IDENTIFICAÇÃO DAS CAUSAS DA EVASÃO E RETENÇÃO PROLONGADA DE ESTUDANTES DO ENSINO SUPERIOR A PARTIR DE DADOS CENSURADOS

Camila da Silva Oliveira
Universidade Federal de Pernambuco, Departamento de Engenharia de Produção - CTG
Av. da Arquitetura, s/n, Bloco Administrativo do CTG, 5º andar, Cidade Universitária, Recife, PE
camilasiol@hotmail.com

Luciano Nadler Lins
Universidade Federal de Pernambuco, Departamento de Engenharia de Produção - CTG
Av. da Arquitetura, s/n, Bloco Administrativo do CTG, 5º andar, Cidade Universitária, Recife, PE
lucianolins@ufpe.br

RESUMO

O objetivo principal do trabalho tem foco no desenvolvimento de uma sistemática para identificação e análise dos fatores condicionantes da saída ou da permanência excessiva de estudantes em cursos de nível superior. O trabalho emprega técnicas estatísticas descritivas e inferências de Análise de Sobrevivência para ressaltar dentre um conjunto de covariáveis, extraídas de um sistema de gestão acadêmica, quais poderiam estar afetando diretamente o encerramento do vínculo devido à formatura ou ao abandono em definitivo do curso. Para ilustrar a aplicação da técnica, foram empregados dados da evolução acadêmica de alunos ingressantes em um conjunto de turmas anuais consecutivas de um mesmo curso superior de uma universidade federal pública brasileira. Tem-se como característica a presença de censuras, na medição do tempo, originárias de alguns estudantes ainda estarem vinculados ao curso no momento da coleta. Os resultados apontam como algumas covariáveis teriam influenciado adversamente os fenômenos da evasão e retenção prolongada discente.

PALAVRAS CHAVE. Evasão, Retenção Prolongada, Ensino Superior.

Área principal. EDU.

ABSTRACT

The main purpose of the work is focused on developing a systematic approach to the identification and analysis of the conditioning factors for student dropout or excessive permanence in higher education courses. Descriptive and inferential statistics from survival analysis were employed to highlight a set of covariates, extracted from an academic management system, which could directly affect the course completion or its abandonment. To illustrate the application of the selected technique, data about the academic progress of groups of students that started a same course in a public federal university in Brazil during consecutive years was used. This data was characterized by the presence of censorship, because some students were still tied to the course at the moment of the data collection. The results show how some covariates could have adversely influenced the phenomenon of dropout and prolonged retention of students.

KEYWORDS. Dropout, Retention, Higher Education.

Main area. EDU.

1. Introdução

Tanto a evasão estudantil como sua retenção prolongada em cursos de nível superior são problemas de relevância internacional, pois afetam sobremaneira o desempenho dos sistemas educacionais. Embora formem processos naturais e muitas vezes inevitáveis, devem ser combatidos pelos gestores das instituições de ensino no sentido de minimizar seus efeitos nocivos.

A perda de estudantes que iniciam, mas não terminam seus cursos, produzem desperdícios não apenas acadêmicos, mas certamente sociais e econômicos. No setor público, recursos públicos investidos na formação acadêmica de alunos acabam não produzindo o devido retorno. No setor privado, representa uma importante redução de receitas. Em ambos os casos, tem-se uma relativa ociosidade de professores, funcionários, equipamentos e espaço físico (SILVA Fo. *et al.*, 2007). Enquanto isso, a retenção prolongada caracteriza-se pela permanência excessiva em um curso de graduação, no qual o aluno, por diversos motivos, leva um tempo maior para concluí-lo do que aquele planejado no respectivo projeto pedagógico de curso (CISLAGHI, 2008).

As causas da evasão são complexas, por serem diversas e geralmente correlatas. Contudo, diversos autores concordam que a taxa de evasão é mais visível no primeiro ano do curso (MACHADO *et al.*, 2005; CISLAGHI, 2008; SILVA Fo. *et al.*, 2007). A evasão acentuada neste primeiro ano pode estar relacionada a uma incorreta escolha do curso de graduação ainda no vestibular, devido à idade dos estudantes, ao pouco conhecimento sobre os cursos e suas diferenças, ou à escolha de outro curso ou turno diferente a fim de encontrar uma menor concorrência no vestibular (MACHADO *et al.*, 2005).

De acordo com Ahlburg *et al.* (2002 *apud* CORRÊA & NORONHA, 2011), os fatores mais influentes na evasão são: o tempo de demora existente entre a saída do aluno do ensino médio e seu ingresso no ensino superior, o apoio familiar, as características pessoais e as condições do mercado de trabalho local. Cunha et al. (2001) separaram as razões da saída dos alunos de seus respectivos cursos em cinco modalidades. São elas, em ordem decrescente de prioridade: desligamento por não cumprimento de condição (44,8%); desligamento voluntário (28,7%); desligamento por abandono (19,7%); mudança de curso (7,2%); e transferência (1,9%). Isto mostra que a evasão é decorrente de fatores diversos, não raro de difícil monitoramento.

Lima (2006) concorda que os motivos geradores da evasão escolar estão ligados a: carência de uma orientação adequada antes da efetiva escolha do curso superior, o que geralmente leva a uma má escolha quanto ao curso desejado; alto índice de reprovação nos primeiros períodos, o que produz desmotivação nos alunos; e dificuldades financeiras, dificultando a permanência do aluno em um curso superior. Santos (2001 *apud* LIMA, 2006) também aponta para a existência de fatores motivacionais da evasão associados com a segurança e o bem-estar do indivíduo, tais como o ajustamento com o curso, a vocação para a carreira escolhida, as oportunidades profissionais, os estímulos sociais e econômicos e o *status* profissional.

Constata-se de uma forma geral a inexistência de consenso acerca das razões que levam alunos de ensino superior a deixarem ou mesmo permanecerem excessivamente em seus respectivos cursos, o que torna o problema de difícil solução. Em todas as situações, pode ser observado um mal aproveitamento dos recursos disponíveis, sejam eles de origem pública ou privada, decorrente do não alcance do objetivo de formar alunos dentro do tempo estabelecido para a conclusão do curso. Faz-se, assim, necessário o emprego de uma metodologia que possibilite identificar preventivamente os alunos com potencial para evasão ou retenção prolongada, a partir de características-chave, o que facilitaria a elaboração de estratégias para a mitigação dos impactos negativos destes fatores em todo o sistema de ensino.

A metodologia selecionada para o desenvolvimento deste presente trabalho visa a estimação e análise das curvas provenientes da técnica estatística de análise de sobrevivência. Esta ferramenta é aplicável quando se pretende analisar dados ao longo do tempo, buscando, entre outras informações, o tempo de ocorrência de um dado evento de interesse. Entende-se por sobrevivência a chance de um indivíduo pertencente a um dado grupo ser capaz de ultrapassar um instante de tempo sem que o evento tenha ocorrido. Em particular, interessa a característica desta técnica em permitir o tratamento de dados censurados, referentes a medições de tempo em que, sabe-se, o evento ainda não ocorreu.

Na análise de sobrevivência aplicada a discentes, os eventos de interesse são em geral a ocorrência de eventos acadêmicos, como a aprovação/reprovação em disciplinas, a realização de estágios curriculares, a participação em programas de mobilidade estudantil, entre outros, mas

principalmente costumam ser relativos ao encerramento do vínculo com o curso por algum motivo. Este último caso configura-se como uma evasão, podendo ser observada quando o aluno não efetua matrícula em algum semestre letivo, ou realiza transferência de curso, ou vem a ser jubilado, ou ainda, por completar o limite de períodos autorizados pelo curso para integralização curricular. Entende-se também que a retenção prolongada é outro evento de interesse identificado pelo aluno que demora a se formar.

No Brasil, essa metodologia ainda não foi muito aplicada na área da educação, devido à dificuldade de se estruturar bancos de dados adequadamente. Um exemplo de trabalhos realizados nesta direção foi o de Gonçalves (2008), que se propôs a identificar e analisar os determinantes probabilísticos de ocorrência da primeira repetência, da primeira reprovação, do primeiro abandono e da evasão entre alunos da 4ª a 8ª série do ensino fundamental, comparando instituições de seis unidades federativas entre os anos de 1999 e 2003. Um outro estudo, conduzido por Alencar (2007), emprega a análise de sobrevivência para defender o programa Bolsa Família do Governo Federal do Brasil como instrumento de combate à abstenção e à evasão escolar. Entretanto, este programa necessita ainda de um maior aprimoramento, pois deixa de divulgar fatores internos importantes de risco, e que podem ser facilmente monitorados e controlados pelas escolas, por influenciarem significativamente a decisão dos alunos em não mais freqüentar a escola antes mesmo de completar o ciclo básico escolar.

Como resultados observados, apresentam-se as curvas de sobrevivência para todos os alunos de um mesmo perfil curricular, ou estratificados com base em alguma característica relevante, tal como o gênero do aluno, seu desempenho acadêmico, a existência de reprovações em seu histórico escolar, solicitações prévias de trancamento de semestre, entre outros fatores. Também, sempre que adequado, foram apresentadas as estimativas do tempo esperado até a ocorrência do evento de interesse.

2. Análise de Sobrevivência

A análise de sobrevivência é uma das áreas da Estatística que mais cresceu nos últimos vinte anos, devido ao expressivo número de aplicações, especialmente na área de saúde, mas também em tecnologia, economia, educação, entre outros. É utilizada para designar o tratamento estatístico de dados quando a variável principal representa o tempo desde um instante inicial bem definido até a ocorrência de determinado acontecimento (KLEIN & MOESCHBERGER, 1983).

A análise de sobrevivência possui como característica fundamental a existência de censuras, que são basicamente observações parciais da resposta. Tem-se que, por alguma razão, que o acompanhamento do objeto investigado foi interrompido antes do final do estudo. Isto significa que toda a informação referente a estas respostas se resume ao conhecimento de que o tempo de falha é superior àquele observado. A grande vantagem da análise de sobrevivência é que ela então permite aproveitar toda informação dos participantes até ao momento em que ocorre o evento de interesse (COLOSIMO & GIOLO, 2006). Assim, a análise de sobrevivência é a técnica ideal para analisar respostas binárias (ter ou não ter um evento) relacionadas aos tempos dos indivíduos. Outra possibilidade é a ocorrência de truncamentos, caracterizado por uma condição que exclui certos elementos da amostra. Neste caso, os participantes não são acompanhados a partir do tempo inicial, mas somente após experimentarem um certo evento (COLOSIMO & GIOLO, 2006).

As observações resultantes são normalmente chamadas de tempo de vida, ou tempo de falha, o que no contexto significa o tempo até o encerramento do vínculo com o curso. O tempo de vida é uma variável aleatória sempre não negativa e pode representar, por exemplo, na área de educação, o tempo até o egresso do aluno tanto devido à formatura como ao abandono prematuro do curso. Estes dois componentes constituem a resposta básica da ferramenta (COLOSIMO & GIOLO, 2006):

a) *Tempo de vida*, constituído pelos seguintes elementos: o instante inicial, a escala de medida e o evento de interesse. O instante inicial deve ser precisamente definido a partir de um marco inicial claramente definido. A escala de medida é quase sempre o tempo real, apesar de existirem outras alternativas, como, por exemplo, intervalos regulares de tempo. E por fim, o evento de interesse que determina o término do acompanhamento, e que na maioria das vezes é algo indesejável ou pelo menos relevante.
b) *Dados censurados*, nomeadamente quando existem longos períodos de monitoração, muitos participantes não atingem o tempo total de acompanhamento previsto. Em

análise de sobrevivência, os dados destes participantes são sempre aproveitados, mesmo que não apresentem o evento. Tal situação pode ocorrer porque eles abandonaram o estudo, foram perdidos na sequência ou o estudo chegou ao fim. Quando isso acontece, designa-se o indivíduo por censurado, independentemente do motivo que provocou sua interrupção.

Os mecanismos de censura são diferenciados em (COLOSIMO & GIOLO, 2006):

- Censura à direita do tipo I: aquela em que o estudo será terminado após um período pré-estabelecido de tempo independente das ocorrências do evento de interesse.
- Censura à direita do tipo II: quando as observações são acompanhadas até se obter um número pré-determinado de ocorrências do evento de interesse.
- Censura aleatória: acontece quando um participante sai ou é retirado do acompanhamento sem que tenha sido observada a ocorrência do evento de interesse.

O objetivo principal da análise de sobrevivência é descrever o comportamento da variável aleatória *tempo de vida* por meio da chamada curva de sobrevivência, uma das principais funções probabilísticas usadas para descrever dados de tempo com a presença de censuras ou truncamentos. Ela é definida como a probabilidade de uma observação não falhar até certo tempo t, ou seja, a chance do evento de interesse ser observação após o tempo t. Isto é descrito como $S(t) = P[T > t]$, sendo uma função sempre monotônica decrescente. A partir desta função $S(t)$ é possível também estimar outras funções ou índices, sendo mais comuns a taxa de falhas ou risco, a taxa de falhas acumulada, o tempo médio de vida, o tempo mediano de vida e a vida média residual (COLOSIMO & GIOLO, 2006).

O estimador de Kaplan-Meier é o mais utilizado na estimação da função de sobrevivência por ser o estimador de máxima verossimilhança. O método consiste em dividir o tempo total de em intervalos, cujos limites correspondem ao tempo em que eventos efetivamente ocorreram. Esta análise permite avaliar corretamente o ritmo com que os eventos vão ocorrendo nos diferentes grupos. Das curvas de Kaplan-Meier pode-se estimar diretamente a proporção relativa a qualquer intervalo temporal.Quando alguma variável de agrupamento permite segregar os dados completos em duas subamostras, é desejável saber se as curvas estratificadas são realmente distintas. Esta constatação pode ser realizada visualmente por meio das estimativas das curvas de sobrevivência ou, de forma mais precisa, pela aplicação de testes estatísticos apropriados (COLOSIMO & GIOLO, 2006).

Uma alternativa popularmente adotada para comparar de curvas de sobrevivência consiste no teste *logrank*, muito apropriado quando as funções de risco de mais de um grupos de dados são proporcionais (COLOSIMO & GIOLO, 2006). A estatística do teste *logrank*, para o caso de covariáveis dicotômicas, segue uma distribuição qui-quadrado com 1 grau de liberdade. No entanto, para que seja possível a aplicação de tal procedimento estatístico, faz-se então necessário o emprego de técnicas que permitam certificar a proporcionalidade dos riscos entre cada subamostra gerada. Embora vários métodos tenham sido propostos para esta finalidade, a proporcionalidade dos riscos será confirmada no presente trabalho tanto graficamente como por meio de teste de hipóteses. O primeiro método consiste na observação visual em um mesmo gráfico para ambos os grupos da função do logaritmo dos riscos acumulados *versus* o tempo. Quando as curvas se mostram com afastamento uma da outra aproximadamente fixo ao longo de todo o eixo temporal, tem-se uma indicação de que os riscos devem ser proporcionais. O segundo procedimento baseia-se no modelo paramétrico de Cox, em que se deve avaliar se o parâmetro estimado para a covariável correspondente não é função do tempo, o que também sinalizaria o atendimento ao pressuposto da proporcionalidade dos riscos. A estatística principal do último teste também possui, para o caso de covariáveis dicotômicas, uma distribuição qui-quadrado com 1 grau de liberdade.

3. Caracterização dos dados

A coleta de informações para realização da análise foi efetuada em uma base de dados de curso de graduação de uma universidade federal pública brasileira. São dados do acompanhamento dos estudantes que ingressaram no curso ao longo de 6 anos, formando 6 turmas consecutivas. A unidade amostral foi determinada como um ingresso de um estudante em algum dos anos indicados, ou seja, um caso na base corresponde a um vínculo. O referido curso de graduação oferecia na época uma entrada única anual com 40 vagas através de processo seletivo vestibular. Como dados relativos às

demais modalidades de ingresso foram desprezados, a base totalizou ao final 240 casos.

A grade curricular vigente à época previa a integralização dos créditos minimamente exigidos durante 8 semestres letivos em turno integral. Tais créditos representavam uma carga horária total de 3270 horas, dispostas em 2205 horas de componentes obrigatórios de formação básica e profissional e 1605 horas de componentes eletivos de formação complementar.

A estrutura da base de dados foi organizada visando atender à proposta de mensuração do tempo, com base no número de semestres letivos, até o egresso do aluno devido à formatura ou por abandono do curso. Todos os dados da evolução acadêmica dos alunos ingressantes nas turmas investigadas foram coletadas até o término de um semestre de referência adotado para este trabalho, equivalente a 22 semestres letivos após o início da primeira turma. O banco de dados completo, no entanto, apresentou um total de 14 variáveis, e que se encontram descritas a seguir.

As variáveis de identificação dos casos foram:

- *caso*: identificação de cada aluno com vínculo no curso nos anos considerados;
- *turma*: ordem cronológica das turmas a partir do primeiro ano de funcionamento da grade curricular analisada.

A variável numérica da medição temporal em intervalos discretos foi:

- *período*: número de semestres letivos em que o aluno esteve vinculado ao curso até o semestre de referência, excluindo os trancamentos de semestre. O período ainda foi corrigido ao se adicionar o número de semestres equivalentes às dispensas obtidas, tendo sido calculado pela divisão desta carga horária dispensada por 360 horas e depois arredondada para o inteiro mais próximo.

As variáveis dicotômicas indicadoras da ocorrência dos eventos de interesse foram:

- *formado*: indica se no semestre de referência o aluno terá se formado no curso;
- *abandonado*: indica se no semestre de referência o aluno terá deixado o curso sem chegar a concluí-lo;
- *desvinculado*: indica se no semestre de referência o aluno terá encerrado o vínculo com o curso, seja por formatura ou por abandono.

Para todas estas variáveis indicadoras de evento, cada aluno ainda vinculado ao curso no semestre de referência representa uma censura à direita tipo I. Além disso, *abandonado* produz uma censura aleatória quando *formado* for utilizada como a variável indicadora de evento, e *formado* produz uma censura aleatória quando *abandonado* for utilizada como a variável indicadora de evento.

As variáveis dicotômicas usadas para agrupamento dos estudantes foram:

- *gênero*: sexo dos alunos ingressantes nas turmas analisadas;
- *dispensa*: indica se o aluno já obteve dispensa de créditos obrigatórios ou eletivos exigidos em seu perfil curricular;
- *trancamento*: indica se em algum semestre letivo o aluno realizou trancamento de semestre, que assegura a manutenção do vínculo, mas não possibilita o cumprimento de créditos pela participação em atividades acadêmicas naquele semestre. Os trancamentos interrompem a contagem do período, mas são limitados a 4 ao todo, sejam eles consecutivos ou não.
- *vínculo*: indica se em algum semestre letivo o aluno efetuou matrícula-vínculo, que assegura a manutenção do vínculo sem permitir a efetiva participação em disciplinas do curso naquele semestre. A matrícula-vínculo se difere do trancamento por não interromper a contagem de períodos e não ter limite no número de solicitações.
- *reprovação*: indica se o aluno já foi reprovado em algum componente curricular em que se matriculou;
- *desempenho*: indica se a média geral do aluno estava acima de 7,0 no semestre de referência.

Estas variáveis de agrupamento são também denominadas de covariáveis, e dentre elas: *gênero* sinaliza uma característica pessoal; *desempenho* e *reprovação* representam performance acadêmica; *trancamento* e *vínculo* indicam ausência temporária do curso; e *dispensa* sinaliza um histórico de atividades complementares ao curso, anterior ou concomitante.

Uma vez que a análise de sobrevivência está relacionada ao tempo de ocorrência do evento, quando é atribuído um valor a esse tempo, escolhe-se na realidade uma escala de tempo e uma origem.

Na presente situação a escala adotada foi a semestral. Por outro lado, a origem foi determinada a partir da efetivação da primeira matrícula dos alunos no curso de graduação.

Dois aspectos relacionados aos dados devem ainda ser mencionados: ao fim do período de acompanhamento parte dos alunos não experimentou os eventos de interesse; entre os anos de vínculo houveram diversos casos de evasão, os seja, o acompanhamento do aluno foi interrompido. A censura corresponde à observação parcial da resposta, mas, apesar de incompletas, as observações censuradas dão informações sobre o tempo de sobrevivência dos alunos. É a presença de dados censurados que requer o uso do método da análise de sobrevivência.

Após finalizada a análise de dados, procura-se saber ao término que cada período se o aluno se apresenta formado ou desvinculado. A idéia foi estimar as curvas de sobrevivência entre a entrada no curso e o evento ou censura. Os parâmetros mais importantes são a sobrevivência no curso a cada instante, isto é, a probabilidade do aluno permanecer no curso desde o período zero até o último período observado.

4. Análise dos Resultados

Deseja-se comparar as curvas de sobrevivência estimadas, visando verificar se os alunos, ao ingressarem no curso de graduação, têm uma maior probabilidade de sobrevivência até o 8º período, o tempo de mínimo para integralização curricular, e mínima probabilidade de sobrevivência após esse período. Variações em relação a este padrão representam, respectivamente, a presença dos fenômenos da evasão e da retenção prolongada. O modelo ideal em comparação ao formato real seria descrito, portanto, conforme a figura 1. Isto significa que se deseja o menor risco de falha possível até o tempo mínimo de integralização curricular e o maior risco de falha possível após este tempo.

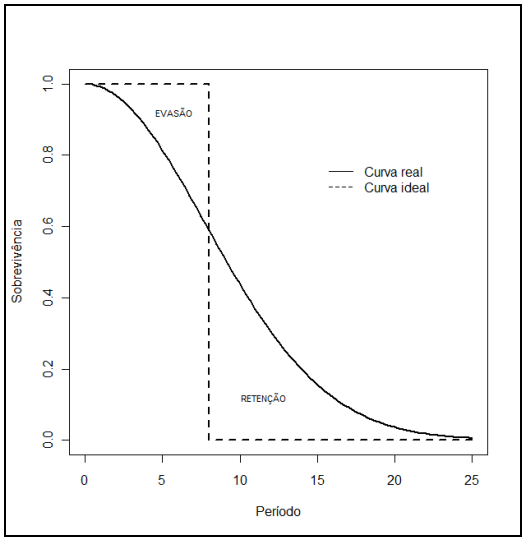

Figura 1 – Curva de sobrevivência real e ideal à perda do vínculo dos alunos no curso

Foram então estimadas as curvas de sobrevivência em relação aos eventos formatura e abandono, separadamente e em conjunto, além do cálculo dos demais índices e funções, quando apropriados. A análise separada se mostrou útil principalmente devido aos alunos que ultrapassaram o limite mínimo de integralização curricular, pois eles ainda corriam o risco de apresentar um dos dois eventos. Todos os resultados descritos foram validados graficamente ou por meio de testes estatísticos, com o auxílio do software estatístico livre R (THE R FOUNDATION, 2011), de modo a serem estatisticamente significativos ao nível de 0,05.

Inicialmente, a figura 2 mostra que a probabilidade de sobrevivência ao término do vínculo nos primeiros períodos de curso é bem menos intensa do que nos últimos períodos de curso. Insta ressaltar, a brusca queda de probabilidade em torno do 8º período, o tempo mínimo para integralização curricular, na qual se pode identificar o início da formatura por uma considerável parcela da turma.

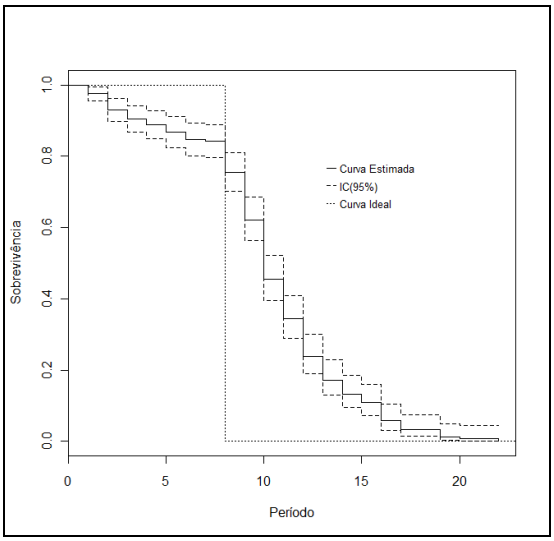

Figura 2 – Curva de sobrevivência considerando conjuntamente alunos que saem do curso por formatura ou por abandono

Dos dados de sobrevivência para o evento do término do vínculo, estimou-se o tempo médio até o encerramento do curso como sendo de 10,22 períodos, enquanto mediana deste mesmo tempo foi estimada em 9,72 períodos, ambas, portanto, acima dos 8 períodos previstos. Em particular, o tempo médio se mostra adequado, pois o maior tempo observado nos dados representava a ocorrência de um evento (formatura), evitando assim uma estimativa enviesada desta grandeza (COLOSIMO & GIOLO, 2006). O tempo mediano também foi determinado, pois mais da metade da turma havia encerrado o curso no semestre de referência.

Os gráficos da figura 3 apresentam curvas de sobrevivência considerando inicialmente apenas os alunos que encerraram o curso por formatura e em seguida apenas os alunos que abandonaram o curso. Pode-se observar, no primeiro gráfico, que a sobrevivência destas turmas à formatura até o oitavo período permanece constante e igual a 1. Após este tempo a probabilidade decresce monotonicamente de maneira assintótica ao eixo do tempo. No segundo gráfico, a curva de sobrevivência por abandono tem um comportamento bem diferente do gráfico citado anteriormente, pois o decrescimento se dá a uma taxa aproximadamente constante, com a sobrevivência ao abandono sendo reduzida em pequenas parcelas desde o ingresso do aluno no primeiro período.

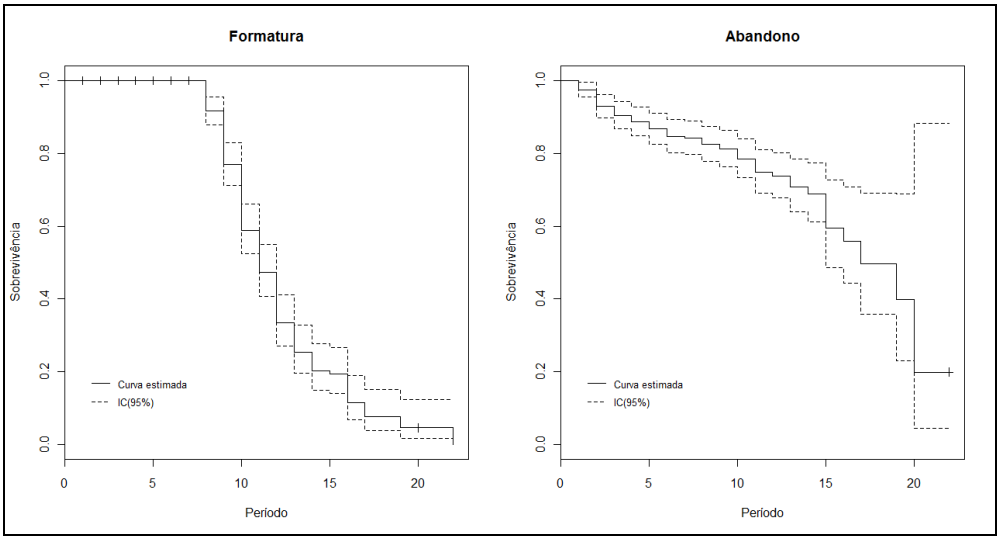

Figura 3 – Curva de sobrevivência à formatura no curso e ao abandono do curso

Nas figuras 4 e 5 estão expostos os gráficos do logaritmo dos riscos acumulados *versus* o período de vínculo dos estudantes, considerando os grupos criados pelas diversas covariáveis.

Observa-se que as funções de sobrevivência categorizadas pela variável *dispensa* poderão não implicar na proporcionalidade dos riscos desejada, pois o afastamento entre as curvas tende a não se manter constante ao longo do tempo, sendo um aspecto mais evidente no gráfico para abandono.

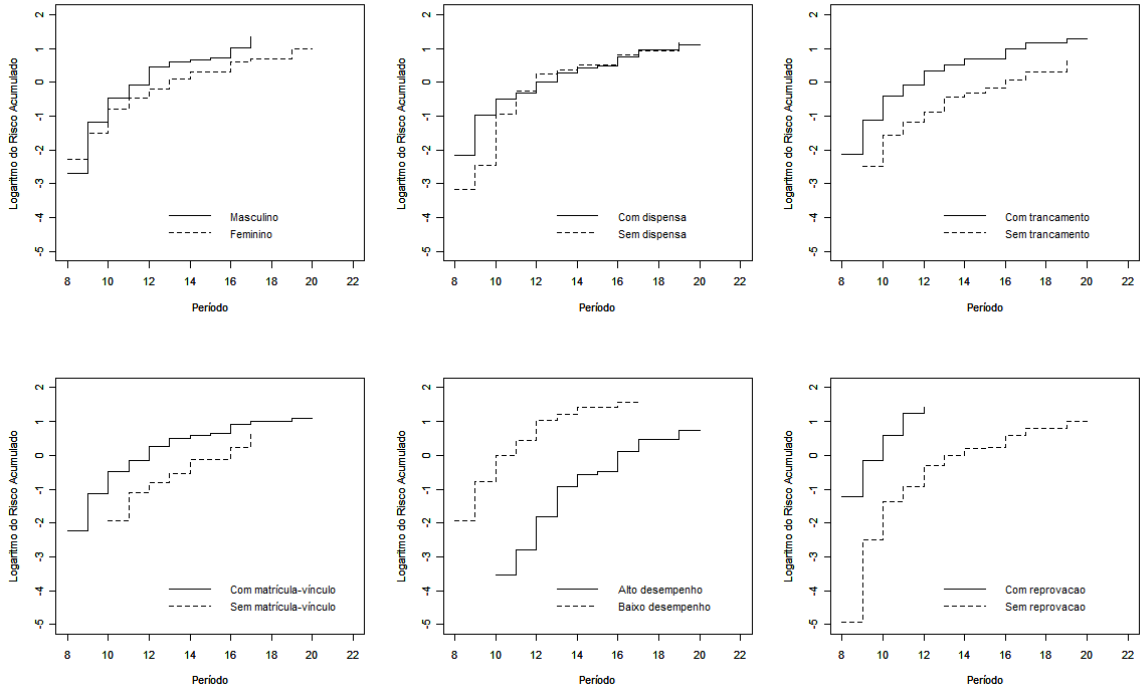

Figura 4 – Logaritmo dos riscos acumulados de formatura

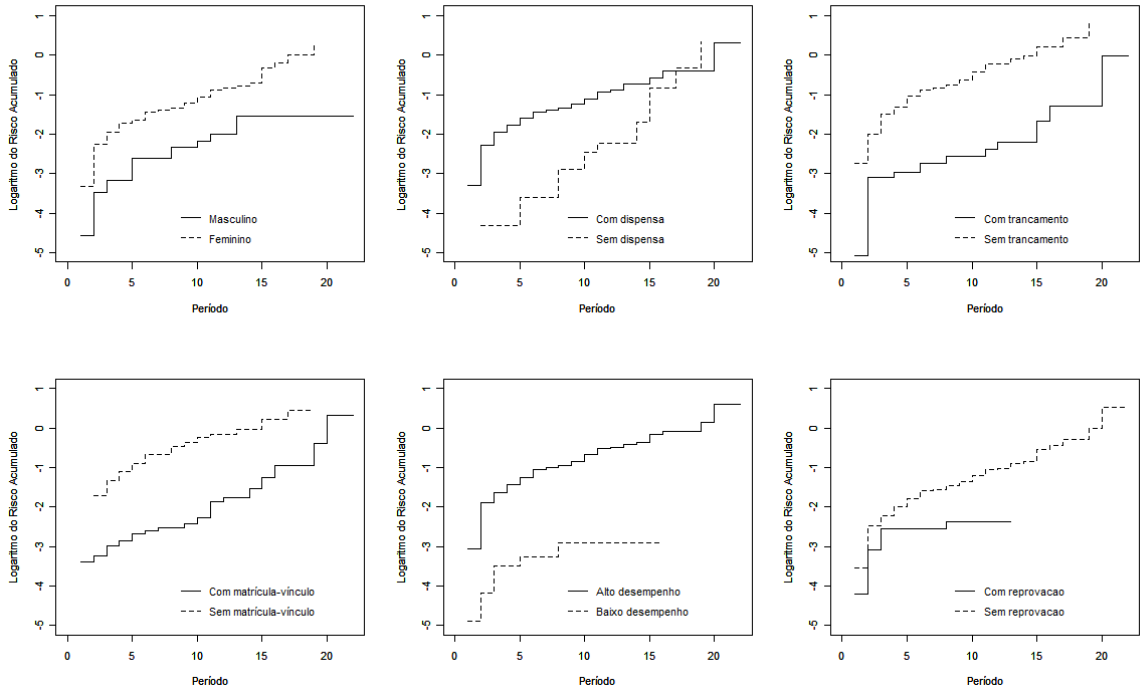

Figura 5 – Logaritmo dos riscos acumulados de abandono

Entretanto, a subjetividade do julgamento gráficos pelas figuras 4 e 5 pode ser superada pelo teste estatístico de proporcionalidade baseado no modelo de Cox, cujo resultados encontram-se indicados na tabela 1, com destaque naqueles significativos, dado o valor crítico desta estatística de

teste em 3,84. O valor *p* confirma que a covariável *dispensa*, de fato, não oferece riscos proporcionais, enquanto para as covariáveis *vínculo* e *desempenho* isto se verifica somente nas curvas de sobrevivência à formatura.

Tabela 1 – Teste de proporcionalidade dos riscos pelo modelo de Cox

Variáveis	**Formatura**		**Abandono**	
	Estatísticas de Teste	**Valor *p***	**Estatísticas de Teste**	**Valor *p***
gênero	$\chi_1^2 = 0{,}24$	0,62	$\chi_1^2 = 0{,}29$	0,59
dispensa	$\chi_1^2 = 4{,}75$	**0,03**	$\chi_1^2 = 8{,}1$	**<0,01**
trancamento	$\chi_1^2 = 2{,}13$	0,14	$\chi_1^2 = 0{,}12$	0,73
vínculo	$\chi_1^2 = 7{,}75$	**<0,01**	$\chi_1^2 = 0{,}98$	0,32
desempenho	$\chi_1^2 = 14{,}50$	**<0,01**	$\chi_1^2 = 037$	0,54
reprovação	$\chi_1^2 = 0{,}74$	0,39	$\chi_1^2 = 0{,}50$	0,48

Para as combinações tidas como de riscos proporcionais, aplicou-se o teste *logrank*, cujos resultados encontram-se indicados na tabela 2, com destaque naqueles significativos, dado o valor crítico desta estatística de teste em 3,84. O valor *p* confirma que, para todas as situações onde os riscos foram aceitos como proporcionais, houve uma diferença entre as curvas de sobrevivência (teste bilateral).

Tabela 2 – Aplicação do teste *logrank* para a diferença entre as curvas de sobrevivência

Variáveis	**Formatura**		**Abandono**	
	Estatística de Teste	**Valor *p***	**Estatística de Teste**	**Valor *p***
gênero	$\chi_1^2 = 6{,}8$	**<0,01**	$\chi_1^2 = 14{,}8$	**<0,01**
dispensa	----	----	----	----
trancamento	$\chi_1^2 = 21{,}2$	**<0,01**	$\chi_1^2 = 62{,}2$	**<0,01**
vínculo	----	----	$\chi_1^2 = 50{,}3$	**<0,01**
desempenho	----	----	$\chi_1^2 = 44{,}8$	**<0,01**
reprovação	$\chi_1^2 = 114$	**<0,01**	$\chi_1^2 = 6{,}8$	**<0,01**

Nas figuras 6 a 10 são apresentadas as curvas de sobrevivência agrupadas segundo as covariáveis levantadas. Como propriedade dos riscos proporcionais, as curvas estimadas tendem a não se tocar, fazendo com que uma esteja sempre acima da outra em todos os períodos. Portanto, para os gráficos da formatura, a curva superior indica uma maior probabilidade do aluno se formar, sendo menor o risco de retenção prolongada. Para os gráficos do abandono, a curva superior indica uma maior probabilidade do aluno deixar o curso, sendo maior o risco de evasão.

Por exemplo, ao se analisar os gráficos da figura 6, fica evidente que as mulheres têm taxas de sobrevivência à formatura menor ao longo dos semestres após o tempo de integralização mínimo. Ou seja, as mulheres se formam com mais regularidade no tempo projetado do curso (8 períodos) em comparação aos homens, que demoram mais tempo para se formar (taxas de probabilidade maior que as das mulheres). No segundo gráfico, a probabilidade de sobrevivência ao abandono por parte dos homens é menor, isto é, eles sobrevivem menos à evasão em relação às mulheres. Os homens também permaneceram mais retidos excessivamente ao longo do curso em comparação às mulheres. Comentários análogos podem ser feitos para os demais gráficos.

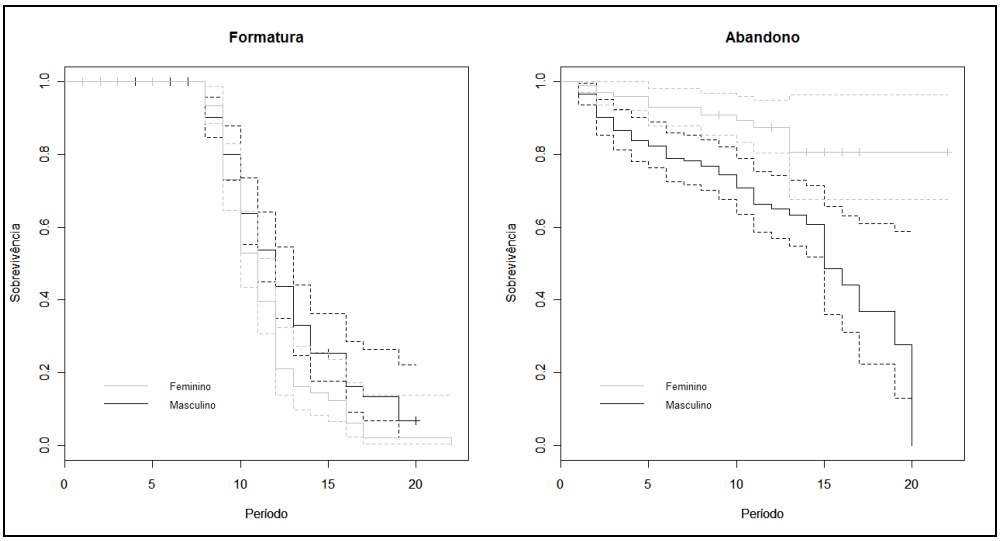

Figura 6– Curvas de sobrevivência à formatura e ao abandono estratificadas por gênero

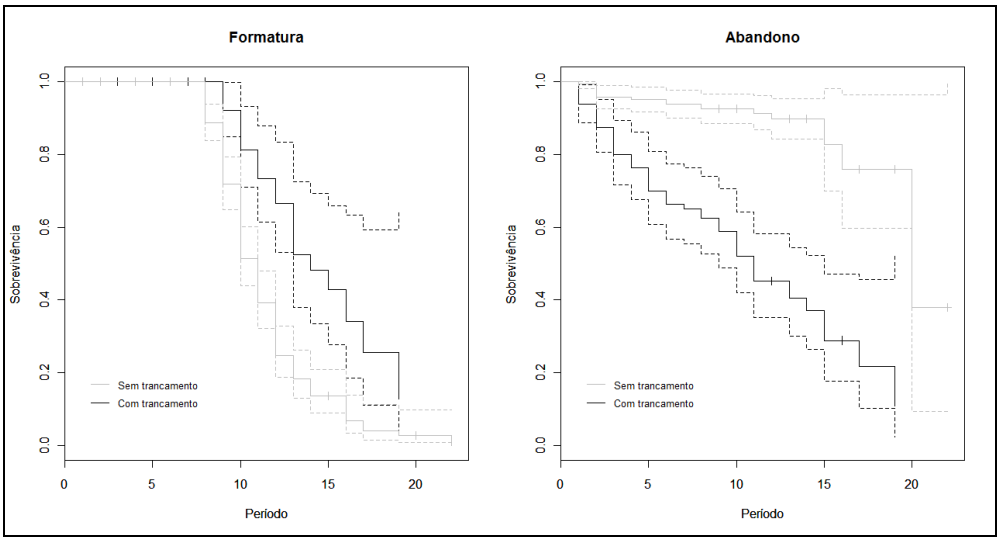

Figura 7 – Curvas de sobrevivência à formatura e ao abandono estratificadas por trancamento

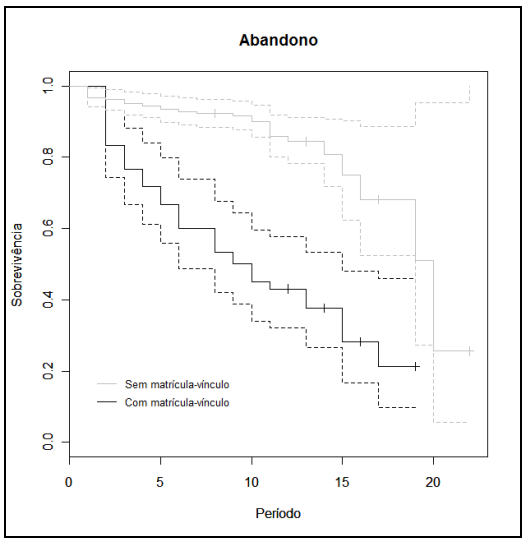

Figura 8 – Curva de sobrevivência ao abandono estratificada por matrícula-vínculo

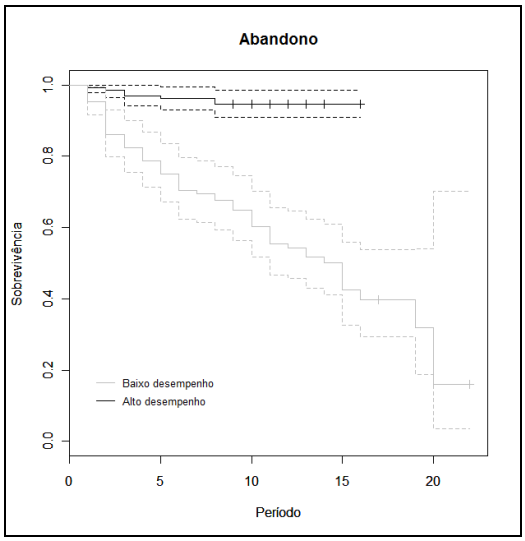

Figura 9 – Curva de sobrevivência ao abandono estratificada por desempenho

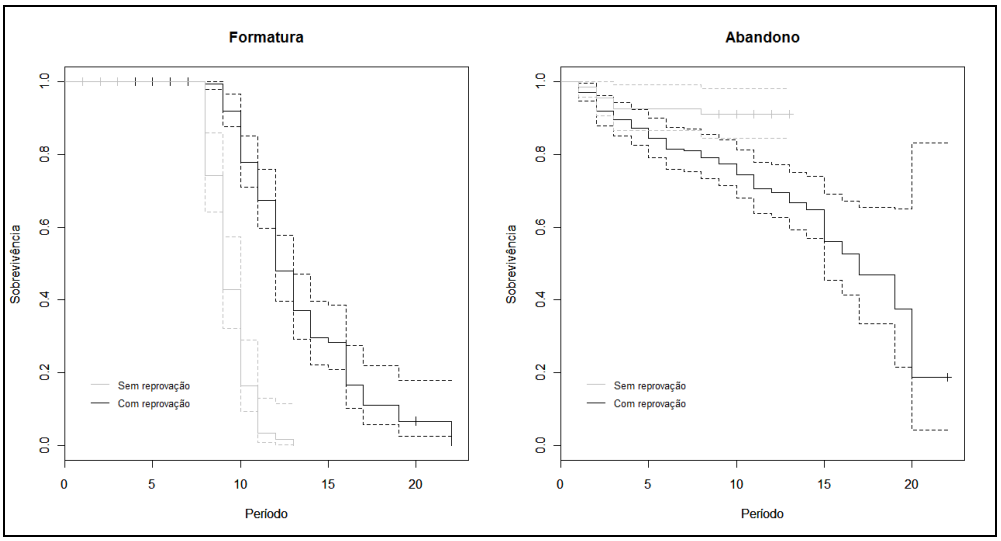

Figura 10 – Curvas de sobrevivência à formatura e ao abandono estratificadas por reprovação

Na tabela 3 apresenta-se um resumo do efeito gerado por cada covariável estudada sobre o risco de um aluno do curso em questão ficar retido ou vir a evadir-se. Entretanto, o método empregado não permitiu realizar inferências sobre o efeito das dispensas de carga horária sobre a evasão e retenção prolongada, bem como o da matrícula-vínculo e do desempenho acadêmico sobre a retenção prolongada.

Tabela 3 – Resumo dos riscos relativos de retenção excessiva e evasão para cada covariável

Variáveis	Retenção Excessiva		Evasão	
	Maior	Menor	Maior	Menor
gênero	Masculino	Feminino	Masculino	Feminino
dispensa	----	----	----	----
trancamento	Com	Sem	Com	Sem
vínculo	----	----	Com	Sem
desempenho	----	----	Baixo	Alto
reprovação	Com	Sem	Com	Sem

5. Conclusões

A análise de sobrevivência demonstrou ser uma técnica muito simples de ser usada, bem como uma poderosa ferramenta de predição da capacidade do aluno se formar ou abandonar um curso.

Quanto maior for a base de dados e os diversos tipos de informação, mais precisa será essa ferramenta. Se poucas informações forem fornecidas, a precisão será incerta, ou, em alguns casos, não ser factível, em particular devido ao grande número de censuras.

Ao longo do acompanhamento de uma turma, ainda faltará investigar a partir de que período será possível analisar os dados de sua evolução acadêmica com a técnica apresentada de análise de sobrevivência. Sabe-se ao menos que parte dos indivíduos já deverão ter experimentado a ocorrência de um dos eventos de interesse. Com isso, embora o acompanhamento do abandono possa ser realizado desde o primeiro período, o da formatura só poderá ser iniciado após ultrapassado o tempo mínimo para integralização curricular.

Salienta-se que o estágio em que as turmas se encontravam favoreceu o cálculo das funções de sobrevivência e termos correlatos. Esta fase da evolução também deverá constituir uma exigência se forem admitidos dados provenientes de uma única turma, e não todos os de um mesmo perfil curricular, como aconteceu no presente trabalho.

Outras variáveis ficaram de fora, pois ocorreram com uma frequência muito baixa, como a mobilidade estudantil, dificultada pela excessiva carga horária em relação ao tempo disponível para conclusão do curso, ou então não eram de fácil acesso, como, por exemplo, a participação em atividades de extensão, pesquisa ou de monitoria, principalmente por não serem regularmente coletadas no sistema de gestão acadêmica da universidade ou apresentarem acesso restrito.

A confirmação de que os riscos se mostraram proporcionais para boa parte das covariáveis admitidas sugere a construção futura de um modelo de Cox que incorpore a transformação ou a composição das covariáveis originais. Sugere-se também o uso de outros métodos paramétricos para estimação dos índices e das funções, além dos testes estatísticos de hipóteses empregados neste artigo

6. Agradecimentos

Os autores agradecem à UFPE por ter apoiado a condução desta pesquisa, por ter concedido bolsa de Iniciação Científica e cedido instalações e recursos físicos para esta finalidade.

Referências

Alencar, F. A. G. *Análise do Programa Bolsa Família sob a Ótica da Redução dos Fatores de Risco de Falhas na Freqüência Escolar*. Brasília: ESAF, 2007. Monografia agraciada com menção honrosa no XII Prêmio Tesouro Nacional – 2007. Qualidade do Gasto Público. Eusébio (CE).

Cislaghi, R. *Um modelo de sistema de gestão do conhecimento em um framework para a promoção da permanência discente no ensino de graduação.* Tese de Doutorado. Universidade Federal de Santa Catarina. Florianópolis, Maio, 2008.

Colosimo, E. A.; Giolo, S. R. *Análise de sobrevivência aplicada*. São Paulo: Edgard Blucher, 2006.

Corrêa, A. C. C. e Noronha, A. B. *Avaliação da Evasão e Permanência Prolongada em um Curso de Graduação em Administração de uma Universidade Pública.* Disponível em: <http://www.ead.fea.usp.br/ >. Acesso: 15/04/2011

Cunha, A. M.; Tunes, E.; Silva, R. R. Evasão do Curso de Química da Universidade de Brasília: A Interpretação do Aluno Evadido. *Quim. Nova*, v. 24, n. 1, p.262-280, 2001.

Gonçalves, M. E. *Análise de sobrevivência e modelos hierárquicos logísticos longitudinais: uma aplicação à análise da trajetória escolar (4ª a 8ª série - ensino fundamental).* Belo Horizonte: UFMG/Cedeplar, 2008.

Klein, J. P. e Moeschberger, M. L. *Survival Analysis: techniques for censored and truncated data*. 2nd ed. New York: Springer, 1983.

Lima, V. M. *Percepções de estudantes de primeiro período sobre o serviço educacional: análise empírica de uma IES privada na cidade do Rio de Janeiro.* Rio de Janeiro: Faculdades Ibmec. 2006.

Machado, S. P.; Melo Filho, J. M.; Pinto; A. C. A evasão nos cursos de graduação de química: uma experiência de sucesso feita no Instituto de Química da Universidade Federal do Rio de Janeiro para diminuir a evasão. *Quim. Nova*, v. 28, Suplemento, S41-S43, 2005.

Silva Fo., R. L. L.; Motejunas, P. R.; Hipólito, O.; Lobo, M. B. C. M.. A evasão no ensino superior brasileiro. Instituto Lobo para o Desenvolvimento da Educação, da Ciência e da Tecnologia. *Cadernos de Pesquisa.* Fundação Carlos Chagas. v. 37, n. 132, 2007.

The R Foundation. *The R Project for statistical computing.* Disponível em: < http://www.r-project.org/>. Acesso: 15/04/2011