

O DESENVOLVIMENTO DE UM MODELO MATEMÁTICO PARA A PREVISÃO DA APROVAÇÃO DA DISCIPLINA DE CÁLCULO 1 UTILIZANDO REGRESSÃO LOGÍSTICA

Igor Rodrigues De Nardi – ir.denardi@hotmail.com⁽¹⁾

Shirley Cristina Cabral Nascimento – sccn@ufpa.br⁽¹⁾

Alessandra Macedo de Souza – amacedos@ufpa.br⁽¹⁾

Universidade Federal do Pará⁽¹⁾
Instituto de Tecnologia
Rua Augusto Corrêa 01 Guamá
66075-110 Belém – Pará – Brasil

Resumo: Este estudo visa identificar quais fatores são significativos no desempenho dos calouros dos cursos de engenharia da Universidade Federal do Pará (UFPA) na disciplina de Cálculo 1, a qual possui um dos maiores índices de reprovação dentro das universidades. Utilizando o software IBM SPSS Statistics, pode-se criar um banco de dados de variáveis dicotômicas e com isso fazer a regressão logística que, por meio dela, obtém-se uma função que possui como entrada as variáveis categóricas e como saída a probabilidade de passar ou não em Cálculo 1. Para tanto, entrevistou-se 281 discentes que já cursaram a disciplina, respondendo à 13 perguntas (variáveis categóricas) cada.

Palavras-chave: Cálculo 1. Variável Categórica. Regressão Logística.

1 INTRODUÇÃO

Considerável parcela dos alunos que iniciam um curso de engenharia se depara com uma situação extremamente desconfortável: aprender cálculo. Vindo diretamente do Ensino Médio, o aluno dificilmente consegue acompanhar as exigências do novo ambiente acadêmico (CASTRO, 2014). Em uma pesquisa feita pela UNICAMP, levantou-se um histórico de 12 anos (1997 a 2009) de informações relativas à esta disciplina que permitiu constatar taxas de até 77,5% de reprovação e evasão dos estudantes (GARZELLA, 2013). Com o objetivo de amenizar este problema, a Pró-Reitoria de Ensino e Extensão (PROEX), juntamente com o Instituto de Tecnologia (ITEC) da Universidade Federal do Pará (UFPA), Campus Belém, criaram o Programa de Cursos de Nivelamento de Aprendizagem (PCNA) para as ciências básicas das engenharias, o qual oferece cursos básicos de matemática, física e química, preferencialmente aos calouros e não aprovados nas disciplinas básicas à engenharia (disciplinas de cálculo, de física e de química do projeto pedagógico de cada curso) (RODRIGUES et al., 2012). Sabe-se também que além das dificuldades com a aprendizagem das disciplinas das ciências básicas, existem os problemas do dia a dia dos alunos, como transporte, renda familiar, moradia entre outros, e que se entende que esses atributos influenciam diretamente na aprovação ou não de qualquer disciplina.

Este trabalho tem como objetivo desenvolver um modelo matemático baseado em regressão logística multivariável que apontará, entre as dificuldades que os alunos passam no dia a dia, a de maior peso na aprovação ou não da disciplina de cálculo 1. Uma pesquisa de campo entre os alunos da engenharia do ITEC que já cursaram a disciplina de cálculo 1 será

realizada. Esses alunos responderão a um questionário cujas perguntas serão sobre gosto pelos estudos, renda, moradia, transporte e participação no PCNA ou outros projetos. Essas questões formarão o conjunto das variáveis qualitativas do modelo matemático pois as respostas a essas perguntas serão apenas sim ou não. Um método para quantificar essas variáveis qualitativas é atribuir valores para elas, no caso 1 ou 0 (indicando a presença ou a ausência de um dado atributo) que são conhecidas na literatura existente como variáveis categóricas. O software IBM SPSS foi utilizado para realizar as análises bem como obter a função que descreve a possibilidade de um aluno passar ou não na disciplina de Cálculo 1 na primeira vez que cursa essa disciplina. De posse do modelo matemático será possível analisar o perfil de um aluno calouro do curso de engenharia e consequentemente informa-lo onde estará sua principal dificuldade na aprovação da disciplina de Cálculo 1. Dessa maneira caberá a ele contornar suas limitações para alcançar o tão sonhado diploma de graduação.

2 METODOLOGIA UTILIZADA

O procedimento para alcançar os objetivos pretendidos neste trabalho foi composto de duas etapas fundamentais. Na primeira etapa realizou-se uma pesquisa de campo entre 281 alunos dos diversos cursos de engenharia do ITEC que já cursaram a disciplina de Cálculo 1. Para esses alunos foi solicitado que preenchessem de maneira sincera e verdadeira um questionário que continha apenas perguntas cujas respostas eram sim ou não. A Tabela 1 apresenta as 13 perguntas utilizadas na pesquisa de campo bem como a quantidade de respostas sim ou não para cada uma delas.

Tabela 1 – Pesquisa de campo entre alunos que já cursaram a disciplina de Cálculo 1.

| Numeração | Pergunta do questionário | Sim | Não |
|-----------|--|-----|-----|
| 1 | Você passou em Cálculo 1 na primeira vez que cursou a disciplina? | 224 | 57 |
| 2 | Você concluiu o ensino médio em escola pública? | 128 | 153 |
| 3 | Você gostava de estudar matemática no ensino médio? | 254 | 27 |
| 4 | Você tem facilidade no aprendizado de matemática? | 229 | 52 |
| 5 | Você concluiu o curso de nivelamento PCNA? | 169 | 113 |
| 6 | Você estudava Cálculo 1 em grupo nos turnos em que não tinha aula na UFPa? | 117 | 164 |
| 7 | Você resolvia exercícios de Cálculo 1 além daqueles resolvidos em sala? | 202 | 79 |

| | | | |
|----|--|-----|-----|
| 8 | Você tinha o hábito de revisar o conteúdo de Cálculo 1 no mesmo dia da aula? | 62 | 219 |
| 9 | Você participou ou participa de algum projeto de pesquisa ou extensão | 179 | 102 |
| 10 | Você usa transporte público para vir à UFPa? | 226 | 55 |
| 11 | Você mora com a família? | 235 | 46 |
| 12 | Você mora em Belém? | 232 | 49 |
| 13 | Você possui renda própria? | 49 | 232 |

Fonte: Elaborado pelos autores

As informações obtidas por meio do questionário foram cadastradas no software IBM SPSS. De posse dos dados registrados iniciou-se a segunda etapa do procedimento, a criação de um modelo matemático obtido via regressão logística. Como o questionário utilizado só possui perguntas cujas respostas são do tipo sim ou não, isto é, 1 ou 0, diz-se então que a regressão logística usa variáveis dicotômicas para a estimativa da função que modela os dados tabelados. As questões de 2 a 13 do questionário foram utilizadas como dados de entrada, formando o conjunto das variáveis independentes do problema. A questão de número 1 foi utilizada como dado de saída, a qual representa a variável dependente do problema. A ferramenta de regressão logística aponta qual variável ou variáveis influenciam mais na aprovação da disciplina de Cálculo 1. O software escolhido para a modelagem deste problema foi o IBM SPSS, o qual dispõe do método de regressão logística.

3 MODELAGEM POR REGRESSÃO LOGÍSTICA

Um modelo de regressão pode ser definido como uma equação matemática em que se expressa o relacionamento de variáveis. Nestes modelos, define-se uma variável dependente (Y), ou variável de saída, e procura-se verificar a influência de uma ou mais variáveis ditas variáveis independentes, causais ou explicativas (X's) sobre esta variável dependente. Na equação (1), vê-se um exemplo de um modelo de regressão linear (ZANINI, 2007).

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_i X_i + \varepsilon_i \quad (1)$$

Onde:

Y_i : representa a variável dependente;

β_i : são os coeficientes de regressão;

X_i : são as variáveis independentes;

ε_i : erro.

A regressão logística consiste em um tipo de regressão aplicável e preferida quando se tem uma variável dependente categórica dicotômica, ou seja, uma variável nominal ou não métrica que possui apenas dois grupos ou classificações como resultados possíveis como, por exemplo, alto ou baixo, homem e mulher, sim ou não, etc. Na regressão logística, a probabilidade de ocorrência de um evento pode ser estimada diretamente (ZANINI, 2007). No caso da variável dependente Y assumir apenas dois possíveis estados (1 ou 0) e haver um

conjunto de p variáveis independentes X_1, X_2, \dots, X_p , o modelo de regressão logística pode ser escrito da seguinte forma:

$$P(x) = \frac{e^{g(x)}}{1 + e^{g(x)}} \quad (2)$$

Onde:

$$g(x) = \beta_0 + \beta_1 X_{1i} + \dots + \beta_p X_{pi}$$

A seguir tem-se as etapas de análise dos dados para se obter o modelo de regressão logística apresentado na equação (2).

3.1 VERIFICAÇÃO DE COLINEARIDADE

Quando trabalha-se com mais de uma variável regressora, é muito importante verificar se essas variáveis explicativas são correlacionadas. Se não houver nenhum relacionamento entre elas, diz-se que são ortogonais. Porém se as variáveis forem muito correlacionadas, as inferências baseadas no modelo de regressão podem ser errôneas ou pouco confiáveis. Por isso, é necessário verificar se as variáveis são altamente correlacionadas (FERREIRA, 2012).

Por meio do software IBM SPSS, é possível verificar a colinearidade por meio do seguinte procedimento: *Analisar* → *Regressão* → *Linear*. Neste novo painel, é possível clicar em *Estatística* e marcar a opção *Diagnósticos de colinearidade*. Feito isso, na aba *Coefficientes* são verificados os índices de colinearidade, que são: Tolerância e VIF (VarianceInflationFactor).

O valor de VIF deve estar abaixo de 10 e o valor de Tolerância ($1/VIF$) deve estar acima de 0,1 (FERREIRA, 2012). A Figura 1 apresenta a janela de saída do software com 12 perguntas do questionário (perguntas de 2 a 13, que são as variáveis de entrada do modelo). Verificando os índices, percebe-se que os valores de VIF e Tolerância colaboram com o intervalo permitido, não registrando alta colinearidade entre as variáveis. Dessa forma, pode-se prosseguir com a regressão logística.

Figura 1 – Verificação de colinearidade

| Modelo | | Coeficientes não padronizados | | Coeficientes padronizados Beta | t | Sig. | Estatísticas de colinearidade | |
|--------|--|-------------------------------|------|--------------------------------|--------|------|-------------------------------|-------|
| | | B | Erro | | | | Tolerância | VIF |
| 1 | (Constante) | ,362 | ,128 | | 2,819 | ,005 | | |
| | Você concluiu o ensino médio em escola pública? | -,130 | ,048 | -,161 | -2,726 | ,007 | ,832 | 1,202 |
| | Você gostava de estudar matemática no ensino médio | ,327 | ,083 | ,240 | 3,928 | ,000 | ,783 | 1,278 |
| | Você tem facilidade no aprendizado da matemática? | ,005 | ,064 | ,005 | ,084 | ,933 | ,757 | 1,320 |
| | Você concluiu o curso de nivalamento PCNA? | ,088 | ,047 | ,107 | 1,871 | ,062 | ,888 | 1,127 |
| | Você estudava C1 em grupo nos turnos em que não tinha aula na UFPA? | -,063 | ,046 | -,077 | -1,366 | ,173 | ,918 | 1,089 |
| | Você resolvia exercícios de C1 além daqueles resolvidos em sala? | ,039 | ,053 | ,044 | ,739 | ,460 | ,833 | 1,200 |
| | Você tinha o hábito de revisar o conteúdo de C1 no mesmo dia da aula? | ,091 | ,057 | ,094 | 1,603 | ,110 | ,855 | 1,170 |
| | Você participou ou participa de algum projeto de pesquisa ou extensão? | ,115 | ,050 | ,138 | 2,324 | ,021 | ,825 | 1,212 |
| | Você usa transporte público para vir pra UFPA? | ,045 | ,057 | ,044 | ,788 | ,431 | ,926 | 1,080 |
| | Você mora com a família? | ,091 | ,061 | ,084 | 1,499 | ,135 | ,938 | 1,066 |
| | Você mora em Belém? | -,048 | ,062 | -,045 | -,763 | ,446 | ,845 | 1,183 |
| | Você possui renda própria? | ,030 | ,058 | ,029 | ,525 | ,600 | ,972 | 1,029 |

Fonte: Elaborado pelo autor

3.2 VALOR-P E OS TESTES DE WALD, OMNIBUS, HOSMER-LEMESHOW

Um importante dado a ser considerado é o valor-p de significância, que aparece como *Sig* no software. O valor-p é definido como a probabilidade de se observar um valor da estatística de teste (hipótese nula) maior ou igual ao encontrado. Tradicionalmente, o valor de corte para rejeitar a hipótese nula é de 0,05, o que significa que, quando não há nenhuma diferença, um valor tão extremo para a estatística de teste é esperado em menos de 5% das vezes. Valores pequenos de valor-p indicam, portanto, que a hipótese nula é falsa (FERREIRA, PATINO, 2015). O teste de Wald verifica o grau de significância de cada coeficiente da equação logística. É definido como a razão entre o coeficiente e o seu respectivo desvio padrão. O seu valor-p admite como hipótese nula que o coeficiente é nulo. Deve ser, portanto, recusado, ou seja, seu valor-p deve estar abaixo de 5%. Isso indica que, em menos de 5% das vezes o coeficiente é nulo. Logo, ele é então significativo para a equação (3).

$$W_j = \frac{B}{SE} \quad (3)$$

Onde:

Wj: Valor-p de Wald

B: Coeficiente da variável

SE: Desvio padrão da variável

O teste de Omnibus possui como hipótese nula que o ajuste do modelo seja igual ao ajuste do modelo sem previsor e como hipótese alternativa que o ajuste do modelo seja diferente do modelo sem previsor. É então recomendado que se escolha a hipótese alternativa (valor-p de Omnibus < 5%), pois significa que um modelo com um determinado previsor (variável) a mais é melhor que um modelo sem esse previsor. Caso não seja possível rejeitar a hipótese nula, deve-se aceitar o modelo sem aquele determinado previsor (GEYER, 2008).

O teste de Hosmer e Lemeshow admite como hipótese nula que o modelo gerado seja igual às categorias observadas. Considerando-se Y como o valor real da variável e Y^{\wedge} como o valor previsto, esse teste é feito com intuito de medir a proximidade de ambos. A hipótese nula (hipótese de teste) é que não existe diferença significativa entre o valor real e o valor previsto, ou seja, equivale a dizer que o modelo tem bom poder de ajuste. Quanto menor é o valor da diferença entre Y e Y^{\wedge} , mais os valores previstos se aproximam dos reais e, portanto, melhor desempenho preditivo tem o modelo. Desta forma, um fator positivo a favor do modelo é quando se aceita a seguinte hipótese nula:

$$H_0: Y = Y^{\wedge} \quad (4)$$

Dessa forma, o valor-p do teste de Hosmer-Lemeshow deve estar acima de 5%, o que afirma que o valor Y do modelo com a nova variável é igual ao valor Y^{\wedge} previsto em mais de 5% das vezes caso a hipótese nula seja verdadeira (ZANINI, 2007).

3.3 ANÁLISE DO VALOR-P DE WALD

No software, clicando em *Analisar* → *Regressão* → *Logística Binária*, pode-se definir os parâmetros a serem considerados pelo programa. Primeiro, clicou-se em *categórico* e todas as variáveis independentes foram inseridas como categóricas. Inicialmente, foi realizada uma única regressão logística com todas as variáveis, a fim de observar quais delas estavam em desacordo com as condições estabelecidas. Dessa forma, obteve-se a saída cuja imagem encontra-se na Figura 2.

Figura 2 – Regressão logística com todas as variáveis

| | | B | S.E. | Wald | df | Sig. |
|----------------------|---|--------|-------|-------|----|------|
| Passo 1 ^a | Você concluiu o ensino médio em escola pública?(1) | 1,052 | ,378 | 7,740 | 1 | ,005 |
| | Você gostava de estudar matemática no ensino médio(1) | 1,617 | ,553 | 8,533 | 1 | ,003 |
| | Você tem facilidade no aprendizado da matemática?(1) | -,096 | ,445 | ,046 | 1 | ,829 |
| | Você concluiu o curso de nivalamento PCNA?(1) | -,695 | ,357 | 3,787 | 1 | ,052 |
| | Você estudava C1 em grupo nos turnos em que não tinha aula na UFPA?(1) | ,612 | ,368 | 2,765 | 1 | ,096 |
| | Você resolvia exercícios de C1 além daqueles resolvidos em sala?(1) | ,297 | ,385 | ,595 | 1 | ,440 |
| | Você tinha o hábito de revisar o conteúdo de C1 no mesmo dia da aula?(1) | -1,278 | ,661 | 3,732 | 1 | ,053 |
| | Você participou ou participa de algum projeto de pesquisa ou extensão?(1) | 1,135 | ,459 | 6,112 | 1 | ,013 |
| | Você usa transporte público para vir pra UFPA?(1) | ,293 | ,445 | ,434 | 1 | ,510 |
| | Você mora com a família?(1) | -,718 | ,426 | 2,844 | 1 | ,092 |
| | Você mora em Belém?(1) | ,286 | ,509 | ,315 | 1 | ,574 |
| | Você possui renda própria?(1) | -,630 | ,497 | 1,606 | 1 | ,205 |
| | Constante | ,528 | 1,115 | ,224 | 1 | ,636 |

Fonte: Elaborado pelo autor

Na coluna *Sig*, pode-se verificar que o valor-p de Wald de muitas variáveis está acima de 5%. Isso indica que a inclusão daquela variável não faz uma diferença significativa no modelo construído. Do contrário, esta variável está poluindo o modelo. Foi optado então por fazer uma nova regressão retirando as variáveis categóricas que estão com índice acima de 5%. Na Figura 3 tem-se a janela de saída do software com apenas 5 entradas para o modelo. Quando realizou-se essa nova regressão, observou-se que as variáveis possuíam valor-p de Wald abaixo de 5%, sendo possível iniciar um novo critério de exclusão para o modelo.

Figura 3 – Regressão logística após a análise do Valor-p de Wald.

| | | B | S.E. | Wald | df | Sig. | Exp(B) | Inferior | Superior |
|----------------------|---|--------|------|--------|----|------|--------|----------|----------|
| Passo 1 ^a | Você concluiu o ensino médio em escola pública?(1) | ,962 | ,354 | 7,389 | 1 | ,007 | 2,618 | 1,308 | 5,240 |
| | Você gostava de estudar matemática no ensino médio(1) | 1,833 | ,481 | 14,496 | 1 | ,000 | 6,253 | 2,434 | 16,067 |
| | Você concluiu o curso de nivelamento PCNA?(1) | -,735 | ,337 | 4,746 | 1 | ,029 | ,480 | ,248 | ,929 |
| | Você tinha o hábito de revisar o conteúdo de C1 no mesmo dia da aula? (1) | -1,321 | ,633 | 4,356 | 1 | ,037 | ,267 | ,077 | ,923 |
| | Você participou ou participa de algum projeto de pesquisa ou extensão?(1) | ,997 | ,438 | 5,170 | 1 | ,023 | 2,709 | 1,147 | 6,398 |
| | Constante | ,588 | ,792 | ,551 | 1 | ,458 | 1,800 | | |

Fonte: Elaborado pelo autor

3.4 ANÁLISE DO TESTE DE OMNIBUS E DO TESTE DE HOSMER-LEMESHOW

Nesta nova etapa, utilizou-se o método de regressão logarítmica hierárquica para estas 5 variáveis de entrada remanescentes. Nesse método, o programa adiciona gradualmente as variáveis uma a uma, possibilitando comparar diferentes modelos em diferentes etapas. Cada vez que se adiciona uma nova variável, o programa adiciona um novo bloco, mostrando os testes de Omnibus e de Hosmer-Lemeshow após a inserção daquela variável. Quando foram inseridas as variáveis ‘Você concluiu o curso de nivelamento PCNA’ e ‘Você tinha o hábito de revisar o conteúdo no mesmo dia da aula’, respectivamente, as variáveis não passaram no teste de Hosmer-Lemeshow, ou seja, em menos de 5% das vezes o valor real é igual ao valor previsto caso a hipótese nula seja aceita. Isso indica que estas variáveis estão poluindo o modelo, e podem ser descartadas. Porém, é importante salientar que essas duas variáveis passaram no teste de Omnibus, ou seja, a inserção dessas variáveis altera de alguma forma o resultado. Mesmo assim, optou-se por retirá-las.

4. ANÁLISE E INTERPRETAÇÃO DO MODELO FINAL

Após a determinação das 3 entradas que mais afetam o modelo de regressão logística, foi feita uma nova regressão a qual aponta para um modelo julgado mais adequado. Os dados obtidos via software são mostrados na Figura 4.

Figura 4 – Variáveis na Equação

| | | B | S.E. | Wald | df | Sig. | Exp(B) | 95% C.I. para EXP(B) | |
|----------------------|---|--------|------|--------|----|------|--------|----------------------|----------|
| | | | | | | | | Inferior | Superior |
| Passo 1 ^a | Você concluiu o ensino médio em escola pública?(1) | -1,116 | ,380 | 8,616 | 1 | ,003 | ,327 | ,155 | ,690 |
| | Você gostava de estudar matemática no ensino médio(1) | 2,417 | ,528 | 20,987 | 1 | ,000 | 11,209 | 3,986 | 31,521 |
| | Você participou ou participa de algum projeto de pesquisa ou extensão?(1) | 2,289 | ,618 | 13,713 | 1 | ,000 | 9,861 | 2,937 | 33,112 |
| | Constante | -,412 | ,573 | ,517 | 1 | ,472 | ,662 | | |

Fonte: Elaborado pelo autor

A coluna B indica os coeficientes de cada variável. Podemos observar que o fato de o aluno ter concluído o ensino médio em escola pública contribui negativamente para a estatística. A coluna Exp(B) indica a escala de OddsRatio(razão de chances). A interpretação deste dado é que indivíduos que dizem 'sim' para aquela variável categórica, possuem chances Exp(B) vezes maior de estarem no grupo que passou em cálculo 1 de primeira.

Nossa base de dados, portanto, conclui que 'gostar de estudar matemática no ensino médio' e 'participar de algum projeto de pesquisa ou extensão' implica em chances maiores de passar em Cálculo 1 de primeira (aumento de 11,209 vezes e 9,861 vezes respectivamente). Em contrapartida, alunos que concluem o ensino médio em escola pública possuem menos chance de passar em Cálculo 1 de primeira (0,327 vezes a mais). Sendo assim, conforme a equação (2), o modelo de regressão logística final deste trabalho fica:

$$P(x) = \frac{e^{-1.116x_1+2.417x_2+2.289x_3}}{1 + e^{-1.116x_1+2.417x_2+2.289x_3}} \quad (5)$$

Onde:

x1 é a resposta (1 ou 0) para a pergunta 'Você concluiu o ensino médio em escola pública?'.

x2 é a resposta (1 ou 0) para a pergunta 'Você gostava de estudar matemática no ensino médio?'.

x3 é a resposta (1 ou 0) para a pergunta 'Você participou ou participa de algum projeto de pesquisa ou extensão?'.

5. CONSIDERAÇÕES FINAIS

O objetivo deste trabalho foi, além de gerar uma equação de maneira mais fiel e precisa possível, identificar estatisticamente quais foram os fatores que mais afetaram na reprovação em Cálculo 1 do grupo em estudo. Embora a regressão logística realizada não deva ser a única forma de análise, corresponde a uma poderosa ferramenta para que os órgãos e instituições competentes possam estudar e definir novas estratégias para evitar a reprovação e evasão de calouros da disciplina, haja visto que alguns fatores aqui identificados como cruciais podem não estar sendo levados com tanta atenção em consideração em detrimento de outros.

Agradecimentos

A equipe realizadora deste trabalho agradece aos alunos de engenharia do Instituto de Tecnologia da Universidade Federal do Pará campus Belém, que colaboraram com o preenchimento do questionário.

REFERÊNCIAS

CASTRO, Yuri. Porque reprova-se tanto em cálculo no ensino superior? Disponível em:< <https://extra.globo.com/noticias/educacao/vida-de-calouro/por-que-reprova-se-tanto-em-calculo-no-ensino-superior-saiba-como-evitar-13356046.html>>. Acesso em: 06/01/2019.

DEMARIS, Alfred. 'A **tutorial in logistic regression**'. Journal of Marriage and Family, Vol. 57, No 4. 1995.

FERREIRA, Juliana Carvalho, PATINO, Cecilia Maria. 'O que realmente significa valor-p?'. São Paulo, 2015.

FERREIRA, Ariane. '**Disciplina de modelos lineares – Regressão Logística**'. Rio de Janeiro, 2012.

FISHER, R. '**A. Statistical methods for research workers**', 12th edition. New York, Hafner, 1954.

GARZELA, Fabiana. '**A disciplina de Cálculo I: a análise das relações entre as práticas pedagógicas no professor e seus impactos nos alunos.**' 2013. 275 f. Tese (Doutorado em Educação), Universidade de Campinas, Campinas, 20/08/2013.

GEYER, Charles J., '**Fitting a Regression Model**'. Disponível em: <<http://www.stat.umn.edu/geyer/aster/short/examp/reg.html>>. Acesso em: 06/01/2019.

HOSMER, LEMESHOW, '**Applied Logistic Regression**', 2nd edition, 2000.

RODRIGUES, Nilton Rodolfo N. M. et. al. Alunos Ingressantes nas engenharias e a aprendizagem básica em matemática. In: XL Congresso Brasileiro de Educação em Engenharia - COBENGE, 2012, Belém, PA.

ZANINI, Alexandre. '**Regressão logística e redes neurais: um problema de estrutura de preferência do consumidor e classificação de perfis de consumo**'. TD. Mestrado em Economia Aplicada. FEA/UFJF. Juiz de Fora, 2007.

THE DEVELOPMENT OF A MATHEMATICAL MODEL FOR THE FORECAST ON APPROVAL IN CALCULUS 1 USING LOGISTIC REGRESSION

Abstract: *This research aims to identify which factors are significant in the performance of freshmen of the Federal University of Pará (UFPA) engineering courses in the discipline of Calculus 1, which has one of the highest failure rates in universities. Using the IBM SPSS Statistics software, a database of dichotomous variables can be created and logistic regression can be used to obtain a function that has as input the categorical variables and as an output the probability of passing through or not in Calculus 1. For that, we interviewed 281 students who have already studied the discipline, answering the 13 questions (categorical variables) each.*

Key-words: *Calculus 1. Categorical variable. Logistic regression.*