UNIVERSIDADE FEDERAL DO PARANÁ
SETOR DE CIÊNCIAS EXATAS
DEPARTAMENTO DE ESTATÍSTICA
CURSO DE ESTATÍSTICA

Beatriz Cristina de Lima
Cleverson Freitas de Paula

ESTUDO SOBRE RETORNO ESCOLAR DE ALUNOS EVADIDOS DE ESCOLAS DA REDE ESTADUAL DO PARANÁ

CURITIBA
2013

Beatriz Cristina de Lima

Cleverson Freitas de Paula

ESTUDO SOBRE RETORNO ESCOLAR DE ALUNOS EVADIDOS DE ESCOLAS DA REDE ESTADUAL DO PARANÁ

Trabalho de Conclusão de Curso apresentado à disciplina Laboratório de Estatística do Curso de Graduação em Estatística da Universidade Federal do Paraná, como exigência parcial para obtenção do grau de Bacharel em Estatística.

Orientadora: Profa. Dra. Suely Ruiz Giolo

CURITIBA

2013

AGRADECIMENTOS

A Deus, por ter nos concedido vida, oportunidade e capacidade para passar por mais esta etapa; sem ele isto não seria possível.

As nossas famílias, pelo apoio e paciência no decorrer desses anos de graduação, assim como em toda a nossa vida.

Aos nossos amigos, pelo companheirismo em bons e maus momentos.

A Vanessa Mazurek e ao Felipe Werner, pela compreensão de nossa ausência durante o período de elaboração deste trabalho.

A Professora Doutora Suely Ruiz Giolo, pela disponibilidade de seu tempo e compartilhamento de seu conhecimento acadêmico.

A Professora Doutora Nívea da Silva Matuda, pela disponibilidade em participar da banca deste trabalho.

LISTA DE TABELAS

LISTA DE FIGURAS

RESUMO

Estudo sobre Retorno Escolar de Alunos Evadidos de Escolas da Rede Estadual do Paraná

Este trabalho apresenta um estudo sobre o retorno dos alunos evadidos de escolas da rede estadual de ensino do Estado do Paraná. Os dados analisados são referentes ao segundo semestre letivo do ano de 2011, oriundos dos levantamentos do Programa Fica Comigo do Governo do Estado do Paraná. Para a análise dos dados foi, primeiramente, ajustado um modelo de regressão logística a fim de estudar a probabilidade de retorno do aluno evadido. Na segunda etapa da análise, técnicas no contexto de análise de sobrevivência foram utilizadas para estudar o tempo até o retorno do aluno. Para isso, fez-se uso do estimador de Kaplan-Meier, do teste *log-rank* e do modelo de riscos proporcionais de Cox. Ambos os modelos ajustados mostraram que as variáveis: turno das aulas, série frequentada pelo aluno e causas da evasão (aluno, social e outros) foram significativas tanto para estimar a probabilidade de retorno do aluno evadido quanto para explicar o tempo até este retorno.

Palavras-Chave: Análise de Sobrevivência. Estimador de Kaplan-Meier. Evasão Escolar. Modelo de Cox. Programa Fica Comigo. Regressão Logística.

SUMÁRIO

1 INTRODUÇÃO

Segundo o artigo 205 da Constituição Federal e o artigo 53 do Estatuto da Criança e do Adolescente, a educação é direito de todos, visando o desenvolvimento da pessoa, o preparo para o exercício da cidadania e a qualificação para o trabalho. Porém, este direito básico do cidadão muitas vezes é violado pela evasão escolar, ou seja, pelo fato do aluno parar de frequentar a escola. Isto ocorre por diversos fatores, como: problemas familiares; distância entre casa e escola; falta de material escolar; transporte ou, até mesmo, falta de infra-estrutura por parte da escola; dentre outros.

Segundo dados divulgados pelo Instituto Nacional de Estudos e Pesquisas Anísio Teixeira (INEP, 2004) na Sinopse Estatística da Educação Básica, o percentual brasileiro de abandono escolar no ano de 2004 no ensino fundamental em escolas estaduais do Brasil foi de 8% e, no ensino médio, chegou a 15%. No estado do Paraná os números foram 5% e 13%, respectivamente.

Diante deste cenário, observou-se a necessidade de combater a evasão escolar. Para isso, o Governo do Estado do Paraná criou em 2005 o Programa Fica Comigo, com o objetivo de auxiliar as escolas a enfrentarem a evasão. Para o Programa, a evasão do aluno é considerada após este se ausentar da escola por 5 dias consecutivos ou 7 dias alternados em um mês.

O instrumento de controle do Programa é a ficha “FICA” (Ficha de Comunicação do Aluno Ausente), que contém informações a serem preenchidas pela escola, tais como: as medidas que foram tomadas para o retorno do aluno ausente, possíveis causas da ausência, dados da escola e do aluno, informações sobre a inserção da família e aluno em outros programas governamentais, além de constar se o aluno retornou ou não à escola. A ficha possibilita o levantamento e a análise das informações sobre o assunto.

Neste trabalho, os dados analisados são referentes às fichas FICA preenchidas no segundo semestre letivo de 2011 (25 de julho a 16 de dezembro) por todas as

escolas da rede estadual de ensino do Estado do Paraná, considerando os ensinos fundamental e médio. A Secretaria de Estado da Educação divide o estado em 32 Núcleos Regionais de Educação (NRE); contudo, devido ao grande volume de informação, foi analisado, neste trabalho, apenas o NRE Curitiba, pois este é o mais representativo.

Para análise das informações disponíveis, foram ajustados modelos de regressão logística a fim de estimar a probabilidade de retorno do aluno à escola após a evasão. Nestes modelos, o retorno do aluno (sim ou não) foi utilizado como variável resposta. Também foram ajustados modelos no contexto de análise de sobrevivência com o objetivo de analisar o tempo (em dias) até o retorno do aluno.

2 REVISÃO DE LITERATURA

Por ser um problema que afeta tanto as escolas brasileiras quanto as de outros países, a evasão escolar é um tema que tem recebido significativa atenção e vem sendo alvo de estudos por vários pesquisadores e órgãos governamentais.

Tanto nos ensinos fundamental e médio, quanto no ensino superior, um dos objetivos desses estudos tem sido identificar os principais fatores que levam à evasão escolar a fim de auxiliar na busca de estratégias para a sua redução.

Nos ensinos fundamental e médio, Ferreira (2000), por exemplo, abordou um estudo de caráter teórico e informativo sobre causas da evasão escolar, além de apresentar maneiras de intervenção. Em seu trabalho, o autor utilizou a divisão a seguir para grupamento de causas de evasão: i) escola: não atrativa, autoritária, professores despreparados, insuficiente, ausência de motivação, etc.; ii) aluno: desinteressado, indisciplinado, com problema de saúde, gravidez, etc.; iii) pais ou responsáveis: não cumprimento do pátrio poder, desinteresse em relação ao destino dos filhos, etc; e iv) social: trabalho com incompatibilidade de horário para os estudos, agressão entre os alunos, violência em relação a gangues, etc.

Já o trabalho de Haddad *et al*. (2011) procurou contextualizar historicamente a educação no Estado do Paraná, que no passado era excludente devido à oferta de vagas não ser suficiente. O trabalho mostrou que este problema foi parcialmente solucionado, dado que uma parcela significativa da população continua sendo excluída, pois se matricula e abandona os estudos. Para estes levantamentos, também foram utilizados os dados do programa Fica Comigo do Estado do Paraná, focando no ensino fundamental.

Quanto à evasão no ensino superior, Batistela *et al*. (2009) utilizaram técnicas de regressão logística para analisar a evasão no primeiro ano do curso de Administração da Faculdade de Filosofia, Ciência e Letras de Ituverava (FFCL), São Paulo. Assim como no presente trabalho, os dados utilizados pelos autores também

foram obtidos por meio de um questionário com a intenção de identificar covariáveis relacionadas com a evasão. O modelo final ajustado pelos autores continha as seguintes covariáveis: residência (em Ituverava ou não), renda familiar e ocupação do pai. Um dos resultados obtidos a partir do modelo mostrou que os alunos que não residiam em Ituverava apresentaram chance de evasão 2,29 vezes a dos alunos que moravam na cidade.

Evasão dos alunos que ingressaram no Curso de Estatística da Universidade Federal do Paraná (UFPR) entre os anos de 1991 e 2011, também foi analisada por Martins e Rocha (2011). Um dos objetivos da análise foi avaliar o processo seletivo estendido (PSE), o qual foi implantado em 2006 em três cursos da UFPR com o intuito de reduzir a evasão. Para o estudo, os autores analisaram os alunos que evadiram e os que não evadiram do curso, separadamente, utilizando técnicas de análise de sobrevivência tais como: o estimador de Kaplan-Meier, o teste *log-rank* e modelos de regressão no contexto de dados de sobrevivência discretos. Com base nos resultados dessas análises, os autores concluíram haver indícios que sugerem que o PSE tem auxiliado na redução da evasão. Apontaram, também, uma tendência sistemática de diminuição no número de ingressos de alunos no curso.

Já Silva *et al*. (2012) analisaram o tempo até a evasão dos alunos do Curso de Estatística da Universidade Federal da Paraíba (UFPB). Segundo os autores, as covariáveis que apresentaram maior associação com a evasão dos alunos foram: naturalidade, formas de ingresso e raça. Para chegar a estas covariáveis, os autores fizeram uso do estimador de Kaplan-Meier, para uma análise descritiva dos dados; o método de Collett para a seleção das covariáveis e; da distribuição Log-Normal para a variável resposta do modelo proposto, além dos resíduos de Cox Snell para verificar o ajuste do modelo.

Visando o retorno dos alunos evadidos, alguns programas vêm sendo adotados nos ensinos fundamental e médio. No Paraná, por exemplo, o Governo do Estado criou em 2005 o Programa Fica Comigo. Desse modo, a fim de contribuir com este tema, o presente trabalho apresenta um estudo sobre o retorno escolar de alunos dos ensinos

fundamental e médio que se evadiram das escolas estaduais de Curitiba, PR, no segundo semestre letivo de 2011.

3 MATERIAL E MÉTODOS

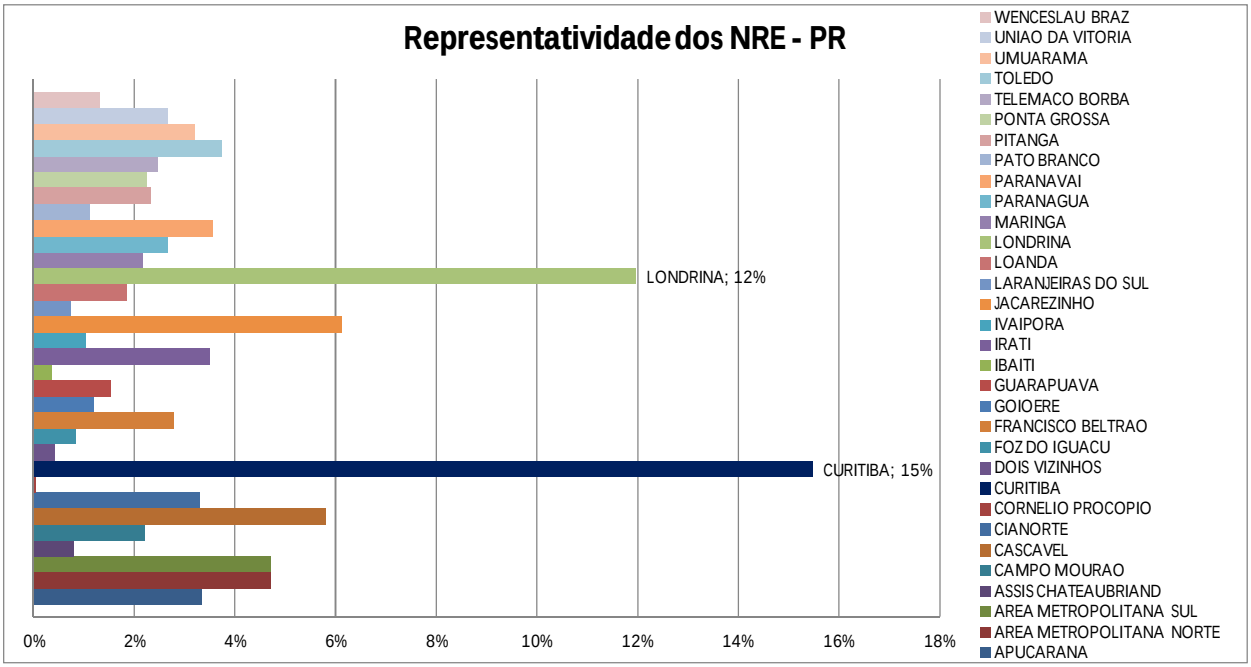
3.1 MATERIAL

A base de dados utilizada neste trabalho refere-se ao programa governamental intitulado Programa Fica Comigo e foi fornecida pela Secretaria de Educação do Estado do Paraná em Abril de 2013. Esta base contém informações de alunos evadidos das escolas estaduais do Paraná (ensino fundamental e médio, ensino de jovens e adultos e educação especial); todos os dados são referentes ao segundo semestre letivo do ano de 2011 (25 de julho a 16 de dezembro).

No geral, o banco de dados contém informações de 12.499 alunos que, de acordo com a divisão do Estado do Paraná em 32 Núcleos Regionais de Educação (NRE) proposta pela Secretaria de Estado da Educação (SEED), apresenta a distribuição mostrada na Figura 1. Observando os resultados nesta figura, nota-se que o NRE Curitiba é o mais representativo na base de dados de alunos evadidos no Estado do Paraná. Assim, e após reuniões conduzidas com a Coordenação de Gestão Escolar – CGE (Coordenação responsável pelo Programa Fica Comigo da SEED-PR), a decisão conjunta foi prosseguir os estudos estatísticos apenas com este núcleo. Um fator relevante para a escolha deste núcleo foi a orientação da CGE que o indicou como sendo homogêneo e menos suscetível às sazonalidades presentes em outros núcleos como, por exemplo, períodos de safras agrícolas.

Desta forma, a amostra para o presente estudo foi constituída com as fichas FICA do segundo semestre letivo do ano de 2011 dos alunos evadidos da rede estadual de ensino de Curitiba, capital paranaense, compreendendo um total de 1.933 alunos.

Figura 1 – Representatividade dos núcleos regionais de educação (NRE) do estado do Paraná



Fonte: Os autores (2013)

Para estes 1.933 alunos, as informações registradas nas fichas FICA estão divididas em seis grupos. São eles: 1) informações do estabelecimento; 2) informações do aluno; 3) formas de convocação; 4) inserção da família e/ou aluno em programas governamentais; 5) possíveis causas da ausência do aluno e; 6) resolução.

A partir desses grupos de informações foram definidas, neste trabalho, diversas covariáveis as quais se deseja avaliar suas respectivas associações com o retorno e o tempo até o retorno dos alunos evadidos. Tais covariáveis estão descritas na Tabela 1 e compreendem basicamente: a série frequentada pelo aluno no momento da evasão; o nível de ensino (fundamental ou médio); o turno das aulas (manhã, tarde etc.); a causa da evasão e; se o aluno e/ou a família estavam inseridos em algum programa social.

Neste trabalho, a definição das “causas de evasão” foi baseada na classificação apresentada por Ferreira (2000), adicionando a categoria “OUTROS”, que engloba as causas que não foram contempladas em nenhuma das quatro (escola, social, aluno e pais ou responsáveis) consideradas pelo autor. Como as causas atribuídas à evasão não eram mutuamente exclusivas, ou seja, era possível atribuir mais de uma causa

para a evasão do aluno, foram definidas variáveis dicotômicas para representar cada uma das cinco causas consideradas, como mostrado na Tabela 1.

Tabela 1 – Descrição das covariáveis definidas a partir das informações disponíveis no banco de dados dos 1.933 alunos evadidos do NRE Curitiba no segundo semestre letivo de 2011

COVARIÁVEL	DESCRIÇÃO
SERIE	Série frequentada pelo aluno evadido (1º ao 9º ano do ensino fundamental, 1º ao 4º ano do ensino médio)
TURNO	Turno das aulas do aluno evadido (manhã, tarde, noite, integral, intermediário)
NIVEL	Nível de ensino frequentado pelo aluno no momento da evasão (ensino fundamental, ensino médio)
CAUSA_ESCOLA	Causa da evasão do aluno foi atribuída à escola (sim, não)
CAUSA_SOCIAL	Evasão do aluno foi atribuída a causas sociais (sim, não)
CAUSA_PAIS	Causa da evasão do aluno foi atribuída aos pais ou responsáveis (sim, não)
CAUSA_ALUNO	Causa da evasão do aluno foi atribuída ao aluno (sim, não)
CAUSA_OUTROS	Causa da evasão do aluno foi atribuída a outros motivos (sim, não)
PROGRAMA_SOCIAL	Aluno e/ou família inseridos em algum programa social (sim, não)

Fonte: Os autores (2013)

Ressalta-se, ainda, que todas as informações contidas no banco de dados foram preenchidas pelas escolas de maneira manual, ou seja, há a possibilidade de erros de digitação que poderão causar possíveis vieses às análises; porém não há maneiras de antever ou prevenir estes equívocos.

Quanto à variável resposta considerada no modelo de regressão logística, que é dicotômica e corresponde ao retorno ou não do aluno evadido à escola, foi observado entre os 1.933 alunos, um total de 352 retornos (18,21%). Quanto ao tempo até o retorno do aluno, utilizado no modelo de sobrevivência, foi considerado o tempo (em dias) entre as datas de evasão e de retorno dos alunos. Para os que não retornaram, os tempos registrados foram aqueles entre a data da evasão e o último dia letivo do segundo semestre de 2011. Tais tempos em análise de sobrevivência são ditos tempos censurados.

3.2 MÉTODOS

A seguir são apresentados os métodos estatísticos utilizados neste trabalho para analisar os dados descritos na Seção 3.1.

3.2.1 Modelo de Regressão Logística

A regressão logística é uma técnica estatística que consiste em descrever a relação entre uma variável resposta Y e uma ou mais variáveis independentes (covariáveis). A variável resposta apresenta duas possibilidades, ou seja, é dicotômica, assumindo o valor 1 se o evento de interesse ocorre (sucesso) e 0, caso contrário (fracasso). No estudo analisado nesse trabalho a variável resposta foi definida como:

$$Y = \begin{cases} 1 \text{ se o aluno evadido retornou} \\ 0 \text{ se o aluno evadido não retornou.} \end{cases}$$

Assim, denotando por $\mathbf{x} = (x_1, ..., x_p)$ os valores das p (k = 1,..., p) covariáveis descritas na Tabela 1, a probabilidade de retorno do aluno evadido, de acordo com o modelo de regressão logística, ficou expressa por:

$$P(Y = 1 \mid \mathbf{X} = \mathbf{x}) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p)} \quad (1)$$

e a probabilidade de não retorno por:

$$P(Y = 0 \mid \mathbf{X} = \mathbf{x}) = \frac{1}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p)},$$

em que β_0 é uma constante e β_k (k = 1, ..., p) são os parâmetros ou coeficientes de regressão associados ao modelo.

O modelo apresentado em (1) em termos do logito, isto é, do logaritmo entre as probabilidades de retorno e não retorno apresenta uma forma linear nos parâmetros, sendo expresso da seguinte forma:

$$\ln\left(\frac{P(Y=1|\mathbf{X}=\mathbf{x})}{P(Y=0|\mathbf{X}=\mathbf{x})}\right)=\beta_0+\beta_1x_1+\beta_2x_2+...+\beta_px_p.$$

Para a estimação dos parâmetros do modelo de regressão logística em (1) foi utilizado o método de máxima verossimilhança, com o auxílio do *software* R, de modo que as estimativas obtidas para os parâmetros β_k (k = 1, ..., p) correspondem aos valores que maximizam a função de verossimilhança $L(\boldsymbol{\beta})$, expressa para uma amostra de n indivíduos independentes (l = 1,..., n) por:

$$L(\boldsymbol{\beta})=\prod_{l=1}^{n}(P(Y=1|\mathbf{X}=\mathbf{x}_l))^{y_l}(P(Y=0|\mathbf{X}=\mathbf{x}_l))^{1-y_l}\ .$$

Após a estimação dos parâmetros, foi utilizado o método *stepwise* para a seleção das covariáveis. Uma covariável é selecionada, segundo este método, se produz uma mudança significativa no logaritmo da verossimilhança associada ao modelo que não contém a mesma. O método basicamente consiste em testar a inclusão e a retirada de covariáveis levando-se em consideração dois níveis de significância (um para inclusão e outro para retirada). Para este trabalho, o nível de significância tanto para inclusão quanto para exclusão das covariáveis foi de 10%.

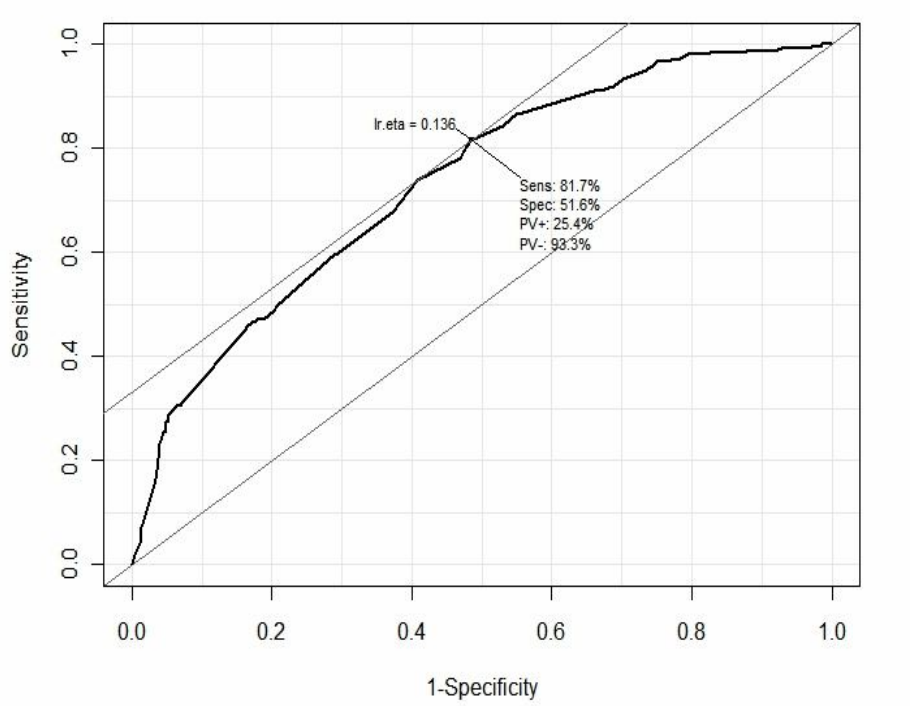
O método utilizado para a comparação dos modelos foi o Critério de Informação de Akaike (AIC), definido por:

$$AIC=2l+2p$$

em que $\mathbf{l}$ é o logaritmo neperiano do máximo da função de verossimilhança e p é o número de parâmetros do modelo considerado. De acordo com esse critério, o melhor modelo será o que apresentar o menor valor de AIC.

Para comparação dos modelos ajustados foi utilizada a curva ROC (*Receiver Operating Characteristic*). Esta é uma ferramenta gráfica que pode ser utilizada para avaliação da qualidade dos modelos (MARTINEZ *et al.*, 2003). O melhor modelo será aquele que apresentar a maior área abaixo da curva, além das melhores sensibilidade e especificidade, sendo sensibilidade a taxa de verdadeiros positivos e especificidade a taxa de verdadeiros negativos. Neste trabalho, a sensibilidade refere-se à taxa de alunos que o modelo indicou que retornariam à escola e que de fato retornaram. Já a especificidade é uma taxa semelhante à sensibilidade, porém referente aos alunos que não retornaram. A Figura 2 apresenta um exemplo de curva ROC, com sensibilidade de 81,7% e especificidade de 51,6%.

Figura 2 – Exemplo de Curva ROC (*Receiver Operating Characteristic*)



Fonte: Os autores (2013)

Por fim, para interpretar os parâmetros que compõem o modelo selecionado, foi analisado o exponencial de cada parâmetro estimado, que é interpretado como a chance de retorno. Assim se, por exemplo, o exponencial do parâmetro associado ao

turno tarde for igual a 2, significa que a chance de um aluno deste turno retornar é duas vezes a de um aluno do turno definido como o de referência para esta covariável.

3.2.2 Métodos e Modelos para Dados de Sobrevivência

Em análise de sobrevivência a variável resposta é o tempo decorrido até o evento de interesse do estudo; neste trabalho a variável resposta é o tempo até o retorno do aluno evadido. Este tempo é denominado tempo de falha; porém a principal característica em análise de sobrevivência é a presença de dados censurados, ou seja, dados incompletos devido à interrupção no acompanhamento de alguns indivíduos no estudo. A esses é dado o nome de tempos de censura.

Para uma amostra de *n* indivíduos (*i* = 1, ...*n*), dados de sobrevivência são usualmente representados pelos tempos de falha ou censura (t_i), pelo indicador de falha ou censura (δ_i) e, na presença de *p* covariáveis, pelos valores das mesmas, isto é, $\mathbf{x_i} = (x_{1i}, ..., x_{pi})'$.

As funções de maior interesse em estudos com dados de sobrevivência são:

i) a função de sobrevivência ($S(t \mid x)$): definida como a probabilidade de uma observação não falhar até certo tempo t, ou seja, a probabilidade de uma observação sobreviver ao tempo t (COLOSIMO; GIOLO, 2006) e;

ii) a função de taxa de falha ($\lambda(t \mid x)$): definida como a probabilidade condicional do indivíduo falhar em um intervalo de tempo $[t, t + \Delta t)$. Assumindo Δt bem pequeno, esta função representa a taxa de falha instantânea no tempo t condicional à sobrevivência até o tempo t.

Para realizar uma analise descritiva dos dados foi utilizado, inicialmente neste trabalho, o estimador não-paramétrico de Kaplan-Meier, também conhecido por estimador produto-limite (COLOSIMO; GIOLO, 2006). Sua expressão é dada por:

$$\hat{S}(t) = \prod_{j:t_j<t} \left(\frac{n_j - d_j}{n_j} \right) = \prod_{j:t_j<t} \left(1 - \frac{d_j}{n_j} \right),$$

sendo j a quantidade de falhas variando de 1 até *k,* t_j os *k* tempos distintos e ordenados de falha, d_j o número de falhas no tempo t_j e n_j o número de indivíduos sob risco em t_j, ou seja, os indivíduos que não falharam ou censuraram até o instante anterior a t_j.

Após estudos descritivos dos dados foi utilizado o teste *log-rank* para comparar curvas de sobrevivência entre categorias de uma mesma covariável, com o objetivo de encontrar as candidatas ao modelo final de sobrevivência. A hipótese nula deste teste estabelece a igualdade entre as curvas de sobrevivência, no qual a estatística de teste para a comparação de duas curvas é expressa por:

$$T = \frac{\left[\sum_{j=1}^{k} (d_{2j} - w_{2j}) \right]^2}{\sum_{j=1}^{k} (V_j)_2},$$

sendo *k* a quantidade de tempos de falha, d_{2j} o número de falhas observado no grupo 2 em t_j ($j = 1, ..., k$) e, w_{2j} e $(V_j)_2$ o valor esperado e a variância de d_{2j} sob H_0, respectivamente. A estatística T segue distribuição qui-quadrado com 1 grau de liberdade para grandes amostras.

Concluída a etapa descritiva dos estudos e conhecendo as covariáveis mais significativas conforme o teste *log-rank*, foi ajustado o modelo de regressão de Cox. Este também é denominado modelo de riscos proporcionais devido à suposição básica de taxas de falha proporcionais, sendo representado por:

$$\lambda(t) = \lambda_0(t) g(\mathbf{x}'\boldsymbol{\beta}),$$

sendo $\lambda_0(t)$ a função de taxa de falha de base, com $\lambda_0(t)$ e g(.) funções não-negativas em que g(.) é especificada de tal forma que $g(0) = 1$. Usualmente g(.) é expressa por:

$$g(\mathbf{x}'\boldsymbol{\beta}) = \exp\{\mathbf{x}'\boldsymbol{\beta}\} = \exp\{\beta_1 x_1 + ... + \beta_p x_p\},$$

sendo $\boldsymbol{\beta}$ o vetor de parâmetros associados às covariáveis.

O modelo de regressão de Cox é caracterizado pelo vetor de coeficientes ***β***, que medem os efeitos das covariáveis sobre a função de taxa de falha (COLOSIMO; GIOLO, 2006). Um método frequentemente utilizado para estimação destes coeficientes é o de máxima verossimilhança; porém com uma função de verossimilhança adequada para este fim devido ao componente não-paramétrico, a qual é denominada função de verossimilhança parcial e expressa por:

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left(\frac{\exp\{\mathbf{x}'_i \boldsymbol{\beta}\}}{\sum_{j \in R(t_i)} \exp\{\mathbf{x}'_j \boldsymbol{\beta}\}} \right)^{\delta_i}.$$

Juntamente ao ajuste dos modelos, foi utilizado o teste de Wald com a finalidade de testar a significância das covariáveis. As hipóteses associadas a este teste são:

$$H_0 : \beta_m = 0$$
$$H_1 : \beta_m \neq 0$$

para *m* = 1,..., *p*, sendo a estatística de teste dada por:

$$W = \frac{\hat{\beta}_m}{\sigma(\hat{\beta}_m)}$$

no qual $\sigma(\hat{\beta}_m)$ denota o desvio padrão de $\hat{\beta}_m$.

Caso a hipótese nula H_0 seja rejeitada, a covariável é dita apresentar efeito significativo, contribuindo, assim, para a explicação da variável resposta.

Para analisar a suposição de riscos proporcionais do modelo ajustado foram utilizados os resíduos de Schoenfeld (1982); estes são obtidos apenas para as falhas e são calculados da seguinte forma:

$$r_{iq} = x_{iq} - \frac{\sum_{j \in R(t_i)} x_{jq} \exp\{\mathbf{x'}_j \hat{\boldsymbol{\beta}}\}}{\sum_{j \in R(t_i)} \exp\{\mathbf{x'}_j \hat{\boldsymbol{\beta}}\}},$$

sendo x_j o vetor com os valores das covariáveis do modelo.

Caso a suposição de riscos proporcionais seja atendida, o gráfico dos resíduos padronizados de Schoenfeld $+\hat{\beta}_m$ (m = 1,..., p) versus t terá a forma de uma linha horizontal.

Os resíduos de Cox-Snell foram também utilizados com o propósito de testar a qualidade de ajuste do modelo de regressão de Cox. Estes são definidos por:

$$\hat{e}_i = \hat{\Lambda}_0(t_i) \exp\left\{\sum_{k=1}^{p} \mathbf{x}_{ip} \hat{\boldsymbol{\beta}}_k\right\} \quad \text{i = 1, ... , n.}$$

Caso o modelo esteja bem ajustado, estes resíduos devem seguir distribuição exponencial padrão.

Para interpretação dos parâmetros deve-se observar o exponencial do coeficiente estimado; este representa a razão de taxas de falha entre as categorias de cada covariável.

4 RESULTADOS E DISCUSSÕES

Conforme foi mencionado na Seção 3.1, as análises foram conduzidas considerando apenas o NRE Curitiba devido a sua representatividade. Entretanto, o percentual de retorno de alunos evadidos foi de 18%, valor que para algumas análises estatísticas é considerado baixo e pode dificultar bons ajustes. A seguir, são apresentadas informações descritivas da base de dados final para o estudo; deixando a ressalva de que para análises específicas foram realizados alguns tratamentos na base, os quais são expostos nos resultados das respectivas análises.

Analisando a Tabela 2 é possível observar alguns aspectos descritivos da base de dados do presente estudo. Quanto à distribuição da covariável nível de ensino, pode ser observado que os níveis com maiores frequências são: Ensino Fundamental (EF) e Ensino Médio (EM), sendo estes 98% da base de dados; o maior volume de alunos evadidos está concentrado no fim do EF e início do EM. Para a covariável série frequentada pelo aluno, nota-se, assim como na covariável nível de ensino, algumas classes sem representatividade (frequências pequenas). Já, observando a covariável turno, tem-se as maiores representatividades nos turnos: manhã, tarde e noite.

Na Tabela 3 são apresentadas as variáveis dicotômicas definidas a partir da base de dados, na qual “sim” denota a presença da covariável para o aluno evadido e “não” a ausência da mesma. A covariável “Programa Social”, definida a partir de 9 informações presentes na base de dados, indica se o aluno ou a família estavam inseridos em algum programa social. Para esta covariável, pode-se observar que mesmo quando o aluno estava inserido em algum programa, o retorno permaneceu baixo (23,38%). Quanto às covariáveis “Causas da Evasão”, pode-se observar que a causa com maior frequência foi àquela atribuída ao próprio aluno (1390 dentre os 1933), com correspondente percentual de retorno de 15,04%.

Também é possível observar a partir da Tabela 3, que os menores percentuais de retorno ocorreram entre os alunos em que as causas da evasão foram atribuídas à escola ou a outros motivos (0% e 7,11%, respectivamente). Em contrapartida, o maior

percentual de retorno ocorreu entre os alunos em que a causa foi atribuída aos pais ou responsáveis (38,59%).

Tabela 2 – Análise descritiva das variáveis categóricas definidas a partir das informações disponíveis no banco de dados dos 1.933 alunos evadidos do NRE Curitiba no segundo semestre letivo de 2011

COVARIÁVEL	CLASSE	TOTAL	RETORNO	%RETORNO
NIVEL	ENSINO FUNDAMENTAL	1.374	263	19,14%
	ENSINO MÉDIO	521	79	15,16%
	EJA	36	10	27,78%
	EDUCAÇÃO ESPECIAL	2	0	0,00%
SERIE	1º ANO ENSINO FUNDAMENTAL	1	0	0,00%
	4º ANO ENSINO FUNDAMENTAL	1	1	100,00%
	5º ANO ENSINO FUNDAMENTAL	16	12	75,00%
	6º ANO ENSINO FUNDAMENTAL	312	65	20,83%
	7º ANO ENSINO FUNDAMENTAL	290	45	15,52%
	8º ANO ENSINO FUNDAMENTAL	386	51	13,21%
	9º ANO ENSINO FUNDAMENTAL	360	81	22,50%
	1º ANO ENSINO MÉDIO	285	41	14,39%
	2º ANO ENSINO MÉDIO	148	19	12,84%
	3º ANO ENSINO MÉDIO	74	13	17,57%
	4º ANO ENSINO MÉDIO	1	0	0,00%
	EDUCAÇÃO ESPECIAL	2	0	0,00%
	EJA ANOS FINAIS	36	10	27,78%
	SEM INFORMAÇÃO	21	14	66,67%
TURNO	MANHÃ	979	176	17,98%
	TARDE	686	153	22,30%
	NOITE	216	12	5,56%
	INTERMEDIARIO	34	2	5,88%
	INTEGRAL	14	5	35,71%
	SEM INFORMAÇÃO	4	4	100,00%

Fonte: Os autores (2013)

Tabela 3 – Análise descritiva das variáveis dicotômicas definidas a partir das informações disponíveis no banco de dados dos 1.933 alunos evadidos do NRE Curitiba no segundo semestre letivo de 2011

COVARIÁVEL	CLASSE	TOTAL	RETORNO	%RETORNO
CAUSA_ESCOLA	NÃO	1.916	352	18,37%
	SIM	17	0	0,00%
CAUSA_SOCIAL	NÃO	1.795	329	18,32%
	SIM	138	23	16,67%
CAUSA_PAIS	NÃO	1.749	281	16,07%
	SIM	184	71	38,59%
CAUSA_ALUNO	NÃO	543	143	26,34%
	SIM	1.390	209	15,04%
CAUSA_OUTROS	NÃO	1.525	323	21,18%
	SIM	408	29	7,11%
PROGRAMA_SOCIAL	NÃO	1.732	305	17,61%
	SIM	201	47	23,38%

Fonte: Os autores (2013)

4.1 RESULTADOS DO MODELO DE REGRESSÃO LOGÍSTICA

Para iniciar os estudos de regressão logística foram realizados alguns tratamentos na base de dados (basicamente em razão de frequências pequenas observadas em certas categorias) a fim de evitar vieses nas análises. Na covariável série freqüentada foram utilizadas as observações da sexta série do ensino fundamental até a terceira série do ensino médio, excluindo-se as demais classes. Devido ao mesmo motivo, o turno intermediário (MT) da covariável turno das aulas foi agrupado com o turno integral (I) e, para a covariável nível de ensino, foram utilizados apenas os ensinos fundamental e médio.

Após esses tratamentos, a base de dados ficou com registros de 1.851 alunos.

Com o auxilio do *software* R (R *Development Core Team*, 2013) foram ajustados modelos de regressão logística múltipla, modelando a variável resposta retorno do aluno evadido. Utilizando o procedimento *stepwise* chegou-se ao melhor modelo com AIC de 1.503, sendo todas as variáveis deste modelo significativas a pelo menos 10% de significância. As estimativas dos parâmetros do modelo ajustado estão apresentadas na Tabela 4.

Tabela 4 – Estimativas associadas ao modelo de regressão logística ajustado aos dados de 1.851 alunos evadidos do NRE Curitiba no segundo semestre letivo de 2011

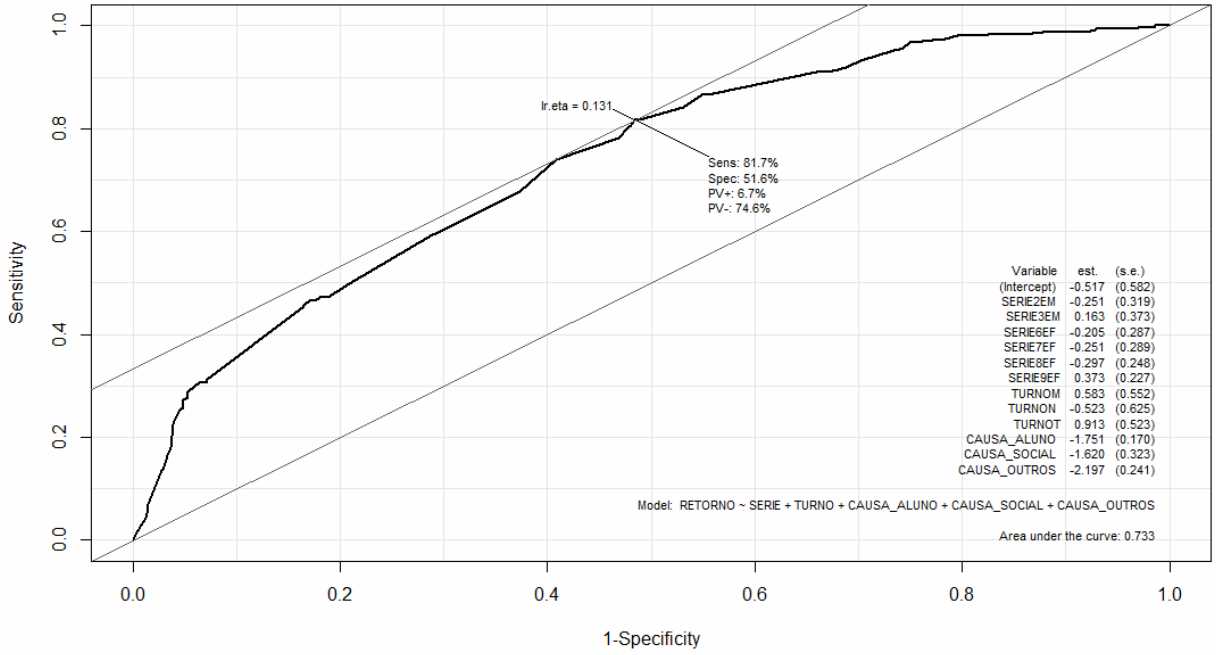
	ESTIMATIVA	**ERRO PADRÃO**	**EXP(ESTIMATIVA)**	**W**	**Pr (>|Z|)**
(Intercepto)	-0,5173	0,5823	--	0,7885	0,3743
SERIE2EM	-0,2507	0,3192	0,78	0,6177	0,4322
SERIE3EM	0,1629	0,3733	1,18	0,1900	0,6626
SERIE6EF	-0,2052	0,2867	0,81	0,5126	0,4742
SERIE7EF	-0,2511	0,2893	0,78	0,7534	0,3855
SERIE8EF	-0,2966	0,2477	0,74	1,4328	0,2311
SERIE9EF	0,3730	0,2265	1,45	2,7093	0,0997
TURNOM	0,5835	0,5517	1,79	1,1193	0,2902
TURNON	-0,5229	0,6251	0,59	0,7005	0,4029
TURNOT	0,9122	0,5226	2,49	3,0520	0,0807
CAUSA_ALUNO	-1,7512	0,1704	0,17	105,5551	<0,0001
CAUSA_SOCIAL	-1,6202	0,3230	0,20	25,1502	<0,0001
CAUSA_OUTROS	-2,1967	0,2406	0,11	83,3386	<0,0001

Fonte: Os autores (2013)

Dado que as covariáveis série e turno das aulas são categóricas (politômicas), foi estabelecido para cada uma delas um nível de referência. Para o turno, o nível de referência foi o integral/intermediário e, para a série, o primeiro ano do ensino médio.

A fim de verificar se o modelo ajustado explica de forma desejável a variável resposta retorno, foi construída a curva ROC apresentada na Figura 3. De acordo com essa figura, há evidências de que o modelo apresenta ajuste satisfatório, pois a área sob a curva foi de 0,733, bem como sensibilidade e especificidade associadas ao ponto de corte 0,131 foram 81,7% e 51,6%, respectivamente. Em termos das probabilidades preditas pelo modelo, tal ponto de corte corresponde ao valor que proporcionou os melhores percentuais de acerto quanto ao retorno ou não dos alunos evadidos.

Figura 3 – Curva ROC associada ao modelo de regressão logística selecionado



Fonte: Os autores (2013)

Após o ajuste do modelo foi possível obter algumas conclusões analisando o modelo final selecionado. O parágrafo a seguir apresenta tais conclusões, as quais têm como objetivo trazer algum auxílio a análises futuras de pesquisadores da área de educação.

Se o aluno apresentar qualquer uma das causas de evasão (aluno, social ou outros), a chance dele retornar é menor que a de um aluno ao qual não tenha sido atribuída nenhuma causa específica para a sua evasão. Por exemplo, um aluno

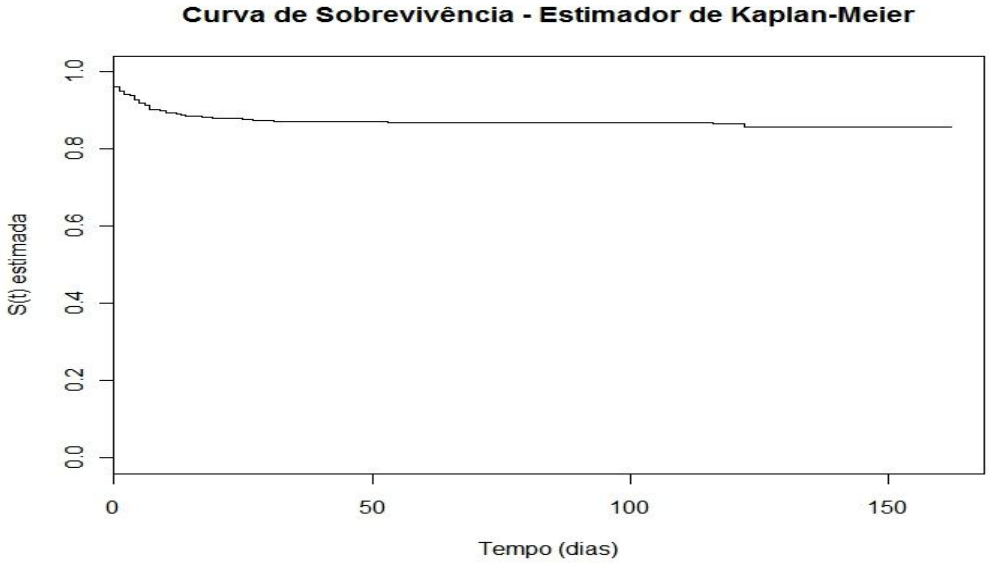
evadido devido à causas sociais, apresenta chance de retorno 5 vezes menor que a de um aluno sem nenhuma causa aparente para sua evasão. Já para um aluno evadido devido a outras causas, tem-se tal chance de retorno 9 vezes menor. Em relação à série frequentada, a chance de retorno aumenta se o aluno estiver nas séries finais de cada nível. Por exemplo, comparado aos alunos no 1º ano do EM (nível de referência), a chance de retorno dos alunos na nona série do ensino fundamental é 1,45 vez maior e, para os no terceiro ano do ensino médio, 1,18 vez maior. Quanto à covariável turno das aulas, a maior chance de retorno foi observada para os alunos do turno tarde; 2,49 vezes em relação aos alunos do turno integral/intermediário.

4.2 RESULTADOS DO MODELO DE ANÁLISE DE SOBREVIVÊNCIA

Para chegar aos resultados da análise de sobrevivência que se seguem, foram realizados os mesmos tratamentos mencionados em regressão logística. Além destes, foi também observado que a data de retorno não havia sido registrada para alguns alunos que retornaram; logo estes alunos foram excluídos dessa análise por não ser possível obter para os mesmos o tempo até o retorno, que é a variável resposta. Com isso, os estudos foram realizados com 1.791 alunos, sendo 13,06% o percentual de retorno.

Na análise de sobrevivência, devido à presença de dados censurados, a estatística descritiva dos tempos até o retorno foi feita por meio do estimador de Kaplan-Meier. Com este estimador, nota-se a partir da Figura 4 que a curva de sobrevivência estimada (na ausência de covariáveis) ficou bem distante do eixo horizontal, o que significa que os maiores tempos presentes nos dados são de observações censuradas. Além disso, pode-se observar também que os degraus nesta curva, que representam os tempos de falha, estão acumulados nos primeiros tempos, significando que a maioria dos retornos ocorreu em menos de 50 dias. Outra informação observada na curva refere-se aos tamanhos dos degraus, que representam a repetição de retornos em alguns tempos.

Figura 4 – Curva de sobrevivência obtida via Kaplan-Meier na ausência de covariáveis

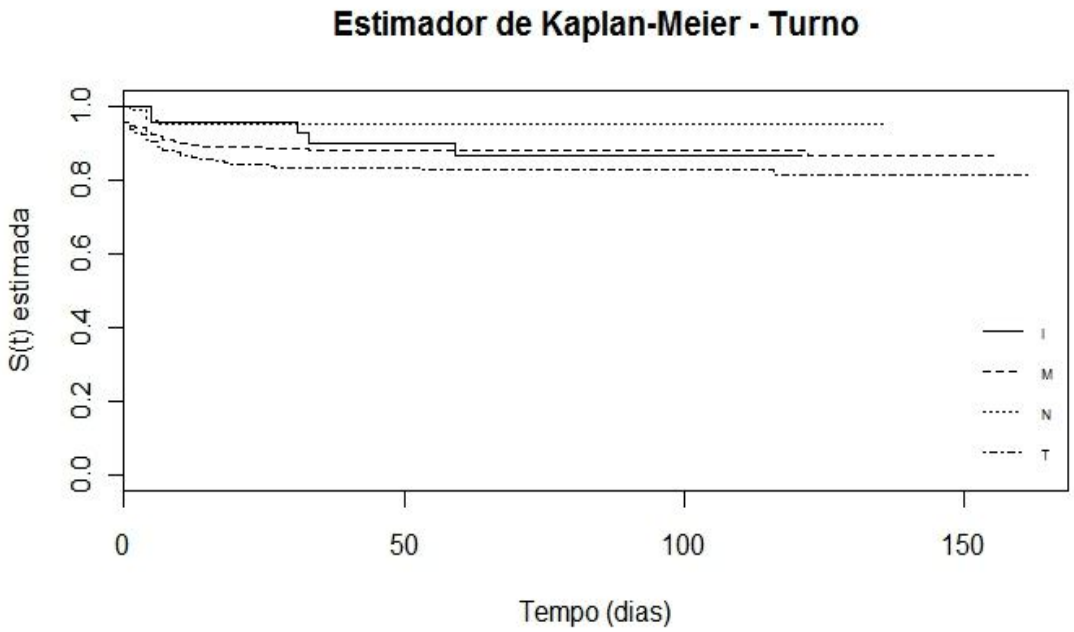


Fonte: Os autores (2013)

Quanto às curvas de sobrevivência considerando cada uma das covariáveis descritas na Tabela 1, estas também foram obtidas pelo estimador de Kaplan-Meier. A Figura 5 mostra tais curvas para a covariável turno, sendo possível observar que a curva com decréscimo mais acentuado está associada ao turno da tarde; o que mostra um retorno maior e em menor tempo para alunos desse turno. Para as covariáveis causas da evasão (aluno, social e outros) as curvas com decréscimos mais acentuados também ocorreram para os alunos em que a evasão não foi atribuída às mesmas. Analogamente, as curvas com decréscimos mais acentuados foram aquelas associadas às séries finais dos ensinos fundamental e médio (9º EF e 3º EM).

Ainda, a fim de avaliar se as curvas de sobrevivência estimadas para as categorias de cada uma das covariáveis apresentaram diferenças significativas entre elas, foi aplicado o teste *log-rank*, que retornou os p-valores apresentados na Tabela 5. Considerando 10% como nível de significância, notou-se que apenas a covariável CAUSA_ESCOLA não seria candidata a entrar no modelo.

Figura 5 – Curvas de sobrevivência obtidas via Kaplan-Meier na presença da covariável turno



Fonte: Os autores (2013)

Tabela 5 – Resultados do teste *log-rank* associados às covariáveis

Covariável	**Estatística *log-rank***	**P-valor**
SERIE	28,9	0,0001
NIVEL	15,9	0,0001
TURNO	22,3	0,0001
CAUSA_ESCOLA	2,4	0,1200
CAUSA_SOCIAL	3,7	0,0553
CAUSA_PAIS	58,6	0,0000
CAUSA_ALUNO	13,4	0,0003
CAUSA_OUTROS	35,0	0,0000
PROGRAMA_SOCIAL	6,5	0,0107

Fonte: Os autores (2013)

Foi, assim, ajustado o modelo de Cox, que com o auxilio do teste de Wald evidenciou as seguintes covariáveis com efeitos significativos: TURNO, SERIE, CAUSA_ALUNO, CAUSA_SOCIAL e CAUSA_OUTROS.

Para o modelo de Cox com as covariáveis mencionadas, foram analisados os gráficos dos resíduos de Schoenfeld mostrados na Figura A1 (Apêndice), sendo possível notar que a suposição de riscos proporcionais do modelo não foi violada, uma vez que não há evidências de tendências em nenhum deles.

Além disso, a Tabela 6 mostra também outra evidência de que a suposição de riscos proporcionais não está sendo seriamente violada. Isso é observado pelos valores de Rho pequenos ou próximos a zero.

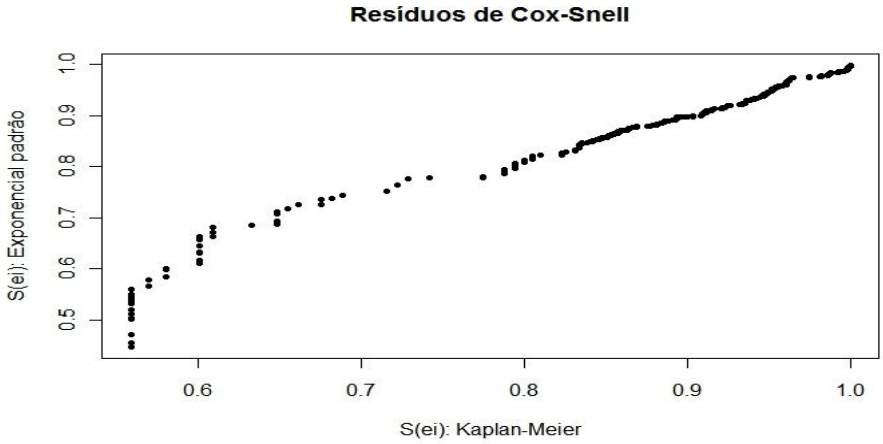
Tabela 6 – Correlação de Pearson entre os resíduos padronizados de Schoenfeld e os tempos t

	RHO		**RHO**
TURNOM	-0,2384	SERIE7EF	-0,0797
TURNON	-0,2348	SERIE8EF	0,0803
TURNOT	-0,2082	SERIE9EF	-0,0624
SERIE2EM	0,0364	CAUSA_ALUNO	0,0817
SERIE3EM	-0,0328	CAUSA_SOCIAL	0,2496
SERIE6EF	-0,0886	CAUSA_OUTROS	0,0523

Fonte: Os autores (2013)

Ainda é possível concluir a partir da Figura 6, que o modelo final apresentou ajuste razoável aos dados, uma vez que os resíduos de Cox-Snell seguem aproximadamente uma distribuição exponencial padrão.

Figura 6 – Resíduos de Cox-Snell associados ao modelo de Cox ajustado



Fonte: Os autores (2013)

Na Tabela 7 são apresentadas as estimativas dos parâmetros do modelo final.

Tabela 7 – Estimativas e intervalos de 95% de confiança associados ao modelo de Cox ajustado aos dados de 1.791 alunos evadidos do NRE Curitiba no segundo semestre letivo de 2011

	EXP(COEF)	EXP(-COEF)	INFERIOR .95	SUPERIOR .95
TURNOM	1,6134	0,6198	0,6113	4,2580
TURNON	0,6453	1,5497	0,2294	2,0518
TURNOT	1,8950	0,5277	0,7605	4,7219
SERIE2EM	0,5488	1,8222	0,2452	1,2283
SERIE3EM	1,2239	0,8171	0,5465	2,7407
SERIE6EF	1,4386	0,6951	0,8126	2,5471
SERIE7EF	1,3870	0,7210	0,7761	2,4791
SERIE8EF	1,1240	0,8896	0,6650	1,9000
SERIE9EF	1,8335	0,5454	1,1315	2,9712
CAUSA_ALUNO	0,2831	3,5318	0,2117	0,3788
CAUSA_SOCIAL	0,2351	4,2533	0,1184	0,4667
CAUSA_OUTROS	0,1176	8,5044	0,0667	0,2072

Fonte: Os autores (2013)

Dado que as covariáveis série e turno das aulas são categóricas (politômicas), foi estabelecido para cada uma delas um nível de referência. Para o turno, o nível de referência foi o integral/intermediário e, para a série, o primeiro ano do ensino médio. Assim, interpretando alguns dos parâmetros mostrados na Tabela 7, segue que a taxa de retorno dos alunos do turno da tarde foi 1,895 vez a dos alunos do turno integral/intermediário. Já a taxa de retorno dos alunos na nona série do ensino fundamental foi 1,8335 vez a dos alunos na primeira série do ensino médio. Analisando as variáveis causas, pode-se notar que a presença das mesmas diminui a taxa de retorno do aluno aos estudos. Nota-se, por exemplo, um risco menor de retorno (consequentemente um tempo maior até o retorno) associado aos alunos em que a causa foi atribuída a eles mesmos (em torno de 3,5 vezes menor), seguido dos alunos em que a causa foi atribuída à causas sociais (aproximadamente 4 vezes menor) e daqueles em que a causa foi atribuída a outras causas (8,5 vezes menor).

5 CONCLUSÕES

A evasão escolar é um problema ainda muito presente na realidade da educação brasileira e de outros países. Para combater este problema no estado do Paraná, o Governo Estadual criou o Programa Fica Comigo, sendo o presente estudo baseado nos dados de evasão e retorno do segundo semestre letivo de 2011 deste programa.

As análises realizadas neste trabalho tiveram como objetivo estudar a probabilidade de retorno dos alunos evadidos, bem como encontrar variáveis que afetam o tempo até o retorno destes alunos. Nesse sentido, foi, inicialmente, realizado um estudo descritivo para um melhor conhecimento dos dados, bem como para avaliar possibilidades de explorar técnicas estatísticas assertivas. Após este estudo, foram ajustados modelos de regressão logística a fim de encontrar variáveis associadas ao retorno dos alunos aos estudos. O modelo selecionado permaneceu com as seguintes variáveis: turno das aulas, série frequentada e causas da evasão (aluno, social e outros).

Para estudar o tempo até o retorno do aluno evadido foi aplicada a análise de sobrevivência. Para um estudo descritivo do tempo até o retorno foi utilizado o estimador de Kaplan-Meier. Na sequência dos estudos de sobrevivência, foi ajustado o modelo de Cox que também permaneceu com as variáveis: turno das aulas, série frequentada e causas da evasão (aluno, social e outros). Foram realizadas análises gráficas a fim de verificar a suposição de riscos proporcionais e a adequação deste modelo, com ambos sendo satisfeitos. O resultado final das análises mostrou que a taxa de retorno dos alunos que estudam no turno da tarde é maior do que a daqueles que estudam em turno integral/intermediário. Também foi observada uma taxa de retorno maior associada aos alunos nas séries finais de cada nível (9º EF e 3º EM). Quanto às causas de evasão (aluno, social, outros), pôde-se notar que a presença das mesmas diminui a taxa de retorno dos alunos evadidos aos estudos.

Tanto em análise de sobrevivência quanto em regressão logística, o perfil do aluno evadido que apresentou maior chance de retornar foi: o que estuda no turno da tarde, nas séries finais (tanto do EM quanto do EF) e que não apresenta nenhuma das três causas de evasão presentes no modelo.

Embora os resultados tenham indicado um percentual de retorno ainda inferior ao desejado, em sendo a educação o alicerce fundamental para o exercício da cidadania, se faz necessário a continuidade de esforços dos órgãos municipais, estaduais, federais e da sociedade em geral, a fim de assegurar o ingresso, regresso, permanência e sucesso de todas as crianças e adolescentes na escola.

REFERÊNCIAS

BATISTELA, G. C.; RODRIGUES, S. A.; BONONI, J. T. Estudo sobre evasão escolar usando regressão logística: análise dos alunos do curso de administração da Fundação Educacional de Ituverava. **Tékhne e Logos,** Botucatu, SP, v.1, n.1, p. 21-34, out. 2009.

BRASIL. Capítulo IV Art. 53, 2010. Do direito à educação, à cultura, ao esporte e ao lazer. **Estatuto da Criança e do Adolescente**. 7.ed., p.40. Disponível em: <http://bd.camara.gov.br/bd/bitstream/handle/bdcamara/785/estatuto_crianca_adolescente_7ed.pdf>. Acesso em: 10 abr. 2013.

BRASIL. Constituição (1988). **Constituição da República Federativa do Brasil**. Brasília, DF: Senado Federal: Centro Gráfico, 1988. Disponível em: <http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm>. Acesso em: 30 jan. 2013.

BRASIL. Instituto Nacional de Estudos e Pesquisas Anísio Teixeira (INEP). **Sinopse Estatística da Educação Básica 2004**. Brasília, 2005. Disponível em: <http://www.inep.gov.br/>. Acesso em: 10 abr. 2013.

CAVALCANTI, G.; SILVA, A. O.; FREITAS, W. W. L. **Modelos de sobrevivência aplicados a evasão dos alunos de estatística da UFPB**. In: I SIMPÓSIO DE MATEMÁTICA E ESTATÍSTICA DO DELTA (SIMED), 2012, Parnaíba, PI. Disponível em:< http://www.simed.estatistico.com/trabalhos/poster/SIMED_Poster010.pdf>. Acesso em: 10 jun. 2013.

COLOSIMO, E. A.; GIOLO, S. R. **Análise de Sobrevivência Aplicada**. São Paulo: Edgard Blucher, 2006. 370p.

COX, D. R. Regression Models and Life Tables (with discussion). **Journal Royal Statistical Society**, Serie B, v. 34, p.187-220, 1972.

FERREIRA, L. A. M. **Evasão Escolar**. Presidente Prudente, 2009. 14p. Disponível em: <http://www.mp.sp.gov.br>. Acesso em: 10 jun. 2013.

HADDAD, C. R.; FRANCO, A. F.; SILVA, D. V. Os motivos da evasão escolar: uma análise do Programa Fica. In: **Congresso Nacional de Educação – EDUCERE**, 10., 2011, Curitiba.

KAPLAN, E. L.; MEIER, P. Nonparametric estimation from incomplete observations. **Journal of the American Statistical Association**, v. 53, p. 457-81, 1958.

MARTINEZ, E. Z.; LOUZADA-NETO, F.; PEREIRA B. B. A curva ROC para testes diagnósticos. **Cadernos Saúde Coletiva**, Rio de Janeiro, v.11, n.1, p. 7-31, 2003.

MARTINS, G. O.; ROCHA, S. H. **Evasão e tempo de permanência no curso de estatística da Universidade Federal do Paraná**: um estudo sobre os alunos que ingressaram no período de 1991 a 2011. Curitiba, 2011. 79p. (Trabalho de Conclusão de Curso) Universidade Federal do Paraná, Curitiba, 2011.

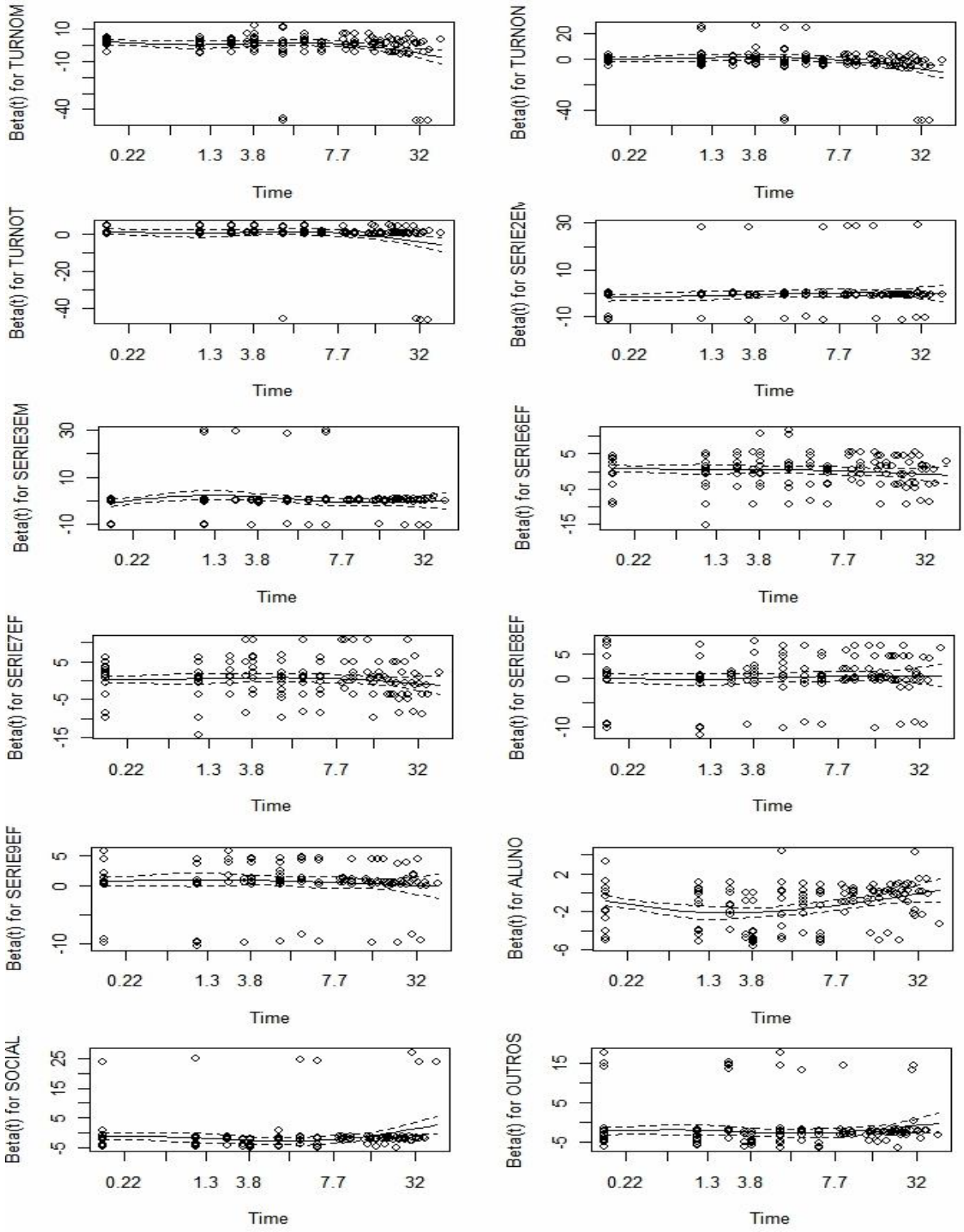
R DEVELOPMENT CORE TEAM (2013). **R: A language and environment for statistical computing.** R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.

SECRETARIA DE ESTADO DA EDUCAÇÃO. **Programa FICA Comigo: enfrentamento à evasão escolar**. Curitiba. 2009. Disponível em: <http://www.gestaoescolar.diaadia.pr.gov.br/arquivos/File/pdf/fica.pdf>. Acesso em: 16 abr. 2013.

APÊNDICES

APÊNDICE A

Figura A1 – Resíduos padronizados de Schoenfeld x tempo para as covariáveis no modelo de Cox



Fonte: Os autores (2013)

APÊNDICE B

B.1 Códigos R para ajuste do modelo de Regressão Logística

```
mod0 <-glm(RETORNO ~ NIVEL + SERIE + TURNO + PROG_SOCIAL + CAUSA_ALUNO +
CAUSA_PAIS + CAUSA_ESCOLA + CAUSA_SOCIAL + CAUSA_OUTROS, family = binomial,
data = BASE_CTBA_REG_LOG)
summary(mod0)
step(mod0,~ NIVEL + SERIE + TURNO + PROG_SOCIAL + CAUSA_ALUNO + CAUSA_PAIS +
      CAUSA_ESCOLA + CAUSA_SOCIAL + CAUSA_OUTROS,
direction=c("both"))

mod3<-glm(formula = RETORNO ~ SERIE + TURNO + CAUSA_ALUNO +CAUSA_SOCIAL +
          CAUSA_OUTROS, family = binomial(link="logit"), data = BASE_CTBA_REG_LOG)
summary(mod3)
attach(BASE_CTBA_REG_LOG)
require(Epi)
ROC(form=RETORNO~SERIE+TURNO+CAUSA_ALUNO+CAUSA_SOCIAL+CAUSA_OUTROS,
plot="ROC")
```

B.2 Códigos R para ajuste do modelo de Cox (Análise de Sobrevivência)

```
require(survival)
ekm <- survfit(Surv(TEMPOS,CENSURA)~1,conf.type="plain")
summary(ekm)
plot(ekm,mark.time=F,conf.int=F, xlab="Tempo (dias)", ylab="S(t) estimada", main="Curva de
Sobrevivência - Estimador de Kaplan-Meier")
ajust9<-coxph(Surv(TEMPOS,CENSURA)~TURNO+SERIE+ALUNO+SOCIAL+ OUTROS,
          data = BASE_FINAL_CTBA_SOB_V4, method = "efron")
summary(ajust9)

resid(ajust9,type="scaledsch")
cox.zph(ajust9,transform="identity")
plot(cox.zph(ajust9))
Ht<-basehaz(ajust9,centered=F)
tempos<-Ht$time
H0<-Ht$hazard
S0<- exp(-H0)
round(cbind(tempos, S0,H0),digits=5)
```

```
rd<-resid(ajust9,type="martingale")
res<-CENSURA-rd
ekm1<-survfit(Surv(res,CENSURA)~1)
t<-ekm1$time
st<-ekm1$surv
sexp<-exp(-t)
par(mfrow=c(1,1))
plot(st,sexp,xlab="S(ei): Kaplan-Meier",ylab="S(ei): Exponencial padrão",pch=20, main="Resíduos
de Cox-Snell")
```