão de matrículas vem se tornando um conceito dominante para todos os líderes institucionais, isso porque essa teoria tem como objetivo a efetivação de matrículas, as quais permitirão a prosperidade acadêmica e financeira das instituições de ensino superior.

As universidades gastam recursos em demasia para ofertar vagas de matrículas para seus alunos. No entanto, nem sempre todas as vagas ofertadas são de fato aproveitadas, assim gerando o uso desnecessário de recursos. Por essa razão, o uso de modelos de gerenciamento de matrículas tem se tornado ferramenta importante para as instituições de educação superior (IES). Neste sentido, Meyer e Murphy (2003) ressaltam que as universidades não podem mais trabalhar apenas com base na intuição e improvisação, sugerindo assim o uso de alguma metodologia de planejamento estratégico.

O modelo de gerenciamento de matrículas possui uma história recente. Sua primeira implementação aconteceu na década de 70 na universidade Boston College, em Boston, Massachusetts, Estados Unidos. A intenção do uso do conceito de administração de matrículas nas instituições de nível superior foi buscar maior eficácia na busca de resultados (matrículas). Infelizmente, no Brasil o uso de sistemas de gestão de matrículas ainda é muito incipiente, não sendo encontrado nenhum modelo como referência nesta área.

O uso de técnicas a favor de uma melhor gestão sobre as matrículas ofertadas pelas instituições tem se tornado uma prática comum nas universidades, principalmente nos Estados Unidos. Comumente conhecido como “*Enrollment Management*”, os modelos de gestão de matrículas têm sido grandes aliados para as universidades otimizarem seus métodos de gestão.

Conforme constatado por Meyer e Murphy (2003), a gestão de matrículas busca uma visão mais estratégica nas instituições, integrando as práticas institucionais com os programas e as políticas desenvolvidas pela universidade. Através dessa visão mais estratégica se obtém uma melhor combinação de alunos matriculados.

Uma definição para *Enrollment Management* de acordo com Hossler e Bean (1990) seria o ato de através do planejamento estratégico influenciar a captação e retenção de alunos através de diversas atividades e práticas institucionais. Esta definição retrata um dos pontos chaves da gestão de matrículas, no que se refere à captação e à retenção de alunos. No entanto, os modelos de gestão de matrículas são mais abrangentes atuando não só para essas finalidades, mas também como um processo informativo que alinha os dados fiscais, acadêmicos, curriculares e recursos de matrículas de uma instituição com as mudanças do mercado para cumprir a missão da instituição e garantir o sucesso e a saúde fiscal ao longo prazo. Os mesmos autores, Hossler e Bean (1990), também definiram *Enrollment Management* como um conceito organizacional que descreve um conjunto sistemático de práticas destinadas a permitir que as instituições de ensino exerçam maior influência sobre suas matrículas. E esta definição reforça a ideia da vasta área de abrangência que esta ferramenta possui nas IES, uma vez que a mesma pode ser aplicada em diversos processos das instituições, otimizando-os.

Considerando a importância do gerenciamento de matrículas, o presente trabalho tem como objetivo propor um modelo estatístico para melhorar o sistema de matrículas da UFRGS, no que tange à efetivação do aluno na matrícula, usando como piloto a disciplina de Cálculo e Geometria Analítica II. Mais especificamente, será construído um modelo preditivo, no intuito de prever a efetivação da matrícula do aluno. O modelo será construído com base na técnica de Regressão Logística, de forma a calcular uma probabilidade de o aluno se matricular na disciplina de Cálculo II. No entanto, vale frisar que a ideia é gerar um modelo piloto de gerenciamento de matrículas, sendo que a disciplina de Cálculo II serve, apenas, como o ponto inicial da construção deste modelo.

O presente trabalho possui cinco seções. Nesta primeira seção, foi apresentado o tema de gestão de matrículas, contextualizando a sua importância bem como justificando sua implementação, e com base nestas informações se introduziu os objetivos do trabalho. Na seção seguinte, será apresentado o referencial teórico, discutindo a técnica estatística que será utilizada na modelagem, bem como o panorama atual das universidades, tanto no Brasil quanto no exterior. Na terceira seção, de materiais e métodos, serão explicados em detalhe as variáveis presentes no banco de dados, assim como o passo a passo da construção do modelo estatístico. Na quarta seção, serão expostos os resultados da análise e sua discussão. Por fim, na quinta seção, serão apresentadas as conclusões, limitações do estudo e também sugestões de estudos futuros.

2. FUNDAMENTAÇÃO TEÓRICA

2.1. Gestão de Matrículas

Embora já se tenha comprovações que o uso da teoria de gestão de matrículas corrobore para a otimização do sistema de matrículas das instituições de ensino superior (IES), o mesmo ainda não é amplamente utilizado. No exterior, as universidades utilizam esta técnica em seus modelos específicos, a fim de estabelecer um planejamento mais estratégico sobre as suas matrículas. A grande maioria dos modelos elaborados nos Estados Unidos são referente à captação e retenção de novos estudantes.

No Brasil, o uso dessa importante e robusta ferramenta é, ainda, incipiente. Atualmente, começam a surgir ideias de implementação da teoria em universidades brasileiras, como é o caso da Pontifícia Católica do Paraná (PUC-PR) que já possui estudos mapeando o impacto gerado pela utilização do sistema de gestão de matrículas. Embora o foco desses estudos seja voltado ao âmbito do marketing, eles mostram um interesse das IES nacionais,

indicando um possível cenário futuro de implementação de técnicas e teorias mais aprimoradas.

Os modelos encontrados em estudos semelhantes ao aqui proposto foram todos com relação a captação e retenção de alunos. Sendo assim, apenas será citado o case de sucesso da DePaul University, a maior universidade católica dos Estados Unidos, localizada em Chicago. A DePaul University impulsionou bastante o crescimento de matrículas a partir da combinação do planejamento estratégico com o marketing e a gestão de matrículas.

Foi desenvolvida uma “pirâmide de matrículas” uma espécie de fluxograma que representava a estratégia de marketing em todo o processo, desde a admissão de novos estudantes até a formação de seus alunos. Essa representação visual mostrava todos os estágios do processo de admissão da universidade, começando pelo mapeamento dos potenciais candidatos existentes no mercado até o último estágio, ou seja, a graduação dos alunos na própria universidade.

O uso da gestão de matrículas permeou todas as etapas desse processo, assim ajudando no monitoramento e na tomada de decisão. O emprego de modelos estatísticos subsidiava decisões, uma vez que identificava os alunos com maior probabilidade de serem aprovados em cada critério. E, justamente em posse dessas informações, se tornava possível dedicar os recursos humanos e financeiros aos alunos mais propensos a efetivarem as matrículas. Esse é um exemplo bastante verossímil que ilustra a aplicabilidade da técnica de gestão de matrículas. Na Figura 1, segue o fluxograma de matrículas criado por Kalsbeek (2003) para a DePaul University.



Figura 1 - Pirâmide de matrículas.

Fonte: Kalsbeek (2003).

2.2. Regressão Logística

O uso de técnicas de regressão para se estudar o comportamento de uma variável resposta (dependente) em relação as variáveis explicativas (independentes) são amplamente utilizadas em diversas áreas. O objetivo do modelo de regressão logística é entender uma variável resposta, geralmente binária, Y , com base em um ou mais preditores ($x_1, x_2, x_3, \dots, x_p$) (HILBE, 2009). A resposta binária ou dicotômica, como é chamada, é quando a variável dependente Y pode assumir, apenas, os valores 0 ou 1, sendo ($Y = 1$) comumente usado para representar a presença de certa característica com probabilidade de ocorrência igual a $\pi(x)$ e o valor 0 ($Y = 0$) a ausência da mesma, com probabilidade de $1 - \pi(x)$.

O modelo de equação da Regressão Logística é definida por:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}.$$

Como a regressão logística pertence à classe de modelos lineares generalizados (MLG) isso quer dizer que se faz necessário aplicar algum tipo de transformação para que o modelo obtenha as desejáveis propriedades dos modelos lineares (HOSMER; LEMESHOW, 1989). Assim, através da transformação logito $g(x)$ se obtém uma relação linear entre as variáveis:

$$g(x) = \ln \left[\frac{\pi(x)}{1 - \pi(x)} \right] = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p.$$

Para a estimação correta dos coeficientes do modelo da regressão logística, dado a não linearidade do modelo logístico é realizado o método de máxima verossimilhança ao invés da estimação pelos mínimos quadrados como é feito na regressão linear.

Assim, a equação do modelo de regressão logística para predição se dá através da fórmula abaixo:

$$E(Y) = \frac{1}{1 + \exp [-(\beta_0 + \sum_{j=1}^k \beta_j X_j)]}$$

Na construção do modelo algumas variáveis preditoras podem sofrer com a multicolinearidade, e isto acontece quando, duas ou mais variáveis são correlacionadas e o uso delas no modelo pode prejudicar o processo de modelagem. Segundo Agresti (2007), retirando os preditores redundantes, ou seja, aqueles que são autocorrelacionados com outros preditores, pode-se reduzir os erros padrão de outros efeitos estimados.

Existem diferentes procedimentos que podem ser utilizados para seleção das variáveis no modelo, tais como o *stepwise*, *backward* e *forward*. Esses métodos consistem em retirar (*backward*), inserir (*forward*) as variáveis explicativas no modelo e verificar a sua significância, identificando as variáveis que mais contribuem para o modelo. Este procedimento acaba por eliminar as variáveis que apresentem indícios de multicolinearidade selecionando apenas as variáveis mais significativas para o modelo.

3. MATERIAIS E MÉTODOS

Para construção do modelo preditivo de efetivação de matrículas da UFRGS, se escolheu a disciplina de MAT01354 - Cálculo e Geometria Analítica II como piloto. Essa escolha foi tomada por diversos motivos, primeiramente devido à dimensão dessa disciplina, isto é, ser oferecida para vários cursos, e

portanto, possuir diversos perfis de alunos. O segundo motivo foi por causa da disciplina possuir pré-requisitos (Cálculo I) e, sendo assim, o aluno já possui um histórico acadêmico. E, por fim, por se tratar de uma das disciplinas com o maior número de alunos matriculados na universidade.

Para ilustrar a problemática que originou o presente trabalho, o fluxograma da Figura 2 representa graficamente o processo de gerenciamento de matrículas, evidenciando o propósito do uso desta técnica nas IES.

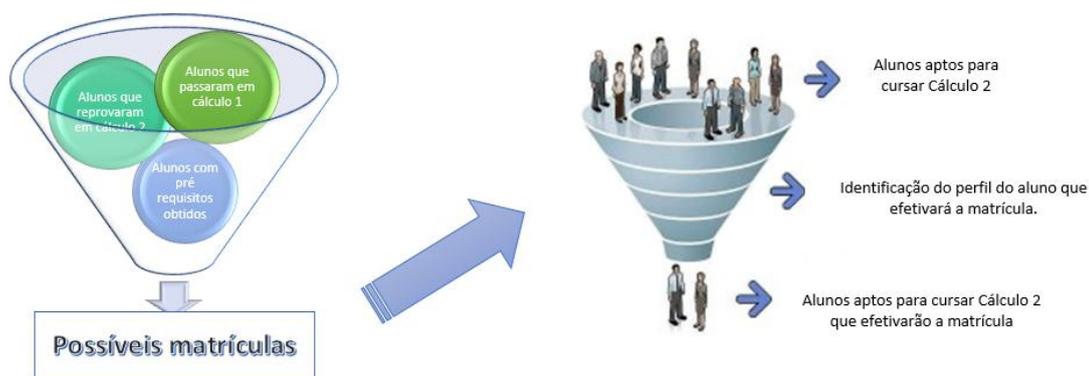


Figura 2 -Processo do modelo preditivo de matrículas.

O processo funciona como se fosse um duplo funil, no qual na primeira “filtragem” tem como objetivo delimitar os alunos aptos a cursarem a disciplina. Dessa primeira etapa se extrai um número total de possíveis matrículas, ou seja, o número de alunos que poderão cursar a disciplina de Cálculo e Geometria Analítica II (MAT01354). Cabe salientar que o sistema da UFRGS, atualmente, se encontra neste estado, ou seja, disponibiliza o número de vagas que permita a inscrição na disciplina para todos os alunos aptos a se matricularem. Mas é intuitivo imaginar que nem todos os alunos aptos irão efetivar as matrículas, e isso pelos mais diversos motivos. Então, é neste momento em que a técnica estatística entra para auxiliar na tomada de decisão da universidade. A técnica de modelagem irá tentar identificar o perfil de alunos que efetivariam a matrícula. Esta análise funciona, como se fosse um novo filtro, ponderando os alunos aptos a cursarem a disciplina de acordo com o seu perfil. Assim, se seleciona os indivíduos que obtiveram um *score* maior no modelo estatístico, ou seja, os alunos mais propensos a realizarem a matrícula.

Interessante notar que a proposta sistemática do uso da técnica de *Enrollment management* no sistema de matrículas da UFRGS se assemelha à pirâmide de matrículas da DePaul University, modelo desenvolvido por Kalsbeek (2003), que se trata de uma referência na área. Esse modelo de Kalsbeek ilustrado na Figura 1 foi desenvolvido para representar todos os processos e fases da estratégia de matrícula da universidade DePaul e, também, para facilitar sua compreensão.

Foi utilizada uma base de dados fornecida pelo Centro de Processamento de Dados (CPD) da UFRGS. A base continha o histórico de 17.827 matrículas, que no período de 2007/1 até 2014/1 estiveram aptos a cursar a disciplina de Cálculo II. A relação de variáveis presentes no estudo está descrita na Figura 3.

Nome da variável	Descrição
Ano/Semestre	Ano e semestre da informação
Aluno	Máscara de identificação do aluno
GrupoMatricula	Grupo de Matrícula (Comissões de Graduação "responsável" por aquela vaga)
NomeCurriculo	Currículo/curso que o aluno está matriculado
ReprovacoesCalc1	Contagem das reprovações no Cálculo I (MAT01353)
ReprovacoesCalc2	Contagem das reprovações no Cálculo II (MAT01354)
CodTurma	Nome (código) da turma em que o aluno se matriculou
Conceito	Conceito obtido na disciplina de Cálculo II
DenominacaoTipoRelacao	Nome da relação do aluno com a turma da disciplina (Aprovado, Reprovado, Cancelado, etc.) na disciplina de Cálculo II
EtapaDoAluno	Índice 1 do ordenamento do aluno naquele semestre
MediaConceitos	Índice 3 do ordenamento do aluno naquele semestre (média dos conceitos obtidos na universidade(I3))
CoeficienteReprovacoes	Índice 4 do ordenamento do aluno naquele semestre
MediaHarmonica	Índice 5 do ordenamento do aluno naquele semestre (Score do vestibular)
AnoIngresso	Índice 6 do ordenamento do aluno naquele semestre
SemestreIngresso	Semestre de ingresso do aluno
TimAluno	Taxa de integralização média do aluno naquele semestre
Nca	Número de créditos aprovados do aluno naquele semestre
Cd	Coeficiente de desperdício do aluno naquele semestre
CHSemestreAnterior	Carga horária de cadeiras realizadas no semestre anterior
CHAprovSemestreAnterior	Carga horária de cadeiras aprovadas no semestre anterior
Sexo	Sexo do aluno (a)
Idade	Idade do aluno (a)
AlunoCalc2Periodo	Indicador se o aluno efetivou matrícula na disciplina de Cálculo 2

Figura 3 - Definição das variáveis da base estudada.

A partir desta base foi construído o modelo de gestão de matrículas e após a sua construção devemos testar o seu poder preditivo em uma outra amostra oriunda da mesma população. Este passo segue como garantia para que o modelo possa ser generalizado para a população. Assim, verifica-se o comportamento do modelo, isto é, a sua eficácia quando alimentado por uma outra amostra da população sem ser a que o ajustou.

Dessa forma, adotou-se a abordagem de particionar os dados em duas amostras, a primeira, para a análise, isto é, para a construção do modelo. E a segunda para a validação, para verificar-se o modelo está bem ajustado aos dados, testando-o com uma base de dados diferente da que o moldou. Normalmente esse particionamento de amostras é realizado com a proporção de 80% para a amostra de análise e 20% para a de validação.

Por se tratar de uma base de dados com um grande volume de informações, verificar a associação das possíveis variáveis preditoras com a variável dependente (efetivou matricula) através do teste de associação do Qui-quadrado não seria o mais apropriado, uma vez que a significância do teste (que indica a presença de associação entre as variáveis) seria devido ao tamanho da amostra e não, necessariamente, pela relação das variáveis testadas. Devido a isso, se analisou a associação das variáveis independentes através dos *Odds*.

Sendo assim o primeiro passo, antes de começar as análises das informações do banco de dados, foi da escolha das variáveis que entraram na análise, podendo vir a integrar o modelo final. Com o uso de tabelas de contingência, calcula-se o *odds* associado aos diferentes atributos (níveis) das variáveis preditoras, dividindo-se o percentual de alunos que efetivaram a matrícula pelo percentual dos que não efetivaram para cada atributo. Como regra geral, Lewis (1992) propõe que os atributos sejam agrupados, segundo o valor do *odds*, em sete classes, conforme apresentado na Figura 4.

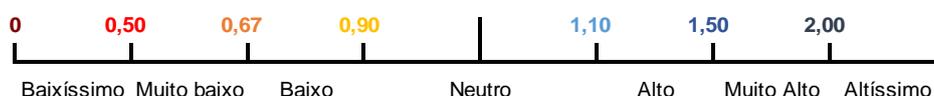


Figura 4 - Classes de níveis de *Odds*.

Após a seleção das variáveis que farão parte da análise, passa-se para a criação de uma variável *dummy* para cada agrupamento construído. Com esse artifício evitam-se problemas decorrentes da não linearidade dos atributos na análise múltipla.

Para analisar a qualidade do modelo gerado, geralmente são utilizadas duas medidas de desempenho, sendo elas o teste de Kolmogorov-Smirnov (KS) para duas amostras e o percentual de classificações corretas do modelo.

O teste de Kolmogorov-Smirnov (KS) para duas amostras é um teste não paramétrico que visa testar os scores (probabilidades) de dois grupos para verificar se os grupos são distintos. O teste é calculado com base nas probabilidades estimadas pela regressão logística dos dois grupos, os que efetivaram a matrícula e os que não efetivaram. O teste consiste em encontrar a distância máxima entre as proporções acumulada ao longo das probabilidades (scores) para os dois grupos, isto é, para o grupo de alunos que efetivaram a matrícula e o grupo de alunos que não efetivaram a matrícula (BOTELHO; TOSTES, 2010).

Segundo Andrade e Oliveira (2002), a discriminação do teste KS pode ser classificada de acordo com a Tabela 1.

Tabela 1 - Nível de discriminação do teste KS.

Valores de KS	Nível de Discriminação
Abaixo de 20%	Baixa discriminação
De 20% a 30%	Discriminação Aceitável
De 30% a 40%	Boa Discriminação
De 40% a 50%	Excelente discriminação
Acima de 50%	Não são muito comuns

O percentual de classificação correta do modelo é constituído pela soma de acertos de classificação de alunos que efetivaram a matrícula e o acerto dos que não efetivaram. Comumente estas razões de acerto são chamadas de sensibilidade e especificidade. A sensibilidade é a proporção de acerto na previsão da ocorrência de um evento nos casos em que ele de fato ocorreu

(efetivaram matrícula). A especificidade é a proporção de acerto na previsão da não ocorrência de um evento nos casos em que ele de fato não ocorreu (não efetivaram a matrícula). Dessa forma, essas duas fontes analisadas conjuntamente formam o percentual de classificação correta do modelo. Além desta medida, é possível analisar a curva ROC (do inglês *Receiver Operating Characteristic*), que se trata de um gráfico onde no eixo abcissas se encontra 1 - especificidade e no eixo das ordenadas a sensibilidade. De acordo com Maroco (2007), a capacidade de discriminação modelo pode ser avaliada de acordo com a Tabela 2.

Tabela 2 - Nível de discriminação da curva ROC.

Valores da área sobre a curva	Nível de Discriminação
ROC = 0,5	Sem poder discriminativo
Entre 0,5 e 0,6	Discriminação Fraca
Entre 0,7 e 0,8	Discriminação Aceitável
Entre 0,8 e 0,9	Discriminação Boa
Acima de 0,9	Discriminação Excepcional

4. RESULTADOS

4.1. Análise da qualidade do banco

O primeiro passo da análise foi verificar a presença de dados faltantes (*missing*), bem como a presença de valores atípicos (*outliers*) para as variáveis predictoras. Com relação aos valores atípicos, embora se tenha observado a existência de alguns valores discrepantes, os mesmos permaneceram na análise, pois se tratavam de casos prováveis e com baixa frequência, não exercendo grandes impactos na análise. Para os valores ausentes, tentou-se mapeá-los a fim de identificar o motivo da ausência de informação para essas ocorrências. Para isso, realizou-se uma distribuição de dados faltantes pelos anos/semestres da base de dados constatando que algumas variáveis possuíam um alto número de *missing* justamente por serem tratadas de variáveis novas que foram adotadas recentemente pela universidade. Sendo assim só

possuem valores a partir de um certo período (no caso, a partir de 2012/1). As variáveis que se enquadram nesta situação são: “Coeficiente de Desperdiço (CD)”; “Taxa de Integralização Média (TIM)” e “Número de Créditos Aprovados (NCA)”. A Tabela 3 evidencia essa constatação.

Ainda assim, permaneceu-se com essas variáveis no estudo, isto é, aplicou-se a mesma análise preliminar nessas variáveis como foi aplicado para as demais. No entanto, quando gerados os modelos, os que continham alguma dessas variáveis acabaram por obter um número de valores válidos (n) bem reduzido devido ao alto índice de valores ausentes dessas variáveis. Por consequência, o modelo ótimo não conteve essas variáveis, uma vez que com as mesmas obtinha uma pior taxa de acerto total.

Tabela 3 - Distribuição de *missing* ao longo dos semestres.

Ano/Semestre	Coeficiente de Desperdiço	TIM	NCA	Total de alunos
2007/1	1.010	1.010	1.010	1.010
2007/2	1.165	1.165	1.165	1.165
2008/1	995	995	995	995
2008/2	1.285	1.285	1.285	1.285
2009/1	1.045	1.045	1.045	1.045
2009/2	1.149	1.149	1.149	1.149
2010/1	1.002	1.002	1.002	1.002
2010/2	1.262	1.262	1.262	1.262
2011/1	1.042	1.042	1.042	1.042
2011/2	1.336	1.336	1.336	1.336
2012/1	27	27	27	1.155
2012/2	74	74	74	1.418
2013/1	42	42	42	1.136
2013/2	57	57	57	1.513
2014/1	50	50	50	1.314
Total de <i>missing</i>	11.541	11.541	11.541	0
Total Válido	6.286	6.286	6.286	17.827

4.2. Análise Preliminar

Para identificar as variáveis que poderiam vir a entrar no modelo, se calculou os *odds*, isto é, o comportamento dos níveis de respostas de cada variável preditora com relação à variável dependente. A partir desses cálculos,

elegem-se as possíveis variáveis preditoras. A Figura 5, mostra o comportamento do *odds* para as quatro, das cinco variáveis presentes no modelo final. Podemos reparar que as variáveis carga horária do semestre anterior e a taxa de aprovação do semestre anterior possuem o mesmo padrão. Já no caso do tempo da faculdade percebe-se que o comportamento é estável e converge para valores próximo de 0,2 conforme aumenta o tempo de permanência na faculdade.

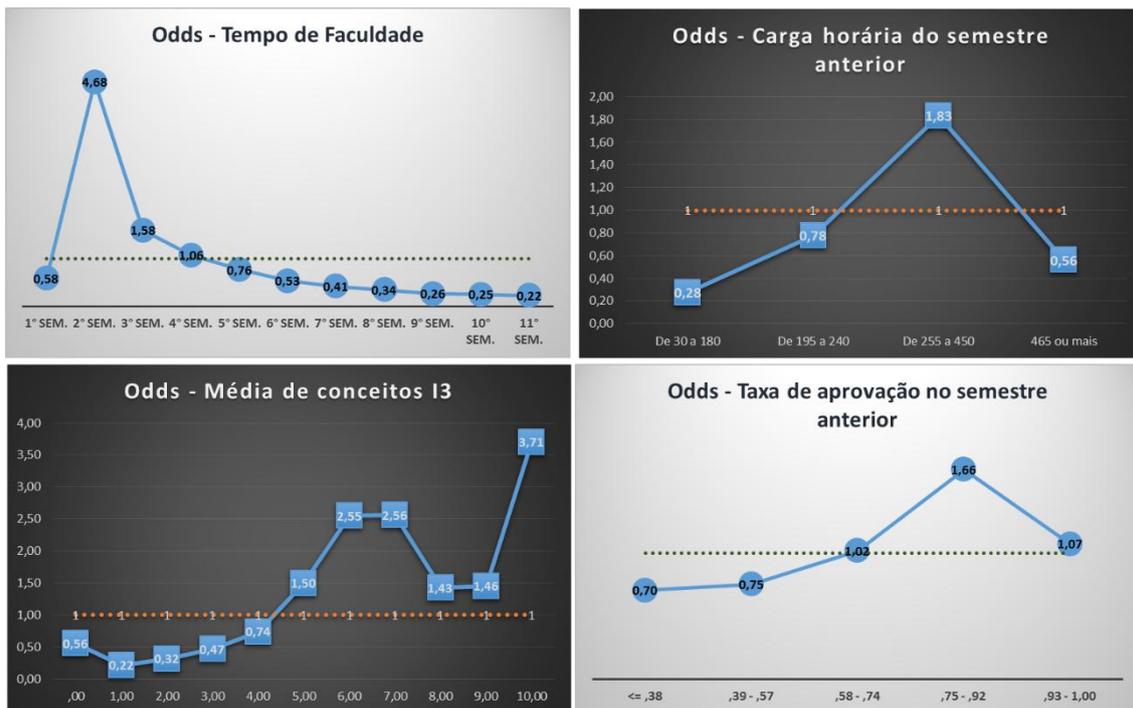


Figura 5 – Odds das variáveis do modelo final.

Ainda nessa etapa, das variáveis originais foram computadas duas novas, a primeira sendo o tempo de faculdade, isto é, o tempo de permanência do aluno na universidade (em anos). Essa transformação foi elaborada através da diferença entre o ano e semestre na qual a informação do aluno consta no banco subtraída do ano e semestre de ingresso do aluno na UFRGS. Já a segunda variável gerada foi a taxa de aprovação do semestre anterior, que se trata da razão de carga horária aprovada sobre o total de carga horária matriculada no semestre anterior. Além das variáveis computadas foram realizadas algumas categorizações para as variáveis: reprovações em cálculo 1; reprovações em cálculo 2; grupo de matrícula (cursos); média de conceitos

(I3); coeficiente de reprovações; média harmônica (nota no vestibular); tim do aluno; número de créditos aprovados; carga horária do semestre anterior; etapa do aluno; coeficiente de desperdício. Assim, as únicas variáveis que foram utilizadas na sua forma original foram idade e sexo, não sendo necessária nenhuma categorização.

Das 5 variáveis que entraram no modelo final, 4 delas eram categorizadas. As variáveis que foram agrupadas seguem abaixo:

- Carga horária do semestre anterior, esta variável foi categorizada nas respostas: Sem Informação; De 30 a 180; De 195 a 240; De 255 a 450 e 465 ou mais.
- Taxa de aprovação, com as respostas categorizadas em: Sem Informação; Menor que 38%; De 38% a 57% de aprovação; De 58% a 74% de aprovação; De 75% a 92% de aprovação e De 93% a 100% de aprovação.
- I3 - Média dos conceitos obtidos na universidade, agrupados em valores discretos de 0 a 10.
- Agrupamento de cursos – Cursos agrupados de acordo com o Odds, isto é, com a propensão de efetivar a matrícula. A Figura 6 mostra os cursos respectivos para cada nível de odds. Além das categorias de acordo com o odds também houve cursos no qual não se tinha a informação, isto é, o curso não foi definido. Para estes há a categoria de “Sem Informação”.

Baixíssimo	Alto
BIOMEDICINA	BACHARELADO EM FÍSICA: MATERIAIS E NANOTECNOLOGIA
DESIGN VISUAL	BACHARELADO EM QUÍMICA
Muito Baixo	ENGENHARIA DE ALIMENTOS
BACHARELADO EM FÍSICA: FÍSICA COMPUTACIONAL	ENGENHARIA METALÚRGICA
BACHARELADO EM FÍSICA: PESQUISA BÁSICA	QUÍMICA INDUSTRIAL
CIÊNCIAS ATUARIAIS - NOTURNO	Muito Alto
DESIGN DE PRODUTO	ENGENHARIA CIVIL
LICENCIATURA EM FÍSICA	ENGENHARIA DE MATERIAIS
LICENCIATURA EM MATEMÁTICA	ENGENHARIA DE PRODUÇÃO
LICENCIATURA EM QUÍMICA	ENGENHARIA HÍDRICA
Baixo	FORMAÇÃO EM QUIMICA
BACHARELADO EM ESTATÍSTICA	Altíssimo
BACHARELADO EM FÍSICA: ASTROFÍSICA	BACHARELADO EM ENGENHARIA FÍSICA
BACHARELADO EM MATEMÁTICA - ÊNFASE MATEMÁTICA APLIC COMPUTACIONAL	ENGENHARIA AMBIENTAL
ENGENHARIA CARTOGRÁFICA - NOTURNO	ENGENHARIA DE COMPUTAÇÃO
Neutro	ENGENHARIA DE CONTROLE E AUTOMAÇÃO
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO	ENGENHARIA DE ENERGIA
BACHARELADO EM FÍSICA	ENGENHARIA ELÉTRICA
BACHARELADO EM MATEMÁTICA- ÊNFASE MATEMÁTICA PURA	ENGENHARIA MECÂNICA
ENGENHARIA DE MINAS	ENGENHARIA QUÍMICA

Figura 6 - Agrupamento de cursos pelo valor do Odds

4.3. Amostra

Como já informado na seção 3, o banco foi separado em duas amostras, sendo a primeira, para análise, ou seja, a base que o modelo será construído. Essa amostra, na qual se realizou toda a modelagem, era composta por 14.262 ocorrências, das quais 9.268 de alunos que efetivaram a matrícula e 4.994 que não efetivaram. Isso representa a proporção de 65% de alunos que efetivaram a matrícula frente a 35% de não efetivações. Essa proporção é a mesma encontrada na base em sua totalidade e por isso foi mantida na hora da seleção das amostras, assim mantendo a proporção original. Já a amostra de validação ficou formada por 2.317 efetivações e 1.248 matrículas não efetivadas, mesma proporção já explanada. É importante salientar que a amostra de validação serve para verificar a qualidade do modelo, uma vez que os dados são retirados da mesma população. Assim, se espera que não haja

grandes perdas de assertividade do modelo frente a esta amostra, o que representaria que o modelo está bem ajustado, assim podendo ser generalizado para o restante da população. A Tabela 4 ilustra esta repartição.

Tabela 4 - Particionamento das amostras.

			Efetivou matrícula		Total
			Não	Sim	
Amostra	Análise	Frequência	4994	9268	14262
		% da linha	35,0%	65,0%	100,0%
		% da coluna	80,0%	80,0%	80,0%
		% do total	28,0%	52,0%	80,0%
	Validação	Frequência	1248	2317	3565
		% da linha	35,0%	65,0%	100,0%
		% da coluna	20,0%	20,0%	20,0%
		% do total	7,0%	13,0%	20,0%
Total	Frequência	6242	11585	17827	
	% da linha	35,0%	65,0%	100,0%	
	% da coluna	100,0%	100,0%	100,0%	
	% do total	35,0%	65,0%	100,0%	

A proporção de efetivações de matrículas neste período estudado foi de 65%. E esta foi a proporção utilizada no particionamento da amostra, assim mantendo o “perfil” de matrículas. No entanto, a Figura 7 mostra a série histórica do percentual de efetivações por semestre, evidenciando uma sazonalidade de efetivações de matrículas, sendo que semestres pares possuem maior demanda de alunos. O que faz sentido com base na disciplina estudada.

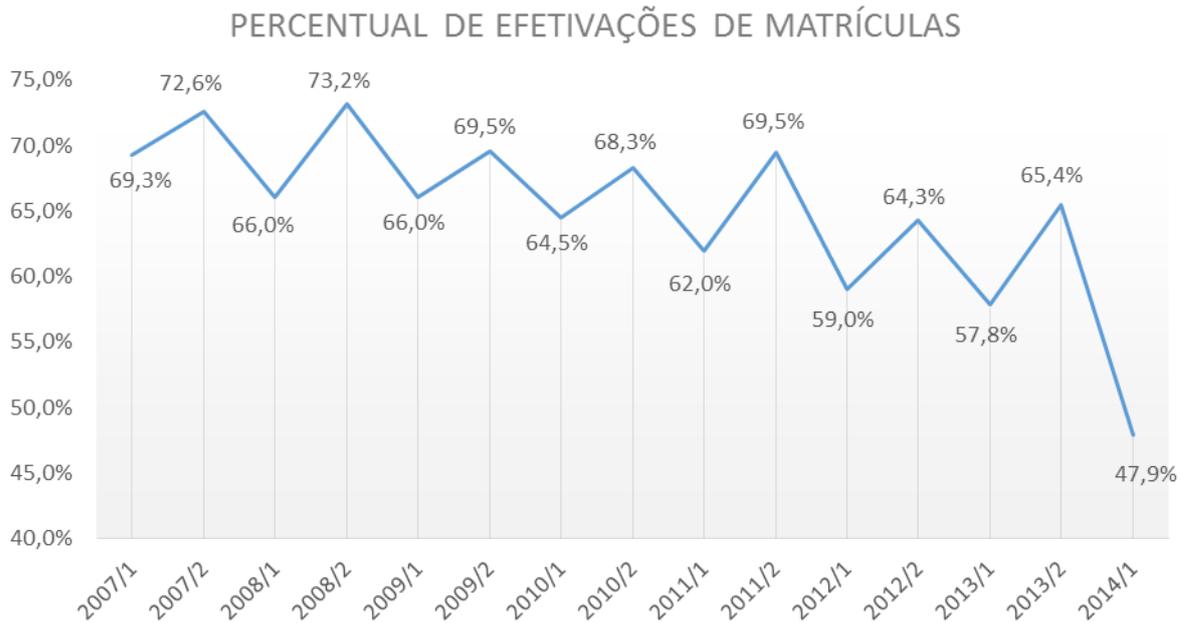


Figura 7 – Série Histórica de efetivação de matrículas

4.4. O modelo

Um pressuposto importante da técnica de regressão logística que deve ser testado é o da multicolinearidade, quando se usa a técnica de *stepwise* como método de seleção acaba-se por reduzir a ocorrência de multicolinearidade porque por meio desta técnica, se prioriza e seleciona as variáveis preditoras mais significativas para o modelo. Uma maneira inicial para verificar a existência da multicolinearidade é justamente analisando as correlações entre as variáveis preditoras, e através deste procedimento se constatou a ausência de indícios de multicolinearidade entre as variáveis preditoras. Porém este método é apenas exploratório, sendo necessário complementá-lo para poder-se concluir pela ausência de multicolinearidade.

Com base nas possíveis variáveis preditoras selecionadas na análise dos *odds*, se testou diversos modelos a fim de verificar o que melhor se ajustava ao banco. O procedimento de escolha do modelo foi realizado com o uso da técnica de regressão logística sendo elencados 18 modelos ao todo. Inicialmente se realizou um modelo para cada variável preditora, verificando assim o poder explicativo de cada uma das variáveis sem a influência de

nenhuma das demais. Posteriormente se verificou o modelo completo, isto é, com a presença de todas as variáveis significativas do modelo. No entanto, o modelo definido foi encontrado a partir da utilização das variáveis nas quais, em seus modelos singulares, obtiveram alto percentual de classificação correta, sendo, portanto, escolhido como modelo final. As variáveis inclusas neste modelo foram:

- Carga horária do semestre anterior;
- Taxa de aprovação;
- Tempo de faculdade;
- Agrupamento de cursos;
- Média de conceitos (I3).

Para a seleção das variáveis inclusas no modelo final, utilizou-se o método de seleção automática *backward* que consiste em iniciar o modelo com todas as variáveis preditoras, e a cada passo retirar uma variável. Assim, o método termina, quando todas as variáveis no modelo são significativas, dessa forma otimizando o percentual de acerto total do modelo.

4.5 Resultados Finais

A interpretação da exponencial do beta se dá através da magnitude. Quanto maior for a magnitude, maior será a chance de efetivação da matrícula. Friso que a comparação da magnitude para cada variável é feita através da categoria de referência utilizada em cada variável. Assim a Figura 8 ilustra as categorias de referência utilizada em cada variável inclusa no modelo.

Variáveis do Modelo	Categoria de Referência
DCURSO Muito baixo ou Baixíssimo	DCURSO Neutro
DCURSO Alto	
DCURSO Muito alto	
DCURSO Altíssimo	
DI3 =1	DI3 = 0
DI3 =2	
DI3 =3	
DI3 =4	
DI3 =6	
DI3 =7	
DI3 =8	
DI3 =9	
D_TX_Menos de 38% de aprovação	D_TX_De 58% a 74% de aprovação
D_TX_De 38% a 57% de aprovação	
D_TX_De 75% a 92% de aprovação	
D_TX_Acima de 93% de aprovação	
D_CH_S_A_De 30 a 180 horas	D_CH_S_A_Acima de 465 horas
D_CH_S_A_De 195 a 240 horas	
D_CH_S_A_De 255 a 450 horas	

Figura 8 - Variáveis de referência

Através da Tabela 5 tem-se que a variável de maior influência na composição do *score* logístico é a carga horária do semestre anterior (D_CH_S_A), de forma que quanto maior for a carga horária (categorias) maior será a influência no *score* logístico. Assim, os alunos com carga horária no semestre anterior de 255 a 450 horas possuem uma $\exp(\beta)$ de 4,948, ou seja, a chance de um aluno efetivar a matrícula dado que obteve no semestre anterior carga horária entre 255 e 450 horas é aproximadamente 395% maior que a chance dos alunos com carga horária acima de 465 horas. Também constata-se que os alunos matriculados nos cursos classificados com *odds* muito baixo e baixíssimo (DCURSO Muito baixo ou Baixíssimo) possuem 31,5% da chance dos alunos matriculados em cursos com *odds* neutro de efetivarem a matrícula. Ao passo que esta mesma comparação, porém com os alunos pertencentes aos cursos classificados com *odds* altíssimo possuem 222% a chance dos alunos de cursos neutro.

Tabela 5 - Estimativas dos coeficientes para o modelo final.

	B	Sig.	Exp(B)	95% C.I. para EXP(B)	
				Inferior	Superior
Tempo_de_faculdade	-0,194	0,000	0,823	0,802	0,845
DCURSO12	-1,278	0,000	0,279	0,246	0,315
DCURSO5	0,148	0,064	1,160	0,992	1,356
DCURSO6	0,426	0,000	1,531	1,323	1,771
DCURSO7	0,660	0,000	1,935	1,683	2,225
D131	-1,341	0,000	0,261	0,195	0,351
D132	-0,883	0,000	0,414	0,350	0,490
D133	-0,496	0,000	0,609	0,521	0,712
D134	-0,266	0,001	0,766	0,655	0,897
D136	0,433	0,000	1,541	1,285	1,849
D137	0,218	0,032	1,244	1,019	1,517
D138	-0,460	0,000	0,631	0,514	0,775
D139	-0,522	0,000	0,593	0,462	0,762
D_TX_1	0,252	0,002	1,286	1,094	1,512
D_TX_2	0,229	0,006	1,258	1,067	1,482
D_TX_4	0,559	0,000	1,749	1,422	2,151
D_TX_5	0,531	0,000	1,701	1,444	2,003
D_CH_S_A_1	0,486	0,000	1,625	1,399	1,888
D_CH_S_A_2	1,192	0,000	3,295	2,789	3,894
D_CH_S_A_3	1,599	0,000	4,948	4,317	5,671
Intercepto	0,249	0,003	1,283		

O primeiro índice utilizado para medir a qualidade do modelo foi analisar o percentual de classificação correta do modelo. Porém, primeiramente tem que se definir o ponto de corte do modelo, isto é, definir a partir de qual probabilidade será considerado que o aluno irá efetivar a matrícula. Para isso se testou valores diferentes de ponto de corte, verificando o acerto do modelo para as três medidas (sensibilidade, especificidade e % total de acerto). Os valores de ponto de corte foram escolhidos, apenas para verificar o que aconteceria com o modelo se o critério de classificação fosse mais brando ou mais exigente. Os valores simulados e suas medidas estão dispostos na Tabela 6.

Tabela 6 - Pontos de cortes simulados.

	Amostra de análise					
	0.4	0.5	0.6	0.65	0.7	0.8
Sensibilidade	49,9%	60,6%	73,6%	78,8%	82,9%	91,2%
Especificidade	91,6%	88,3%	82,6%	78,1%	74,7%	64,6%
% de acerto	77,0%	78,6%	79,5%	78,3%	77,5%	74,0%

Com base na simulação se definiu o ponto de corte no valor de 0.60 pois neste valor otimizou-se o percentual de acerto total. A Tabela 7 mostra a comparação dos valores preditos com os observados no modelo proposto com a ponto de corte escolhido. Percebe-se que o modelo possui maior assertividade para identificar os casos de efetivação de matrícula (82,6%) o que para o presente trabalho é mais interessante, uma vez que o intuito da gestão de matrículas é reduzir o número de cadeiras ociosas, mas procurando garantir vagas para os alunos que possuem o perfil de se matricular no semestre analisado. A Tabela 7 mostra o percentual de acerto do modelo. No qual se verifica que o modelo prevê a efetivação de 8.976 matrículas quando na realidade houve 9.268, ou seja, faltariam vagas para alguns, porém se economizaria 5.286 vagas que estariam, em sua grande maioria, ociosas.

Tabela 7 - Percentual de acerto.

			Classificação da logística		Total
			Não efetivou	Efetivou	
Efetivou matrícula	Não	Contagem	3676	1318	4994
		% dentro de Efetivou_matricula	73,6%	26,4%	100,0%
	Sim	Contagem	1610	7658	9268
		% dentro de Efetivou_matricula	17,4%	82,6%	100,0%
Total		Contagem	5286	8976	14262
		% dentro de Efetivou_matricula	37,1%	62,9%	100,0%

Com base na especificidade e na sensibilidade se obtém a curva ROC, que se trata de uma representação gráfica da variável resposta com a

característica de interesse (efetivação da matrícula) *versus* a variável resposta sem a característica de interesse (não efetivo).

A área sobre a curva foi de 0,852 o que indica uma boa discriminação do modelo de regressão. A Figura 9 mostra a curva ROC.

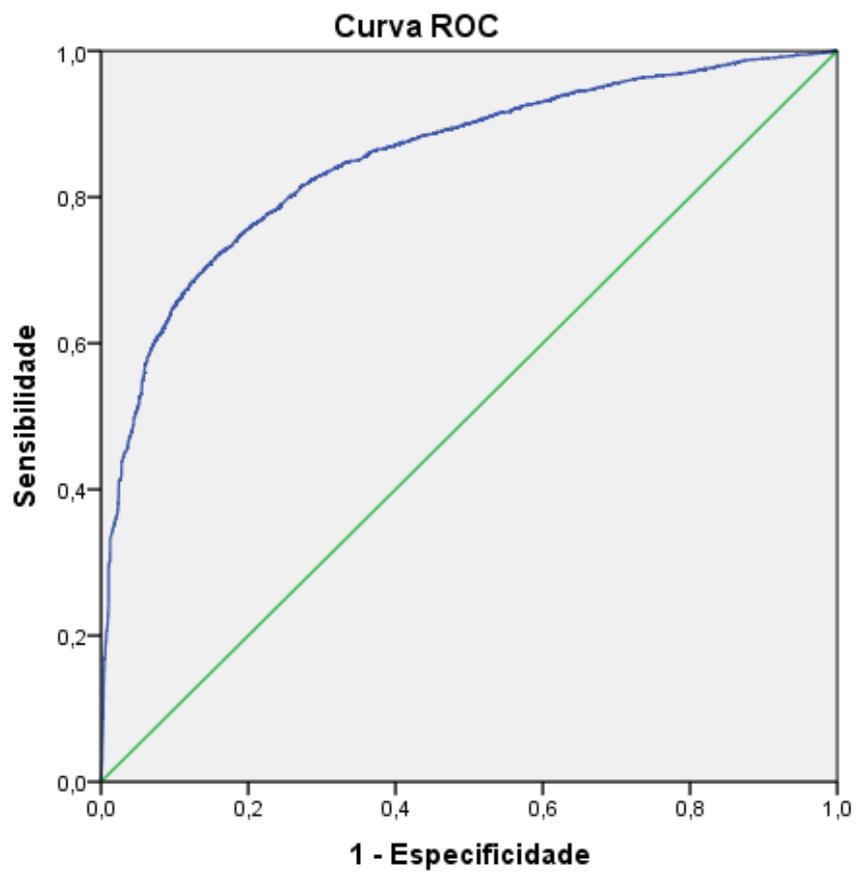


Figura 9 - Curva ROC.

Realizando o teste KS para duas amostras obteve-se o valor de 56,3%. O que representa uma excelente discriminação entre os grupos de alunos que efetivaram a matrícula dos que não efetivaram, ou seja, os grupos (efetivaram ou não) são distintos. A Figura 10 mostra o resultado do teste KS.

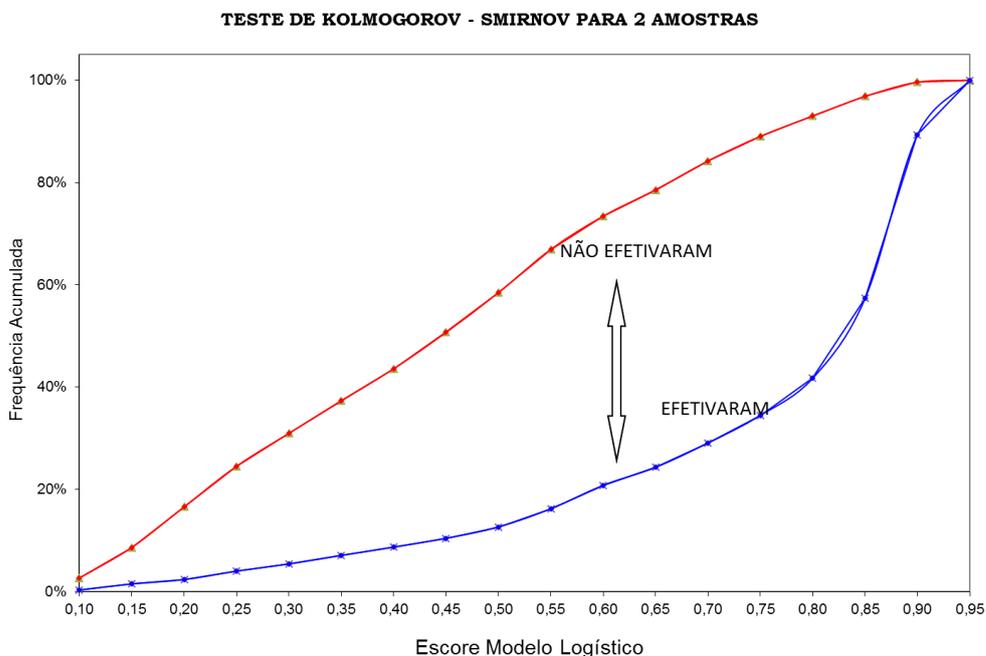


Figura 10 - Nível de discriminação do teste KS.

4.6. Discussão

É interessante avaliar a coerência do modelo escolhido, isto é, verificar se as variáveis incluídas no modelo estatístico condizem com a prática por trás do modelo, ou seja, tentar analisar se o modelo proposto contém as variáveis que se imaginavam exercerem influência na classificação do perfil dos alunos que venham a efetivar a matrícula. E pensando neste sentido o modelo possui uma coerência prática satisfatória, isso porque as variáveis identificadas com maior grau de associação fizeram total sentido com a teoria. Contextualizando com a prática as cinco variáveis definidas como preditoras no modelo, se percebe que as mesmas possuem, no âmbito acadêmico, caráter discriminatórios de perfil de alunos e, isso evidencia o acerto do modelo proposto. Um exemplo de que os resultados encontrados no estudo do

comportamento da variável preditora frente a variável resposta foi o tempo de faculdade que apresentou ter associação decrescente, ou seja, conforme aumenta o tempo de faculdade do aluno, menor será a probabilidade do mesmo se matricular na disciplina de Cálculo II, isso tanto no modelo logístico quanto nos *odds*. Esse resultado é intuitivo, uma vez que a disciplina de cálculo ocorre, em sua grande maioria, até a metade do curso, ou seja, não se espera que os alunos que estão a mais tempo na faculdade venham de fato a cursar a disciplina de Cálculo.

5. CONSIDERAÇÕES FINAIS

5.1. Conclusões

A gestão de matrículas possibilita um melhor controle e aproveitamento dos recursos das instituições. Embora ainda seja incipiente no Brasil, o uso da técnica tende a gerar bons resultados para as instituições de ensino superior, poupando recursos financeiros e de pessoal.

Atualmente o sistema de concessão de vagas utilizado pela UFRGS visa atender todos os alunos aptos a se matricularem, porém é intuitivo e o estudo comprovou que nem todos os alunos concretizam a matrícula. Dessa forma, o uso de um modelo de gestão de matrículas integrado com as políticas e práticas da universidade quando administradas com planejamento estratégico tendem otimizar o sistema de matrículas da UFRGS.

5.2. Limitações

Por se tratar de um estudo piloto algumas limitações foram encontradas. Dentre as variáveis estudadas algumas delas possuíam uma série histórica menor do que o período estudado, não sendo utilizadas no modelo preditivo. Dentre estas variáveis constavam a TIM, o CD, o NCA. Possivelmente estas

variáveis seriam úteis para a análise, no entanto devido ao alto índice de *missing* se tornou inviável seu uso neste primeiro momento.

Por se tratar de uma variável dependente dicotômica (efetivação da matrícula), o uso da técnica de regressão logística seria o mais adequado. No entanto, os alunos que compõem o estudo podem possuir mais de uma ocorrência no banco devido as reprovações na disciplina de Cálculo II e a não matrícula quando disponível. Devido a isso, faz-se necessário levar em consideração a não independência dos casos. Essa implicação necessita de modelos mais robustos com é o caso do uso de equações de estimação generalizadas (GEE).

5.3. Trabalhos Futuros

Dadas as limitações apontadas, sugere-se como estudo futuro aplicar este modelo em um lapso temporal mais recente (período de 2012/1 até 2014/2), a fim de conseguir utilizar as variáveis como a TIM, que possuem série histórica recente, testando os seus efeitos.

Como a ideia da gestão de matrículas é criar um sistema único, seria essencial replicar este estudo com uma maior abrangência, isto é, envolvendo mais disciplinas. Assim se teria um resultado mais assertivo, pois conseguiria ajustar um modelo considerando todas as especificidades das diversas disciplinas oferecidas pela universidade.

Aplicar está técnica para desenvolver um modelo preditivo a fim de identificar os alunos que, de fato, aproveitariam a matrícula, isto é, dado que o aluno se matriculou analisar se ele irá ou não aproveitar a matrícula. Uma utilização mais eficiente dos recursos poderia ser obtida ao integrar esse modelo de aproveitamento com o modelo de efetivação de matrícula apresentado nesse trabalho.

REFERÊNCIAS

AGRANONIK, M. **Técnicas de Diagnóstico Aplicadas ao Modelo de Regressão Logística**. 2006. Universidade Federal do Rio Grande do Sul: monografia de conclusão de curso. 68 pg. 2006.

AGRESTI, A. **An Introduction to Categorical Data Analysis**, SecoAGR nd Edition. 400 pg. 2007.

ANDRADE, F. W. M.; OLIVEIRA, J. G. C. Comparação entre Medidas de Performance de Modelos de *Credit Scoring*. **Revista Tecnologia de Crédito**. n. 33, 2002.

BOTELHO, D.; TOSTES, F.D. Modelagem de Probabilidade de Churn. **RAE**, São Paulo, v. 50, n. 4, p. 396-410. out./dez. 2010.

HILBE, J. M. **Logistic regression models**. Boca Raton: CRC Press, 2009.

HOSMER, D. W.; LEMESHOW, S. **Applied logistic regression**. New York: John Wiley & Sons, 1989.

HOSSLER, D.; BEAN, J. B. **The Strategic Management of College Enrollments**. San Francisco, Cal: Jossey-Bass Publishers, 1990.

KALSBEEK, D. H. **Marketing e gerenciamento de matrículas na educação superior**. In: MEYER Jr. V.; MURPHY, J. P. **Dinossauros, gazelas e tigres: novas abordagens da administração universitária. Um diálogo Brasil e Estados Unidos**. 2.Ampl. Florianópolis: Insular, 2003.

LEWIS, E. M. **An Introduction to Credit Scoring**. San Rafael: Fair, Isaac and Co., Inc. 1992.

MAROCO, J. **Análise Estatística – com Utilização do SPSS**. Lisboa: Ed. Sílabo, 2007.

MEYER Jr. V.; MURPHY, J. P. **Dinossauros, gazelas e tigres: novas abordagens da administração universitária. Um diálogo Brasil e Estados Unidos**. 2.Ampl. Florianópolis: Insular, 2003.