

**ESTUDO SOBRE A EVASÃO ESCOLAR USANDO REGRESSÃO LOGÍSTICA:  
ANÁLISE DOS ALUNOS DO CURSO DE ADMINISTRAÇÃO DA FUNDAÇÃO  
EDUCACIONAL DE ITUVERAVA**

**STUDY ON THE TRUANCY USING LOGISTIC REGRESSION: ANALYSIS OF  
THE STUDENTS THE MANAGEMENT EDUCATIONAL FOUNDATION OF  
ITUVERAVA**

**ESTUDIO SOBRE LA EVASIÓN ESCOLAR MEDIANTE REGRESIÓN  
LOGÍSTICA: ANÁLISIS DE LOS ESTUDIANTES DE ADMINISTRACIÓN DE LA  
FUNDACIÓN PARA LA EDUCACIÓN DE ITUVERAVA**

**GISLAINE CRISTINA BATISTELA<sup>1</sup>**

**SERGIO AUGUSTO RODRIGUES<sup>2</sup>**

**JÚLIA T. CARRER MARTINELLI BONONI<sup>3</sup>**

Recebido em setembro de 2009. Aprovado em setembro de 2009.

---

<sup>1</sup> Professora Assistente da Faculdade de Tecnologia de Botucatu e do Centro Universitário Central Paulista - São Carlos, SP. Licenciada em Matemática pela UNESP de São José do Rio Preto e Mestre em Estatística pela UFSCar. End.: Av. José Ítalo Bacchi, S/N, CEP: 18606 – 855. Tel. (14) 3814-3004, Botucatu-SP. E-mail: [gbatistela@fatecbt.edu.br](mailto:gbatistela@fatecbt.edu.br)

<sup>2</sup> Professor Assistente da Faculdade de Tecnologia de Botucatu e da ASSER de Porto Ferreira. Bacharel e Mestre em Estatística pela Universidade Federal de São Carlos. Pós-graduado em Administração com ênfase em Marketing pela Universidade São Francisco. End.: Av. José Ítalo Bacchi, S/N, CEP: 18606 – 855. Tel. (14) 3814-3004, Botucatu-SP. E-mail: [sergio@fatecbt.edu.br](mailto:sergio@fatecbt.edu.br)

<sup>3</sup> Licenciada em Matemática pela Fundação Educacional de Ituverava. E-mail: [juliathomazini@yahoo.com.br](mailto:juliathomazini@yahoo.com.br)

**ESTUDO SOBRE A EVASÃO  
ESCOLAR USANDO REGRESSÃO  
LOGÍSTICA: ANÁLISE DOS  
ALUNOS DO CURSO DE  
ADMINISTRAÇÃO DA FUNDAÇÃO  
EDUCACIONAL DE ITUVERAVA**

**RESUMO**

Este trabalho propõe um estudo sobre a relação entre as variáveis de uma pesquisa realizada com os alunos ingressantes do curso de Administração da Faculdade de Filosofia, Ciências e Letras de Ituverava (FFCL). Os dados foram obtidos através da aplicação de um questionário socioeconômico com os alunos ingressantes, com o objetivo de analisar as variáveis que levam à desistência no final do primeiro ano. Foram utilizadas as técnicas estatísticas de tabelas de frequências simples, tabelas de contingência, coeficiente de correlação, estatística qui-quadrado e regressão logística múltipla. Através do estudo do relacionamento das variáveis levantadas neste questionário, foi possível concluir que a chance de desistência de um aluno que não mora na cidade onde se localiza a faculdade é aproximadamente 2 vezes a chance de um aluno que mora na própria cidade. Foi possível verificar também que a chance de evasão é maior entre os alunos

com renda familiar maior e alunos cujo pai é empresário, profissional liberal ou funcionário público. Com base em informações deste tipo, é possível propor políticas para retenção destes alunos antes que ocorra a evasão escolar.

**PALAVRAS-CHAVE:** Regressão logística. Relação entre variáveis. Risco de desistência.

**STUDY ON THE TRUANCY USING  
LOGISTIC REGRESSION: ANALYSIS  
OF THE STUDENTS THE  
MANAGEMENT EDUCATIONAL  
FOUNDATION OF ITUVERAVA**

**ABSTRACT**

This paper proposes a study on the relationship between the variables of a survey with the students entering of the course of Management of the School of Philosophy, Sciences and Letters of Ituverava (FFCL). Data were collected through a socioeconomic questionnaire with commencing students in order to examine the variables that lead to withdrawal at the end of the first year. It was used statistical techniques of simple frequency tables, contingency tables, correlation coefficient, chi-square and logistic regression. By studying the relationship of the variables raised in this questionnaire, it was concluded that the chance of withdrawal of a student who does not live in the town where the college is located approximately two times the chance of a student who lives in the city. It was also verified that the chance of evasion is higher among students with family income and students whose father is a businessman, professional or public official. Based on such information, it is possible to propose policies for retention of these students before the occurrence of truancy.

**KEYWORDS:** Logistic regression. Relationship between variables. Risk of withdrawal.

**ESTUDIO SOBRE LA EVASIÓN ESCOLAR MEDIANTE REGRESIÓN  
LOGÍSTICA: ANÁLISIS DE LOS ESTUDIANTES DE ADMINISTRACIÓN DE LA  
FUNDACIÓN PARA LA EDUCACIÓN DE ITUVERAVA**

**RESUMEN**

Este artículo propone un estudio sobre la relación entre las variables de una encuesta de los estudiantes que ingresan al curso de Administración de la Facultad de Filosofía, Ciencias y Letras de Ituverava (FFCL). Los datos fueron recolectados a través de un cuestionario socio-económico a los estudiantes que ingresaron, con el objetivo de examinar las variables que conducen a la retirada al final del primer año. Hemos utilizado técnicas estadísticas de tablas de frecuencias simples, tablas de contingencia, coeficiente de correlación, chi-cuadrado y regresión logística. Mediante el estudio de la relación de las variables planteadas en este cuestionario, se concluyó que la posibilidad del retiro de un estudiante que no vive en la ciudad donde la universidad se encuentra, es aproximadamente a dos veces la probabilidad de que un estudiante que vive en la ciudad. También se verificó que la posibilidad de evasión es más alta entre los estudiantes con ingresos de la familia y los estudiantes cuyo padre es un empresario, persona profesional o funcionario. Con base en esa información, es posible proponer políticas para la conservación de estos estudiantes antes que ocurra la evasión escolar.

**PALABRAS CLAVE:** Regresión logística. Relación entre las variables. El riesgo de deserción.

## 1 INTRODUÇÃO

A regressão logística foi descoberta no século XIX para descrever o crescimento das populações e as reações químicas de autocatálise. A ideia básica do desenvolvimento logístico é simples e usada nos dias atuais para modelar o crescimento populacional e muitas outras técnicas, tais como descrever a chance de ocorrência de um determinado evento, ou seja, a chance de desistência de um aluno, a chance de um cliente de um banco vir a se tornar inadimplente, entre outras.

A regressão logística vem sendo bastante utilizada em várias áreas. Carpenter (2006) analisou os modelos de risco de crédito a empresas não financeiras. Kanso (2003) utiliza a regressão logística para determinar os fatores que influenciam a família estar acima ou abaixo da linha de pobreza. Mai (2006) realizou um estudo sobre o perfil do empreendedor versus a mortalidade das micro e pequenas empresas comerciais do município de Aracruz-ES, onde se aplicou o método de regressão logística para obter a chance do empreendedor que possui um determinado perfil vir a ter seu negócio falido. Já Pessoa, Silva e Duarte (1997) realizaram um estudo sobre “Análise Estatística de Dados de Pesquisa por Amostragem: Problemas no uso de Pacotes-Padrão”.

A regressão logística pode ser

considerada uma extensão da regressão linear, pois assim como na regressão linear, ela estuda relações entre variáveis, buscando as variáveis que podem influenciar de alguma forma em uma variável dependente, sendo que na regressão logística essa variável dependente deve ser categórica. Enquanto a regressão linear dá uma resposta em valor numérico, a regressão logística dá uma resposta em probabilidade de chances de ocorrer o fato que está sendo estudado.

Neste trabalho foi estudada a relação das variáveis levantadas em um questionário de perfil socioeconômico dos alunos da Faculdade de Filosofia, Ciências e Letras de Ituverava (FFCL), do curso de Administração em Gestão de Negócios, com uma variável binária denominada “*status*” representando a desistência ou não do aluno no final do primeiro ano. Desta forma, através de um modelo de probabilidade chamado de Modelo de Regressão Logística, serão identificadas as variáveis que apresentam uma grande relação com a variável “*status*”, com o objetivo de prever a probabilidade de um novo aluno desistir do curso no final do primeiro ano, dado algumas características.

## 2 METODOLOGIA

Segundo Pagano e Gauvreau (2004), na regressão linear a resposta “*y*” é

sempre contínua, mas existem muitas situações em que a resposta é dicotômica, isto é, assumem dois valores possíveis. Esses valores são 1 e 0, onde 1 representa um sucesso e 0 representa um fracasso. A média desses valores é, portanto, a proporção de uns, chamada de “p”, ou seja, a probabilidade de sucesso,  $p=P(\text{sucessos})$ . O modelo de regressão logística procura estimar a probabilidade “p” de sucesso explicado através de variáveis independentes.

Portanto, no estudo de regressão logística, a variável resposta ou variável dependente admite um valor categorizado, podendo assumir 2, 3 ou mais valores categorizados. Quando a variável resposta está disposta em categorias, existem três procedimentos distintos para manipular os dados: regressão logística binária, ordinária e nominal. A escolha depende do número de categorias e característica da variável resposta.

A regressão logística binária é utilizada quando a variável apresenta duas categorias e dois níveis de resposta, por exemplo: sim e não, desistente e não desistente do curso. Já a regressão logística ordinária é utilizada quando apresenta três ou mais categorias e existe ordenação natural dos níveis de resposta.

O princípio de Bernoulli determina que os possíveis resultados apresentem ou não certa característica, ou seja, no caso da

variável apresentar a característica, ela apresenta o valor igual a um, que representa sucesso e, no caso contrário, valor igual a zero que representa o fracasso (SOUZA, 2006). Assim, para o caso da regressão logística binária as variáveis explicativas são de qualquer valor, porém a variável resposta será do tipo de Bernoulli. O objetivo da análise de regressão logística é encontrar um modelo que tenha um bom ajuste para descrever este relacionamento entre variável resposta e variáveis explicativas.

## 2.1 Regressão Logística Simples

A regressão logística é um método estatístico que descreve as relações entre uma variável resposta e uma ou mais variáveis independentes. A variável resposta ( $y$ ) é dicotômica, ou seja, é uma variável que apresenta duas possibilidades de resposta (sucesso e fracasso), atribuindo o valor 1 para o acontecimento de interesse (sucesso) e o valor 0 para o acontecimento complementar (fracasso).

A probabilidade de sucesso do modelo logístico simples é dada por:

$\pi_i = \pi(x_i) = P(y = 1   x = x_i) = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}$	(1)
---	-----

e a probabilidade de fracasso:

$1 - \pi_i = 1 - \pi(x_i) = P(y = 0   x = x_i) = \frac{1}{1 + \exp(\beta_0 + \beta_1 x_i)}$	(2)
---	-----

onde  $\beta_0$  e  $\beta_1$  são parâmetros desconhecidos.

Segundo Toledo e Ovalle (1995), em qualquer problema de regressão, a quantidade a ser modelada é o valor médio da variável resposta, dado os valores das variáveis independentes. Esta quantidade é chamada de média condicional, denotada por  $E(y/x = x_i)$ , sendo que “y” é a variável resposta e  $x_i$ , os valores das variáveis independentes. Na regressão logística, devido à natureza da variável resposta  $0 \leq E(y/x = x_i) \leq 1$ , ou seja, a esperança condicional está variando entre 0 e 1.

Na regressão logística, o valor da variável resposta dado  $x_i$  é expresso por  $y_i = \pi_i + e_i$ , como a quantidade  $e_i$ , que pode assumir somente um de dois possíveis valores, isto é,  $e_i = 1 - \pi_i$  para  $y_i = 1$  ou  $e_i = -\pi_i$  para  $y_i = 0$ , segue que  $e_i$  tem distribuição com média zero e variância dada por  $\pi_i(1 - \pi_i)$ , de acordo com Hosmer *et al* (1989).

Segundo Moore (2006), a regressão logística modela a média “p” em termos de uma variável explicativa “x”. Tentar relacionar “p” e “x” como numa regressão

linear simples:  $p = \beta_0 + \beta_1 x$ . Infelizmente esse não é um bom modelo. Sempre que  $\beta_1 \neq 0$ , valores extremos de “x” fornecerão valores para  $\beta_0 + \beta_1 x$  que ficam fora do conjunto de possíveis valores de “p”,  $0 \leq p \leq 1$ . O modelo de regressão logística remove essa dificuldade, operando com o logaritmo natural da chance,  $p/(1-p)$ .

Utiliza-se o termo logaritmo (ou log) da chance para essa transformação. À medida que “p” se afasta de 0 em direção a 1, o logaritmo da chance assume todos os valores numéricos possíveis negativos e positivos. Modela-se o logaritmo da chance como função linear da variável explicativa, isto é, o modelo para a regressão logística simples:

$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x + e$	(3)
---	-----

onde “p” é uma proporção binomial e “x” é a variável explicativa. Os parâmetros do modelo logístico são  $\beta_0$  e  $\beta_1$ , os quais devem ser estimados pelo método de máxima verossimilhança.

## 2.2 Regressão Logística Múltipla

A regressão logística múltipla é uma generalização do modelo de regressão logística simples, isto é, na regressão logística simples, trabalha-se com uma

variável explicativa e na regressão logística múltipla com mais de uma variável explicativa.

Para modelar o logaritmo da chance como função de duas ou mais variáveis explicativas, ajusta-se um modelo da forma:

$$\ln \left[ \frac{\hat{p}}{1 - \hat{p}} \right] = \hat{\alpha} + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_q x_q \quad (4)$$

onde  $x_1, x_2, \dots, x_q$  são as variáveis explicativas e  $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_q$  os parâmetros do modelo logístico.

### 3 RESULTADOS

Os dados utilizados para análise foram coletados através de um questionário socioeconômico aplicado na ficha de inscrição do vestibular, sendo utilizados somente os dados dos alunos que se matricularam no curso de Administração em Gestão de Negócios. Para algumas variáveis, houve a necessidade de agrupar em categorias de respostas, pois o baixo número de respostas em algumas categorias e também o tipo de resposta dificultava algumas análises. As variáveis analisadas e codificadas podem ser verificadas no anexo A.

Através das variáveis obtidas pelo questionário socioeconômico aplicado no

vestibular, foi analisada a relação de cada variável qualitativa com a variável binária “*status*” através do teste usando a estatística qui-quadrado ( $\chi^2$ ). O p-valor, resultado deste teste, é uma medida que dá a probabilidade de cometer um erro ao rejeitar a hipótese de que as variáveis são independentes. Portanto, se esse valor é muito pequeno, isso traz evidências de que a hipótese de independência das variáveis pode ser rejeitada. Com um nível de significância de 10%, verifica-se pelo teste qui-quadrado que somente as variáveis “residência” (p=0,02), “ENEM” (p=0,06) e “ocupação do pai” (p=0,06) são estatisticamente associadas com a variável “*status*”. As variáveis “período” e “trabalha” obteve um p-valor de 0,11. Apesar do p-valor destas variáveis ser superior ao nível de significância de 10%, elas também são consideradas relacionadas com a variável “*status*”, pois, 0,11 está próximo do nível de significância de 10%. Já a variável “renda” apresentou um nível de significância de 0,18, o qual não é considerado estatisticamente significativo, ou seja, não está relacionada com o “*status*” do aluno. No entanto, por não ser um valor muito discrepante de 10%, esta variável será incluída na análise do modelo logístico para confirmação ou não desta conclusão inicial. As demais variáveis não foram significativas.

Após a análise bivariada realizada através do teste qui-quadrado para verificar a relação entre as variáveis levantadas na pesquisa com a variável binária “*status*” “*y=1*”, ou seja, o aluno desistiu do curso em até um ano e “*y=0*”, o aluno não desistiu do curso, o modelo de regressão logístico será aplicado somente com as variáveis que apresentaram alguma relação com a variável “*status*”, ou seja, aquelas variáveis que apresentaram um p-valor próximo do nível de significância de 10%. Desta forma, a partir do banco de dados de 83 alunos que iniciaram o curso de Administração, será utilizado como variáveis explicativas no modelo de regressão logístico as variáveis descritas na Tabela 1

Para a estimação do modelo logístico, utilizou-se o software SPSS versão 13. Foi considerado como variável dependente a variável “*status*” e as variáveis explicativas estão definidas na Tabela 1. Para a estimação do modelo, os casos sem dados, ou seja, os valores *missing* foram desconsiderados. Dos 83 alunos analisados, 6 (7,2% da amostra) foram excluídos desta análise pois estavam sem dados em alguma variável explicativa (*missing*), ou seja, deixaram de responder alguma questão.

O modelo de regressão logístico foi estimado utilizando o método *Forward Stepwise: Wald*. Neste método cada variável é incluída separadamente em cada passo do modelo e analisado seu desempenho em estimar o “*status*” de um aluno. O processo termina até as variáveis que melhor explicam conjuntamente a variável binária “*status*” forem selecionadas, chegando ao modelo final.

**Tabela 2:** Teste *Omnibus* do modelo

	Teste Qui-Quadrado	Grau de liberdade	P-valor
Passo 1	8,504	2	0,014
Passo 2	10,702	3	0,013
Passo 3	12,644	4	0,013

Pela Tabela 2, verifica-se que as variáveis foram selecionadas em 3 etapas e observa-se que, no passo 3, o modelo apresenta adequação estatisticamente significativa, com um nível de significâncias de 2%, rejeitando a hipótese de que as variáveis explicativas não explicam a variável dependente “*status*”.

**Tabela 1:** Variáveis explicativas do modelo e suas respectivas categorias de respostas.

Variável	Categoria de resposta
Residência ( <i>Código da Variável: Residência</i> )	1-Mora em Ituverava    2-não mora em Ituverava
Período do curso ( <i>Código da Variável: Período</i> )	1-diurno    2-noturno
Renda Familiar ( <i>Código da Variável: Renda</i> )	1-até 5 salários    2-mais que 5 salários
Enem ( <i>Código da Variável: Enem</i> )	1-sim    2-não
Trabalha ( <i>Código da Variável: Trabalha</i> )	1-sim    2-não
Ocupação do pai ( <i>Código da Variável: OcupaçãoPai</i> )	1 - empresário, prof. liberal, func. público 2- militar, comerciante, agropecuarista, autônomo 3 - professor, assalariado e sem ocupação e outra

Outro teste importante antes do ajuste do modelo é o teste de adequação do modelo logístico. De acordo com Hair (2005), um modelo é bem ajustado quanto menor for o valor de “ $-2 \log Likelihood$ ”. Verificou-se pelos resultados deste teste que a cada passo esse valor foi reduzindo indicando que conforme as variáveis foram sendo introduzidas no modelo, sua adequação foi melhorando, passando de 98,24 para 94,1 no passo 3. As outras medidas (*Cox & Snell e Nagelkerke*) também indicam a qualidade do ajuste geral do modelo. Segundo Hair (2005), estas medidas de adequação de ajuste comparam as probabilidades estimadas com as probabilidades observadas, sendo que os valores mais altos significam um melhor ajuste do modelo.

Para Hair (2005), a medida final de ajuste do modelo pode ser analisada através do teste de “*Hosmer and Lemeshow*”. Esse teste mede a correspondência entre os valores reais e os previstos da variável dependentes, ou seja, a variável “*status*”. Um bom ajuste de modelo é indicado por um

valor de qui-quadrado não significativo. Esse teste verifica a hipótese de que os valores previstos estão próximos dos valores observados de chance contra uma hipótese de que isso não acontece. O resultado deste teste pelo SPSS obteve um p-valor igual a  $p=0,507$ , indicando que não há indícios para rejeitar a hipótese inicial, ou seja, os valores previstos estão próximos dos valores observados de chance.

Para avaliar a capacidade do modelo logístico prever as chances de um aluno desistir ou não do curso, verifica-se na Tabela 3 a classificação dos casos previstos pelo modelo comparando com os dados reais da variável “*status*”.

Verifica-se que de uma forma geral, 70,1% dos alunos são classificados de forma correta, ou seja, para um aluno que desistiu, o valor previsto é desistência e para um aluno que não desistiu o valor previsto foi não desistência. O percentual de classificação correta para alunos desistentes é de 55,2% e para alunos não desistentes é de 79,2%. Portanto, o modelo estimado possui

uma capacidade melhor de classificar corretamente alunos que não irão desistir.

**Tabela 3:** Tabela de classificação

Valores observados da Variável Status			Valores Previstos		% Correto
			0-Não desistiu	1-Desistiu	
Passo 3	Status	0-Não desistiu	38	10	79,2
		1-Desistiu	13	16	55,2
Percentual de acerto					70,1

Quanto maior a capacidade de previsão correta, melhor é o modelo estimado.

A Tabela 4 mostra as variáveis que entraram no modelo, gerada pela ferramenta de regressão logística bem como os seus coeficientes e os resultados do teste de significância das variáveis independentes. Na coluna “B (coeficientes)”, são apresentados, no passo 3 onde todas as variáveis já foram selecionadas, os coeficientes do modelo para cada variável explicativa. Já na coluna Erro Padrão, é apresentada uma medida de variabilidade destes coeficientes. A estatística *Wald*, bem como o resultado do teste (colunas “*Est Wald*” e “*p-valor*”), é utilizada para testar a hipótese de cada coeficiente do modelo de regressão. Segundo Hair (2005), esse teste irá identificar o quanto à variável explicativa participa individualmente da explicação da variável dependente Status. A estatística de *Wald*, testa para cada coeficiente das variáveis do modelo a seguinte hipótese de que o coeficiente da variável explicativa é igual a

zero contra uma hipótese alternativa de que não são iguais a zero.

Observa-se, na Tabela 4, que o p-valor das categorias de respostas das variáveis “Residência” (mora fora), “Ocupação do Pai” (ocupação 1), “Ocupação do Pai” (1-ocupação 2) e o “Ocupação do Pai” (2 – ocupação3) são menores ou igual a 0,10, indicando que há evidências para rejeitar  $H_0$ , ou seja, os coeficientes não são iguais a zero. Para a categoria da variável “Renda Familiar” (mais de 5 salários), o p-valor é 0,166 e portanto o coeficiente pode ser considerado igual a zero, logo, poderia sair do modelo. No entanto, resolveu-se manter essa variável no modelo por ser uma variável importante para o perfil financeiro do aluno. A coluna “*Exp(B)*” é o exponencial dos coeficientes estimados de cada categoria das variáveis do modelo e indica a chance de desistência de um aluno que está na categoria de resposta da linha correspondente na tabela, comparado com a primeira categoria da variável, tomada como base.

**Tabela 4:** Variáveis explicativas que entraram no modelo

	B Coeficientes)	Erro padrão	Est. Wald	G.l.	P-valor	Exp(B)	Intevalor de confiança de 95,0% para EXP(B)	
							Limite Inferior	Limite Superior
<b>Passo3</b> Residencia(1)	,829	0,510	2,641	1	0,104	2,292	0,843	6,232
RendaFamiliar(1)	,705	0,509	1,920	1	0,166	2,024	0,747	5,487
OcupacaoPai			5,584	2	0,061			
OcupacaoPai(1)	-1,666	0,717	5,405	1	0,020	0,189	0,046	0,770
OcupacaoPai(2)	-1,175	0,634	3,436	1	0,064	0,309	0,089	1,070

As duas últimas colunas apresentam o intervalo de confiança destes coeficientes. Por exemplo, o valor da coluna “*Exp (B)*” da variável “Residência” no passo 3 é 2,292, indicando que a chance de desistência de uma alunos que mora fora de Ituverava (categoria 2 desta variável) é 2,292 a chance desistência de um aluno que mora em Ituverava.

Analisando a Tabela 4, a chance de um aluno desistir pode ser explicada pelas variáveis: “Residência”, “Renda familiar” e “Ocupação do Pai”. Utilizando somente essas variáveis como variáveis explicativas do modelo de regressão logístico, pode-se prever a chance de um novo aluno desistir do curso até o fim do primeiro ano.

Pelos resultados apresentados na Tabela 4, pode-se verificar que a chance de um aluno que mora fora de Ituverava desistir do curso é 2,29 vezes a chance de um aluno que mora em Ituverava. Percebe-se, também, que a chance de um aluno com renda superior a 5 salários mínimos é 2,02 vezes a chance de um aluno com renda inferior a 5 salários desistir. Outro ponto interessante é

que a chance de um aluno que a ocupação do pai está na categoria ocupação 2 é 18,9% da chance de uma aluno cujo pai está na categoria de Ocupação 1, ou seja, aluno cujo pai é empresário ou político ou profissional liberal ou funcionário público tem chance de desistir 5,29 (1/0,189) a chance de um aluno cujo pai é militar ou agro pecuarista ou assalariado ou autônomo.

#### 4 CONCLUSÃO

Através das análises dos dados de uma pesquisa realizada na Faculdade de Filosofia e Letras de Ituverava (FFCL), com os alunos pretendentes do curso de Administração em Gestão de Negócios, verificou-se através do modelo de regressão logística que as variáveis significativamente importantes para explicar a desistência de um aluno até o final do primeiro ano são: local da residência, renda familiar e ocupação do pai. A partir do modelo logístico, foi possível concluir que a chance de desistência de um aluno que não mora em Ituverava é

aproximadamente 2 vezes a chance de um aluno que mora em Ituverava. Pode-se dizer também que a chance de desistência de um aluno cujo pai é empresário ou político ou profissional liberal ou funcionário público é aproximadamente 5 vezes a chance de desistência de um aluno cujo pai é militar ou agro pecuarista ou assalariado ou autônomo. Com relação à renda familiar, o aluno com renda acima de 5 salários mínimos sua chance de desistência é aproximadamente 2 a chance do aluno com renda abaixo de 5 salários mínimos.

## REFERÊNCIAS

CARPENTER, E.M.L. **Um modelo de análise de risco de crédito de clientes em relações B2B.** Tese (Doutorado em Administração de Empresas) PUC Rio, Rio de Janeiro, 2006.

HAIR, J. *et al.* **Análise multivariada de dados**, 5. ed. Porto Alegre: Bookman, 2005.

HOSMER, D. W.; LEMESHOW, S. **Applied Logistic Regression.** New York: John Wiley, 1989.

KANSO, S. **Utilização da regressão logística para a classificação de famílias**

**quanto à condição de pobreza nas RMs do Rio de Janeiro e Recife nos anos de 1970, 1980 e 1991**, Rio de Janeiro, 2003.

MAI, A.F. **O Perfil do Empreendedor versus a Mortalidade das Micro e Pequenas Empresas Comerciais do Município de Aracruz / ES.** Dissertação (Mestrado em Ciências Contábeis) – Fundação Instituto Capixaba de Pesquisas em Contabilidade, Economia e Finanças, Vitória, 2006.

MOORE, D. *et al.* **A prática da Estatística Empresarial: como usar dados para tomar decisões**, Rio de Janeiro, LTC, 2006.

PAGANO, M.; GAUVREAU, K. **Princípios de Bioestatística.** 2. ed. São Paulo: Pioneira Thomson, 2004.

PESSOA, D. C.; SILVA, P.L.N; DUARTE, R.N **Análise Estatística de Dados de Pesquisas por Amostragem: Problemas no uso de Pacotes-Padrão.** Rio de Janeiro: **Revista Brasileira de Estatística**, vol.58, n. 210, ano 1997.

SOUZA, E.C. **Análise de Influência Local no Modelo de Regressão Logística.** Dissertação (Mestrado em Estatística e experimentação agrônômica) – ESALQ/USP, Piracicaba, 2006.

TOLEDO, L.G; OVALLE, I.I. **Estatística Básica.** São Paulo: Atlas, 1995.

## ANEXO A

**Tabela:** Variáveis levantadas e suas categorias de respostas.

Variável	Categoria de resposta
<i>Sexo</i>	1-Masculino e 2-Feminino
<i>Idade</i>	1-17 a 18 Anos, 2-19 a 20 Anos e 3-Mais de 20 Anos
<i>Estado civil</i>	1-Solteiro, 2-Outros: Casado/ Divorciado/ Amasiado/ Viúvo
<i>Residência (Código da Variável: Residência)</i>	1-Mora em Ituverava 2-não mora em Ituverava
<i>Escola que frequentou no Ensino Médio</i>	1-Pública, 2-Particular e Pública
<i>Curso médio concluído</i>	1-Técnico/ Supletivo/ Magistério e 2-Ensino Médio comum
<i>Período do curso (Código da Variável: Período)</i>	1-diurno 2-noturno
<i>Se fez cursinho</i>	1-Sim e 2-Não
<i>Tempo de cursinho</i>	1- De um semestre até mais de 2 anos e 2-Não fez cursinho
<i>Número de vestibulares prestados</i>	1-Um e 2-Dois ou mais
<i>Curso superior</i>	1-Sim e 2-Não
<i>Enem (Código da Variável: Enem)</i>	1-sim 2-não
<i>Motivo que levou para optar pela FFCL</i>	1-Oferecer o curso da minha escolha, 2-Próxima residência/pelos amigos/não conseguiu vaga em outras Faculdades e 3-Não tenho recurso para estudar em outro local
<i>Decisão quanto ao curso escolhido</i>	1-Decidido e 2-Indeciso
<i>Ocupação do pai (Código da Variável: OcupaçãoPai)</i>	1 - empresário, prof. liberal, func. público 2- militar, comerciante, agropecuarista, autônomo 3 - professor, assalariado e sem ocupação e outra
<i>ocupação do pai e mãe</i>	1-Empresário/Político/Profissional liberal e Funcionário público, 2-Militar/Agro pecuarista/ Comerciante/ Autônomo e 3- Professor/Assalariado/Sem ocupação no momento
<i>Renda Familiar (Código da Variável: Renda)</i>	1-até 5 salários 2-mais que 5 salários
<i>Trabalha (Código da Variável: Trabalha)</i>	1-sim 2-não
<i>Quantas horas trabalha</i>	1-De 20 a 30 horas semanais e 2-Até 44 horas semanais/ trabalha eventualmente
<i>Domínio de língua estrangeira</i>	1-Inglês/ Espanhol e 2-Outra/ nenhuma
<i>Se visita feiras de vestibular</i>	1-Sim, 2-Não
<i>Onde que viu a propaganda da FFCL</i>	1-Jornal/ TV/ Rádio /Revista e 2-Outros/ Internet/ Outdoor/ nenhuma
<i>Que jornal que lê</i>	1-Estado de São Paulo/Folha de São Paulo e 2-Outros/ Jornal do Brasil/ O Globo/ Jornal da cidade ou Região/ nenhum
<i>Que emissora que assiste</i>	1-Globo e 2-Outros/ SBT/ Bandeirantes/ Record/Rede TV/ Cultura/ TV Educativa/ Rede Mulher/ Rede Vida/Rede Família/ MTV/ TV por Assinatura/ Não assiste TV
<i>Horário que assiste TV</i>	1-Diurno (Manhã e Tarde) e 2-Noturno: Noite e Madrugada
<i>Horário que ouve rádio</i>	1-Todos os dias e 2-Às vezes