



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

Bruno Rosevics
Estevão Batista do Prado

**ESTUDO SOBRE A EVASÃO E O TEMPO ATÉ A FORMATURA DOS
ALUNOS DO CURSO DE ESTATÍSTICA DA UFPR**

CURITIBA
2013

**Bruno Rosevics
Estevão Batista do Prado**

**ESTUDO SOBRE A EVASÃO E O TEMPO ATÉ A FORMATURA DOS
ALUNOS DO CURSO DE ESTATÍSTICA DA UFPR**

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**

RESUMO

Estudo sobre a Evasão e o Tempo até a Formatura dos Alunos do Curso de Estatística da UFPR

Nesse trabalho, foram analisados os dados dos alunos que ingressaram no Curso de graduação em Estatística da Universidade Federal do Paraná (UFPR) no período de 1991 a 2010. Como dois currículos nortearam o curso neste período (1991 e 2006), análise foi realizada para cada um deles. O currículo 1991 compreende alunos ingressos entre 1991 e 2005 (período anterior à implantação do Processo Seletivo Estendido (PSE)) e o currículo 2006, os ingressos entre 2006 e 2010, já com o PSE. O processo seletivo estendido compreende, além das duas fases que compõem o vestibular da UFPR, uma terceira fase em que o aluno é avaliado durante um semestre letivo em duas disciplinas. Tal processo tem como um de seus objetivos a redução da taxa de evasão. O objetivo desse trabalho foi analisar características que estão associadas à evasão e ao tempo até a formatura dos alunos mencionados. Para isso, foi utilizado o modelo de mistura com fração de cura. De modo geral, o modelo mostrou ajuste satisfatório para ambos os currículos, tendo como principais variáveis para o componente de sobrevivência o IRA (Índice de Rendimento Acadêmico) e o tipo de ingresso e, para o componente logístico, o IRA e a idade.

Palavras-chave: Evasão. Estatística. Análise de sobrevivência. Regressão logística. Modelo de mistura. Fração de cura.

LISTA DE FIGURAS

Figura 1 - Quantidade de ingressos, evadidos e formados por ano para o currículo de 1991	12
Figura 2 - Proporção de alunos formados e evadidos para o currículo de 1991	13
Figura 3 - Quantidade de ingressos, evadidos e formados por ano para o currículo de 2006	13
Figura 4 - Melhor e pior cenário para taxa de graduados no currículo de 2006	14
Figura 5 - Proporção para melhor e pior cenário para taxa de graduados no currículo de 2006	15
Figura 6 - Curvas de sobrevivência obtidas pelo estimador de Kaplan-Meier para as categorias da variável forma de ingresso na universidade.....	16
Figura 7 - Curvas de sobrevivência obtidas pelo estimador de Kaplan-Meier para as categorias da variável Índice de Rendimento Acadêmico (IRA).....	16
Figura 8 - Curvas de sobrevivência obtidas pelo estimador de Kaplan-Meier para as categorias da variável Idade.....	17
Figura 9 - Curva ROC associada ao componente logístico do modelo de mistura com fração de evasão ajustado aos dados do currículo de 1991.	19
Figura 10 - Análise dos resíduos de Pearson e deviance associados ao componente logístico do modelo de mistura ajustado aos dados do currículo de 1991	20
Figura 11 - Curvas de sobrevivência observada e estimada pelo modelo de fração de evasão ajustado para (a) $S(t U=1,x)$ e (b) $S(t x,z)$ para uma das combinações das variáveis IRA e forma de ingresso na universidade para currículo 1991.....	22
Figura 12 - Curva ROC associada ao componente logístico do modelo de mistura com fração de evasão ajustado aos dados do currículo de 2006.....	22
Figura 13 - Análise dos resíduos de Pearson e deviance associados ao componente logístico do modelo de mistura ajustado aos dados do currículo de 2006.	22
Figura 16 - Curvas de sobrevivência observada e estimada pelo modelo de fração de evasão ajustado para (a) $S(t U=1,x)$ e (b) $S(t x,z)$ para uma das combinações das variáveis IRA, gênero e forma de ingresso na universidade para currículo 2006.....	25

LISTA DE TABELAS

Tabela 1 - Frequências de alunos em 31/12/2013 de acordo com o <i>status</i> e currículo	9
Tabela 2 - Análise descritiva da variável gênero para o currículo 1991 (ingressos 1991-2005)	10
Tabela 3 - Análise descritiva da variável gênero para o currículo 2006 (ingressos 2006-2010)	10
Tabela 4 - Análise descritiva da variável forma de ingresso no curso para o currículo 1991	10
Tabela 5 - Análise descritiva da variável forma de ingresso no curso para o currículo 2006	10
Tabela 6 - Análise descritiva da variável IRA para o currículo 1991	11
Tabela 7 - Análise descritiva da variável IRA para o currículo 2006	11
Tabela 8 - Análise descritiva da variável idade para o currículo 1991	11
Tabela 9 - Análise descritiva da variável idade para o currículo 2006	12
Tabela 10 - Estimativas associadas ao componente logístico $\pi(z)$ – Currículo 1991	18
Tabela 11 - Estatísticas associadas ao componente logístico – Currículo 1991	18
Tabela 12 - Estimativas associadas ao componente de sobrevivência do modelo de mistura ajustado aos dados do Currículo de 1991	21
Tabela 13 - Estatísticas associadas ao modelo escolhido (componente de sobrevivência) – Currículo de 1991	21
Tabela 14 - Correlação estatística entre as funções de sobrevivência estimadas e observadas (Kaplan-Meier) para cada uma das combinações da variáveis IRA e forma de ingresso na universidade para o currículo 1991	23
Tabela 15 - Estimativas obtidas com o modelo de mistura com fração de evasão ao longo do tempo, para cada combinação das variáveis do currículo 1991	25
Tabela 16 - Estimativas associadas ao componente logístico $\pi(z)$ – currículo 2006	21
Tabela 17 - Estatísticas associadas ao componente logístico – currículo 2006	21
Tabela 18 - Estimativas associadas ao componente de sobrevivência do modelo de mistura ajustado aos dados do Currículo de 2006.	24
Tabela 19 - Estatísticas associadas ao modelo escolhido (componente de sobrevivência) – currículo de 2006	24
Tabela 20 - Correlação estatística entre as funções de sobrevivência estimadas e observadas (Kaplan-Meier) para cada uma das combinações da variáveis IRA e forma de ingresso na universidade para o currículo 2006	26
Tabela 21 - Estimativas obtidas com o modelo de mistura com fração de evasão ao longo do tempo, para cada combinação das variáveis do currículo 2006	28
Tabela 22 - Estimativas obtidas com o modelo de mistura com fração de evasão ao longo do tempo, para cada combinação das variáveis do currículo 1991 e 2006.	30

Sumário

1 INTRODUÇÃO	1
2 REVISÃO DE LITERATURA	3
3 MATERIAL E MÉTODOS	5
3.1 Material.....	5
2.1.2 Recursos Computacionais.....	5
3.2 Métodos.....	6
3.2.1 Análise de sobrevivência.....	6
3.2.2 Estimador Não-Paramétrico de Kaplan-Meier	6
3.2.4 Modelo de Mistura com Fração de Cura	7
4 RESULTADOS	9
4.1 Análise Descritiva.....	9
4.2 Ajustes de Modelos de Sobrevivência.....	15
4.2.1 Análise Exploratória: Kaplan-Meier	15
4.2.2 Modelo de Mistura com Fração de Evasão para o Currículo 1991.....	17
4.2.3 Modelo de Mistura com Fração de Evasão para o Currículo 2006.....	20
4.2.4 Comparação.....	29
5 CONCLUSÃO	28
REFERÊNCIAS.	30
APÊNDICES	31

1 INTRODUÇÃO

A evasão escolar é um problema que tem atingido instituições de ensino em todos os níveis da educação, da básica à superior. Com isso, muitos trabalhos e pesquisas têm sido desenvolvidos no intuito de entender as causas que levam a esse evento.

A perda de estudantes que iniciam, mas não terminam seus cursos, produz desperdícios sociais, acadêmicos e econômicos. No setor público, são recursos públicos investidos sem o devido retorno. No setor privado, é uma importante perda de receitas. Em ambos os casos, a evasão é uma fonte de ociosidade de professores, funcionários, equipamentos e espaço físico (LOBO *et al.*, 2007).

Segundo o Instituto Nacional de Estudos e Pesquisas Anísio Teixeira (2006), muitas instituições criaram cursos de nivelamento, com o intuito de amenizar a taxa de evasão dos cursos de graduação (INEP, 2006). Nesse contexto, em 2006, os Cursos de graduação em Estatística, Matemática e Matemática Industrial da Universidade Federal do Paraná (UFPR) incorporaram, além das duas fases que compõem o vestibular da UFPR, uma terceira fase. Nesta terceira fase, os alunos convocados cursam duas disciplinas ao longo do primeiro semestre letivo do ano de ingresso, sendo submetidos a quatro provas em cada uma das disciplinas. Para os candidatos ao Curso de Estatística, as disciplinas previstas são: Fundamentos Básicos para Estatística e Estatística Descritiva e Inferencial Básica.

Por ser composto de uma fase adicional aos dos demais cursos, este processo de seleção é denominado Processo Seletivo Estendido (PSE). Ao final do PSE, o cálculo do desempenho final de cada candidato é realizado pelo Núcleo de Concursos (NC) da UFPR com base nos pontos por ele obtidos na primeira, segunda e terceira fases, bem como na prova objetiva do ENEM.

Historicamente, os Cursos de graduação em Estatística no Brasil apresentam altas taxas de evasão. Assim, um interesse quanto ao PSE, diz respeito à avaliação de seu impacto na redução da taxa de evasão.

Para contribuir com tal avaliação, a proposta deste trabalho foi a de dar continuidade ao estudo publicado por Martins e Rocha (2011), realizando análises estatísticas adicionais à desses autores com o intuito de entender o perfil dos alunos formados e evadidos do Curso de graduação em Estatística da Universidade Federal do Paraná. Para isso, foram considerados os alunos que ingressaram no referido

curso entre os anos de 1991 e 2005 (período anterior ao PSE), bem como entre os anos de 2006 e 2010 (período em que o PSE já estava vigente). O *status* desses alunos (formado, evadido, cursando, etc.), foi o que constava na data de 31 de dezembro de 2013 no Sistema de Informações para o Ensino (SIE) da UFPR.

Para efeitos de comparação é importante ressaltar que existem poucos trabalhos publicados com o mesmo intuito deste. Um dos trabalhos foi sobre o PSE no Curso de Matemática da UFPR, conduzido por Souza (2008). Neste estudo, contudo, não foram apresentadas análises estatísticas; mas sim a parte conceitual e a percepção dos alunos, professores e Coordenadores de Curso sobre o PSE.

2 REVISÃO DE LITERATURA

Este capítulo traz um breve relato de alguns estudos realizados para avaliar a evasão dos Cursos de Estatística de universidades públicas brasileiras.

Com dito anteriormente, o objetivo deste trabalho foi o de dar continuidade ao estudo realizado por Martins e Rocha (2011). Em tal estudo, os autores analisaram, separadamente, por meio de técnicas de análise de sobrevivência, as características que estariam associadas à evasão e ao tempo de permanência dos alunos evadidos e não evadidos do Curso de Estatística da UFPR. Um dos objetivos do estudo foi avaliar o Processo Seletivo Estendido (PSE), o qual foi implantado em 2006 com a proposta de reduzir a evasão. Com base nos resultados obtidos sobre a evasão de alunos que ingressaram no Curso de Estatística da UFPR de 1991 a 2005 (antes do PSE e tempo mínimo de integralização de 4,5 anos) e de 2006 a 2011 (já com o PSE e tempo mínimo de integralização de 4 anos), os autores concluíram haver indícios de diminuição na taxa de evasão, bem como destacaram uma tendência de diminuição no número de alunos ingressos.

Em outro estudo realizado por Fonseca *et al.* (2013), foi também analisada a evasão no período de 1997 a 2004 de alunos do Curso de Estatística da Universidade Federal do Rio Grande do Norte (UFRN). Nesse estudo, foi aplicado um questionário aos alunos com 5 questões relacionadas às suas condições de vida. Para as análises realizadas foram consideradas as variáveis: gênero, idade, meio de transporte (público ou veículo próprio), grau de escolaridade da mãe (superior ou maior ou até nível médio completo) e moradia (vive com os pais ou não). Com os resultados obtidos, os autores apresentaram os perfis dos alunos com maior e menor chance de evadir do Curso de Estatística da UFRN. De forma geral, alunos do gênero feminino, cujo grau de escolaridade da mãe correspondia ao ensino superior ou maior, foram os que apresentaram menor taxa de evasão. Por outro lado, alunos do gênero masculino, cujo grau de escolaridade da mãe era até o nível médio completo, foram os que apresentaram maior taxa de evasão.

Ainda, no estudo realizado por Cavalcanti *et al.* (2012) foi analisado o tempo até a evasão dos alunos do Curso de Estatística da Universidade Federal da Paraíba (UFPB). Neste estudo, os autores tinham à disposição uma base de dados com informações sobre 132 alunos com as seguintes variáveis disponíveis: gênero, procedência (João Pessoa ou outras cidades da Paraíba), forma de ingresso

(vestibular ou outras), raça (branca ou outras) e instituição em que cursou o ensino básico (pública, particular ou ambas). De modo geral, alunos de outras cidades da Paraíba, que ingressaram por outras formas de ingresso (transferência, etc.), e que se autodeclararam de outra raça (que não branca) apresentaram menor tempo até a evasão. Analogamente, alunos procedentes da cidade de João Pessoa, que ingressaram por vestibular e de raça branca, apresentaram maior tempo até a evasão.

Biurrum e Nunes (2010) também realizaram uma pesquisa com alunos evadidos do Curso de Estatística da UFRGS que compreendeu o período de 2000 a 2007. O objetivo dos autores foi traçar o perfil do aluno evadido. Para isso, os autores realizaram amostragem por conveniência e desenvolveram um questionário com 30 perguntas. O questionário foi enviado (por e-mail) para 145 alunos evadidos, de um total de 186, obtendo-se 43 respondentes. Ao final do trabalho, os autores elencaram as principais causas de evasão apontadas pelos alunos evadidos. São elas: *i)* colisão de horários das disciplinas com o trabalho, *ii)* dificuldades nas disciplinas e, *iii)* não terem gostado do curso. Quando perguntados sobre um possível retorno ao Curso de Estatística da UFRGS, 31,3% responderam que não voltariam. Entre os que responderam que voltariam, 77,3% responderam que voltariam se o curso fosse oferecido no período noturno.

3 MATERIAL E MÉTODOS

3.1 Material

O banco de dados utilizado neste estudo contém informações referentes aos alunos que ingressaram no Curso de Estatística da UFPR desde 1991 até 2010. Neste período, dois currículos nortearam este curso, sendo que o primeiro vigorou de 1991 a 2005 e o segundo de 2006 a 2010 (este segundo coincidindo com o início do PSE).

A decisão em considerar 2010 como o último ano de ingresso para o estudo proposto neste trabalho se justifica pelo fato de um novo currículo ter entrado em vigor no Curso de Estatística em 2011, de modo que não haveria tempo hábil até 31/12/2013 para conclusão do curso entre os alunos que ingressaram após 2010.

No geral, o banco de dados contém informações de aproximadamente 1600 alunos, sendo que 1259 ingressaram sob o currículo de 1991 (1991 a 2005) e 297 sob o currículo de 2006 (2006 a 2010).

Dentre as informações (variáveis) disponíveis no banco de dados, tem-se: data de ingresso no curso; forma de ingresso no curso (vestibular, transferência, ou outros); sexo do aluno; idade do aluno na data de ingresso no curso; IRA (Índice de Rendimento Acadêmico) do aluno e *status* do aluno (formado, cursando, etc.) em 31/12/2013.

O tempo desde o ingresso até a conclusão do curso (em semestres) foi considerado como variável resposta no componente de sobrevivência. Para aqueles alunos em que a conclusão do curso não havia ocorrido até 31/12/2013 (evadidos, cursando, etc.), o tempo considerado foi o decorrido desde a data de ingresso até 31/12/2013. Em análise de sobrevivência, esses tempos são denominados tempos censurados (COLOSIMO; GIOLO, 2006).

2.1.2 Recursos Computacionais

Para as análises e tratamento do banco de dados foram utilizados os recursos dos *softwares*: *Statistical Analysis System* (SAS INSTITUTE INC, 2012), versão 4.3 *SAS Enterprise Guide*, R (R DEVELOPMENT CORE TEAM, 2013) e *Microsoft Excel*, versão 2010.

3.2 Métodos

3.2.1 Análise de sobrevivência

Em análise de sobrevivência a variável resposta é o tempo até a ocorrência de um evento de interesse, usualmente denominado tempo de falha. A diferença entre a análise de sobrevivência e as técnicas estatísticas clássicas é a presença de censuras que ocorrem, por exemplo, devido a uma fração de indivíduos com a ausência do evento (falha) até a data final de acompanhamento (COLOSIMO; GIOLO, 2006). Ou seja, indivíduos que apresentam observações parciais da resposta são considerados censuras.

Nesse estudo, falha está associada ao término do Curso de graduação em Estatística; ou seja, falhas associam-se aos alunos que se formaram. Por outro lado, as censuras referem-se àqueles que ainda não se formaram, ou que nunca irão se formar, pois desistiram ou mudaram de curso, entre outros. Logo, a variável indicadora de falha é assim definida, no contexto desse trabalho,

$$\delta_i = \begin{cases} 1 & \text{se o aluno } i \text{ se formou} \\ 0 & \text{se o aluno } i \text{ não se formou,} \end{cases} \quad i = 1, \dots, n.$$

3.2.2 Estimador Não-Paramétrico de Kaplan-Meier

Em um contexto exploratório, o estimador não-paramétrico de Kaplan-Meier (KAPLAN; MEIER, 1958) foi utilizado para estimar as curvas de sobrevivência, $S(t) = P(T > t)$ associadas às categorias de cada covariável disponível, com o intuito de avaliar possíveis associações destas covariáveis com a variável resposta. T denota a variável aleatória não-negativa tempo até o evento.

A expressão deste estimador, também conhecido como estimador produto-limite, é dada por:

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

em que, t_j ($j = 1, \dots, k$) são os k tempos distintos e ordenados de falha, d_j é o número de falhas em t_j , para $j = 1, \dots, k$, e n_j é o número de indivíduos sob risco em t_j .

Para testar se as curvas de sobrevivência diferem entre as categorias de cada covariável avaliada foi utilizado o teste *logrank* (MANTEL, 1966).

3.2.4 Modelo de Mistura com Fração de Cura

O uso de modelos de mistura com fração de cura tem sido proposto na literatura quando se deseja analisar dados de indivíduos que são acompanhados por um longo período de tempo, observando-se, no final do estudo, que parte deles não experimentou o evento de interesse. Tais indivíduos são denominados sobreviventes de longa duração, ou curados, ou simplesmente indivíduos imunes ao evento (EUDES *et al.*, 2012).

Modelos de fração de cura assumem que a população de interesse é dividida em duas subpopulações (uma composta de indivíduos suscetíveis, que podem experimentar o evento de interesse, e outra composta de indivíduos não suscetíveis, ou simplesmente curados, que nunca irão experimentar o evento de interesse). Dessa forma, a estimação dos parâmetros é feita simultaneamente para os indivíduos suscetíveis e não suscetíveis (CORBIÈRE; JOLY, 2007).

Para formulação do modelo, considere U uma variável que indica se o indivíduo é suscetível ($U = 1$) ou não suscetível ($U = 0$) ao evento de interesse, e T uma variável aleatória não negativa que denota o tempo até o evento de interesse, definida somente para indivíduos suscetíveis ($U = 1$). Logo, o modelo de fração, em termos da função de sobrevivência, é dado por:

$$S(t | \mathbf{x}, \mathbf{z}) = \pi(\mathbf{z})S(t | U = 1, \mathbf{x}) + 1 - \pi(\mathbf{z}),$$

sendo $S(t | \mathbf{x}, \mathbf{z})$ a função de sobrevivência para toda a população, $\pi(\mathbf{z})$ a probabilidade do indivíduo ser suscetível, ou simplesmente, a probabilidade do indivíduo apresentar o evento de interesse ($U = 1$), dado o conjunto de variáveis explicativas \mathbf{z} , e $S(t | U = 1, \mathbf{x})$ a função de sobrevivência para os indivíduos suscetíveis, dado o vetor de variáveis \mathbf{x} (CORBIÈRE; JOLY, 2007).

A probabilidade $\pi(\mathbf{z})$ pode ser modelada utilizando modelos de regressão para dados binários com função de ligação logito, probito ou complemento log-log, dentre outras. Neste trabalho, optou-se por utilizar o logito, de modo que:

$$\pi(\mathbf{z}) = \frac{\exp\{\boldsymbol{\beta}'\mathbf{z}\}}{1 + \exp\{\boldsymbol{\beta}'\mathbf{z}\}} = \frac{\exp(\beta_0 + \sum_{k=1}^p \beta_k z_k)}{1 + \exp(\beta_0 + \sum_{k=1}^p \beta_k z_k)}.$$

Com relação à função $S(t | U = 1, \mathbf{x})$, esta pode ser modelada via modelos paramétricos (exponencial, Weibull, lognormal, etc.) ou pelo modelo semiparamétrico

de Cox. Neste trabalho, optou-se pelo modelo de Cox, em que se tem:

$$S(t | U = 1, \mathbf{x}) = S_0(t | U = 1)^{\exp(\boldsymbol{\gamma}'\mathbf{x})},$$

sendo que $S_0(t | U = 1)$ é a função de sobrevivência de base e $\boldsymbol{\gamma}$ o vetor de parâmetros associadas às covariáveis \mathbf{x} .

A função de verossimilhança associada ao modelo de fração de cura é dada por:

$$L(\boldsymbol{\gamma}, \boldsymbol{\beta}) = \prod_{i=1}^n \{\pi(\mathbf{z}_i) f(t_i | U = 1, \mathbf{x}_i)\}^{\delta_i} \{1 - \pi(\mathbf{z}_i) + \pi(\mathbf{z}_i) S(t_i | U = 1, \mathbf{x}_i)\}^{1-\delta_i},$$

em que δ_i é o indicador de falha (1 se falha e 0 se censura) e t_i é o tempo até a ocorrência da falha (CORBIÈRE; JOLY, 2007).

A estimação dos vetores de parâmetros $\boldsymbol{\gamma}$ e $\boldsymbol{\beta}$, bem como para $S_0(t | U = 1)$ é obtida via método EM (do inglês, *estimation and maximization*), que maximiza a função de verossimilhança. Neste trabalho, para obter tal estimação dos parâmetros, foi utilizado a macro PSPMCM desenvolvida por Corbière e Joly (2007).

Para a seleção de covariáveis do modelo foi utilizado uma mistura do método *stepwise* em ambas as direções (para inclusão e exclusão), testes com variáveis selecionadas manualmente e o Critério de Akaike (AIC). Para o componente logístico apenas as variáveis que apresentaram nível de significância menor ou igual a 0,10 foram mantidas no modelo final, enquanto que para o componente de sobrevivência fixou-se 0,20. Foram fixados níveis de significância diferentes porque para a estimação dos parâmetros do componente logístico considera-se toda a amostra, enquanto que para o componente de sobrevivência somente a amostra de indivíduos que apresentaram o evento de interesse. Desta forma, o número de observações em cada componente varia.

Para verificar a qualidade de ajuste do modelo, foram utilizados o coeficiente de correlação de Pearson e o R^2 (coeficiente de correlação de Pearson ao quadrado). O coeficiente de Pearson mede a correlação entre as curvas de sobrevivência obtidas via método Kaplan-Meier e pelo modelo ajustado para todas as combinações das variáveis. Coeficientes próximos de 1 (um) indicam ajuste satisfatório do modelo.

4 RESULTADOS

4.1 Análise Descritiva

Inicialmente as variáveis disponíveis para cada um dos dois bancos de dados (currículos 1991 e 2006) foram analisadas a fim de verificar a existência de tendência no comportamento das mesmas. Neste contexto, a Tabela 1 apresenta as frequências absolutas e relativas de alunos formados, evadidos e cursando na data de 31/12/2013.

Tabela 1 - Frequências de alunos em 31/12/2013 de acordo com o *status* e currículo

Currículo	Cursando	Evadidos	Formados	Totais
1991	0 (0%)	911 (72,4%)	348 (27,6%)	1259 (100%)
2006	54 (22,2%)	132 (54,3%)	111 (45,7%)	297 (100%)

Fonte: Os autores (2013).

A partir da Tabela 1, pode-se verificar que o currículo de 2006 apresenta em 31/12/2013 uma taxa de alunos formados maior do que a do currículo anterior. Se levado em consideração que no currículo 2006 há 22,2% de alunos ainda cursando, pode-se esperar um aumento na taxa de formados, que pode chegar a 67,9%, caso todos consigam terminar sua graduação ou, no pior cenário, ficar em 45,7%, caso todos eles venham a evadir.

Quanto à variável gênero, é possível observar a partir das Tabelas 2 e 3 que alunos ingressos sob o currículo de 1991 (currículo com maior quantidade de alunos) apresentam uma diferença significativa entre os gêneros (número maior de ingressos do gênero masculino, 64% *versus* 36%), com as mulheres, contudo, apresentando taxa de conclusão maior que a dos homens. Já para os alunos do currículo de 2006, observam-se taxas de conclusão em 31/12/2013 maiores para ambos os sexos.

Ressalta-se, contudo, que apesar de uma taxa maior de conclusão ter sido observada nesta data entre as mulheres (45% *versus* 32%), há mais homens com *status* "cursando", o que pode vir a alterar esse resultado. De modo geral, os alunos cursando podem ainda influenciar consideravelmente os percentuais deste currículo.

Tabela 2 - Análise descritiva da variável gênero para o currículo 1991 (ingressos 1991-2005)

Gênero	Frequências Absolutas		Percentuais (%)		Totais	
	Evadidos	Formados	Evadidos	Formados	#	%
Feminino	301	155	66%	34%	456	36%
Masculino	610	193	76%	24%	803	64%

Fonte: Os autores (2013).

Tabela 3 - Análise descritiva da variável gênero para o currículo 2006 (ingressos 2006-2010)

Gênero	Frequências Absolutas			Percentuais %			Totais	
	Cursando	Evadidos	Formados	Cursando	Evadidos	Formados	#	%
Feminino	21	46	54	17%	38%	45%	121	41%
Masculino	33	86	57	19%	49%	32%	176	59%

Fonte: Os autores (2013).

No que diz respeito às formas de ingresso na UFPR, outras formas de ingresso (Tabelas 4 e 5) apresentou, em ambos os currículos, taxa de conclusão muito menor quando comparada com a do vestibular.

Tabela 4 - Análise descritiva da variável forma de ingresso no curso para o currículo 1991

Forma de Ingresso	Frequências Absolutas		Percentuais %		Totais	
	Evadidos	Formados	Evadidos	Formados	#	%
Outros	200	26	88%	12%	226	18%
Vestibular	711	322	69%	31%	1033	82%

Fonte: Os autores (2013).

Tabela 5 - Análise descritiva da variável forma de ingresso no curso para o currículo 2006

Forma de Ingresso	Frequências Absolutas			Percentuais %			Totais	
	Cursando	Evadidos	Formados	Cursando	Evadidos	Formados	#	%
Outros	2	39	9	4%	78%	18%	50	17%
Vestibular	52	93	102	21%	38%	41%	247	83%

Fonte: Os autores (2013).

Quanto a variável IRA (Índice de Rendimento Acadêmico), pode-se notar a partir das Tabelas 6 e 7, uma grande quantidade de alunos evadidos com IRA abaixo de 50 em ambos os currículos. Para as demais classes, a taxa de conclusão aumenta com o aumento do IRA. Para as duas classes de IRA acima de 70, as diferenças entre as taxas de conclusão não diferem muito entre os currículos.

Porém, o tempo até a conclusão do curso entre os alunos com IRA acima de 85 é menor do que o do grupo de alunos com IRA acima de 70 conforme será detalhado em capítulo futuro sobre análise de sobrevivência.

Tabela 6 - Análise descritiva da variável IRA para o currículo 1991

IRA	Frequências		%		Total Geral	
	Evadidos	Formados	Evadidos	Formados	#	%
0 – 50	795	15	98%	2%	810	64%
50 – 70	91	151	38%	62%	242	19%
70 – 85	22	146	13%	87%	168	13%
85 – 100	3	36	8%	92%	39	3%

Fonte: Os autores (2013).

Tabela 7 - Análise descritiva da variável IRA para o currículo 2006

IRA	Frequências			%			Total Geral	
	Cursando	Evadidos	Formados	Cursando	Evadidos	Formados	#	%
0 – 50	29	104	8	21%	74%	6%	141	47%
50 – 70	20	21	44	24%	25%	52%	85	29%
70 – 85	4	7	51	6%	11%	82%	62	21%
85 – 100	1	0	8	11%	0%	89%	9	3%

Fonte: Os autores (2013).

Como esperado, o grupo de alunos com menor IRA em ambos os currículos apresenta uma taxa baixa de formados. Contudo, no currículo 2006 há ainda uma quantidade significativa de alunos cursando, podendo ocorrer, inclusive para esta classe de alunos (IRA < 50), uma alteração significativa ao fim do histórico dos mesmos na universidade. Têm-se, também, taxas de conclusão elevadas entre os alunos com IRA maior, as quais também apresentam alterações devido aos alunos *cursando*.

Para a variável idade, a análise foi conduzida utilizando as mesmas faixas de idade do estudo realizado por Martins e Rocha (2011). Assim, de acordo com o apresentado nas Tabelas 8 e 9, pode-se verificar que conforme a idade dos alunos aumenta também aumenta a taxa de evadidos, porém com pouca diferença entre as faixas etárias.

Tabela 8 - Análise descritiva da variável idade para o currículo 1991

IDADE	#		%		Total Geral	
	Evadidos	Formados	Evadidos	Formados	#	%
17 - 19	187	112	63%	37%	299	24%
19 - 22	202	85	70%	30%	287	23%
22 - 26	199	73	73%	27%	272	22%
> 26 anos	323	78	81%	19%	401	32%

Fonte: Os autores (2013).

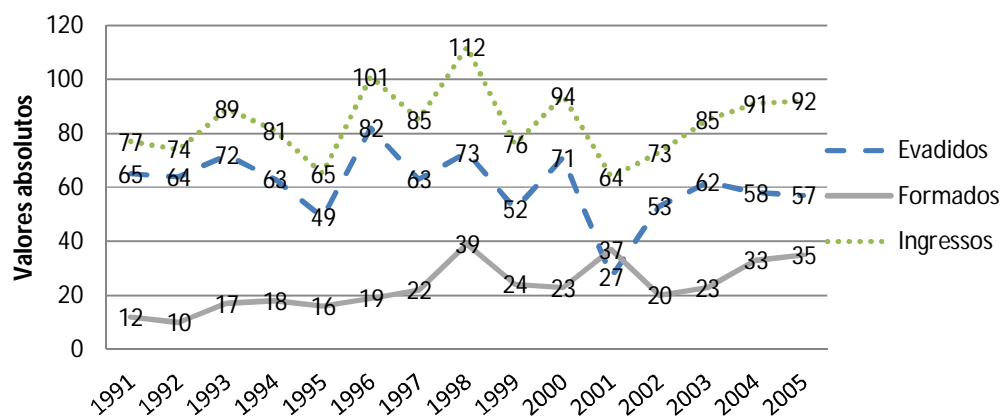
Tabela 9 - Análise descritiva da variável idade para o currículo 2006

IDADE	#			%			Total Geral	
	Cursando	Evadidos	Formados	Cursando	Evadidos	Formados	#	%
17 – 19	30	31	63	24%	25%	51%	124	42%
19 – 22	9	21	20	18%	42%	40%	50	17%
22 – 26	7	27	13	15%	57%	28%	47	16%
> 26 anos	8	53	15	11%	70%	20%	76	26%

Fonte: Os autores (2013).

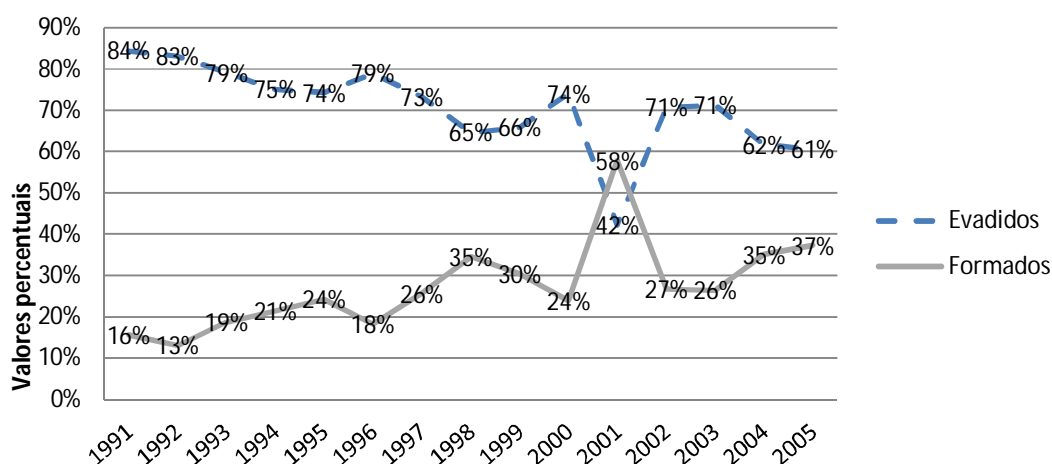
A seguir, é conduzida uma análise descritiva dos perfis de entrada, formados e evadidos por ano a fim de verificar se houve alguma mudança pontual na taxa de conclusão e/ou alguma tendência ao longo dos anos.

Desse modo, são apresentados na Figura 1 os perfis mencionados associados ao currículo de 1991, podendo-se notar uma tendência de aumento na quantidade de alunos formados, acompanhando o crescimento na quantidade de ingressos ao longo dos anos. Em 2001 ocorreu um cenário atípico, com mais alunos formados do que evadidos, fruto de um problema com o processo de vestibular que causou que alunos pudessem se beneficiar de artifícios legais para entrar no curso mesmo não se classificando, assim como uma queda na quantidade de alunos ingressos. Quando analisado em termos de proporções, nota-se (Figura 2) que a tendência de aumento na taxa de formados fica mais evidente.

Figura 1 - Quantidade de ingressos, evadidos e formados por ano para o currículo de 1991

Fonte: Os autores (2013).

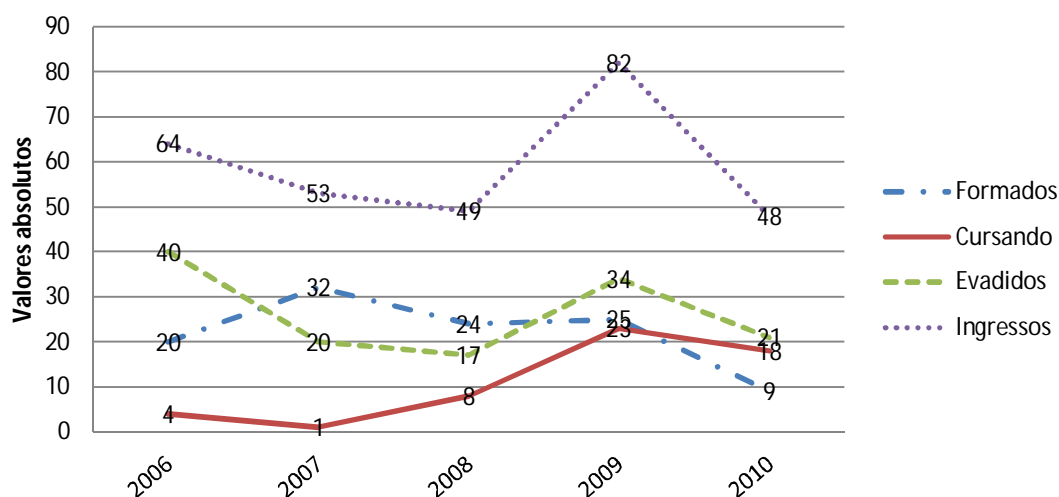
Figura 2 - Proporção de alunos formados e evadidos para o currículo de 1991



Fonte: Os Autores (2013).

A Figura 3 mostra o comportamento observado para os alunos ingressos no currículo de 2006. Neste caso vale ressaltar que as frequências e percentuais apresentados estão ainda sujeitos a mudanças, tendo em vista haver um total de 54 (22,2%) alunos com o curso em andamento (regularmente matriculados até 31/12/2013). A partir da figura mencionada, pode-se observar que o número de alunos que ingressaram por ano neste currículo decresceu quando comparado com o do currículo de 1991.

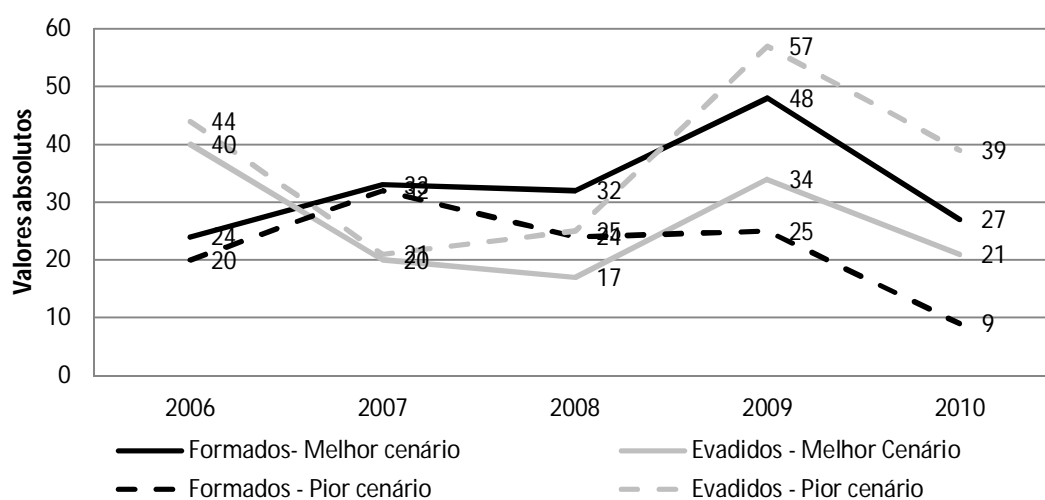
Figura 3 - Quantidade de ingressos, evadidos e formados por ano para o currículo de 2006



Fonte: Os autores

Uma observação importante quanto ao currículo de 2006 é que os percentuais de conclusão de curso entre os alunos que ingressaram em 2009 e 2010 (anos mais recentes) devem sofrer mudanças significativas, tendo em vista muitos alunos desses anos estarem com o curso ainda em andamento. Com o intuito de mostrar graficamente mudanças que podem ocorrer no cenário apresentado na Figura 3, são mostradas nas Figuras 4 e 5 duas projeções do cenário atual. Uma delas correspondendo ao melhor cenário (no qual todos os alunos atualmente cursando concluem o curso) e a outra ao pior cenário (no qual todos os alunos atualmente cursando se evadem). Com base nesses cenários é aconselhável, portanto, olhar com cautela os resultados atuais.

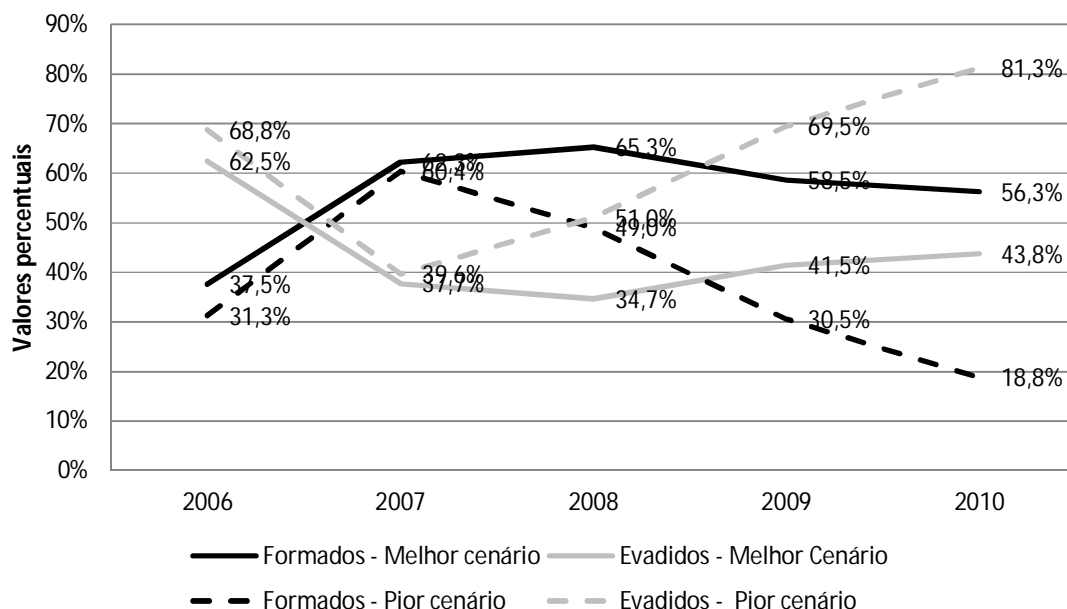
Figura 4 – Frequências para melhor e pior cenário de graduados no currículo de 2006



Fonte: Os autores (2013).

Conforme se pode observar para os anos mais recentes o efeito dos alunos ainda cursando traz maior incerteza às taxas acima demonstradas.

Figura 5 - Proporções para melhor e pior cenário de graduados no currículo de 2006



Fonte: Os autores (2013).

Novamente, quando analisados os dados pela proporção de cada status, tem-se uma visão que mostra a influência dos alunos ainda cursando de maneira mais explícita.

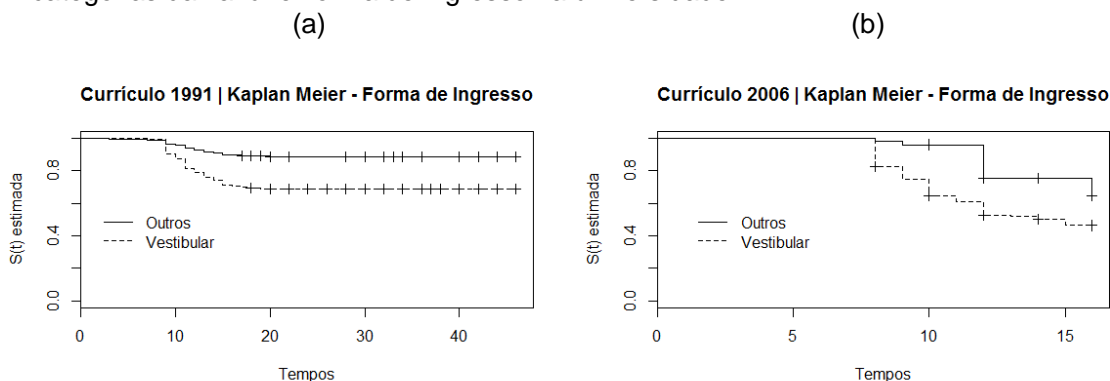
4.2 Ajustes de Modelos de Sobrevida

4.2.1 Análise Exploratória: Kaplan-Meier

Com o objetivo de analisar o tempo até a formatura dos 1259 alunos que ingressaram no Curso de Estatística da UFPR entre 1991 e 2005 (currículo 1991), foi realizada uma análise descritiva utilizando o estimador de Kaplan-Meier para cada uma das variáveis disponíveis. O intuito desta análise inicial foi investigar a existência de possíveis associações entre as variáveis explicativas e a variável resposta, auxiliando, assim, na seleção das variáveis a compor o modelo final.

Iniciando com a variável forma de ingresso na universidade, pôde-se observar, a partir da Figura 6 (gráficos a e b), indicativos de diferenças entre as curvas de sobrevivência associadas às categorias desta variável nos currículos 1991 e 2006. Alunos que entraram por vestibular apresentaram, de acordo com o teste *logrank* um tempo menor até a formatura do que aqueles que entraram por outras formas (LR = 33,4 e 12, p-valores < 0,0006, respectivamente).

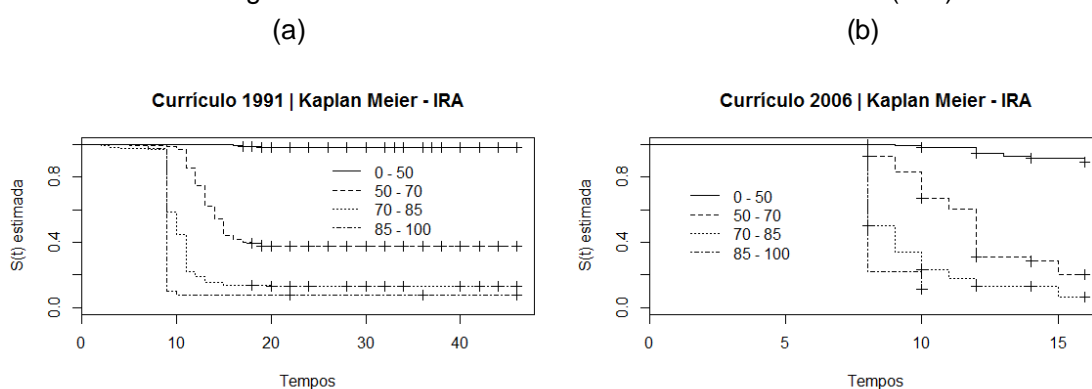
Figura 6 - Curvas de sobrevivência obtidas pelo estimador de Kaplan-Meier para as categorias da variável forma de ingresso na universidade



Fonte: Os autores (2013).

Quanto à variável IRA (Índice de Rendimento Acadêmico), observam-se a partir da Figura 7 (gráficos a e b), evidências de diferenças entre as curvas associadas às categorias desta variável para ambos os currículos (LR = 1311 e 197, p -valores $< 0,001$). É possível notar que alunos com IRA menor que 50 apresentam um tempo maior até a formatura do que àqueles com IRA maior. Ainda, à medida que o IRA aumenta ocorrem mais falhas (i.e., mais alunos se formam). Alunos com IRA entre 85 e 100, por exemplo, apresentam baixa probabilidade de sobrevivência (i.e., de não se formarem) em $t = 9$ semestres. A partir de $t = 10$ semestres, a probabilidade destes alunos não se formarem tende a zero.

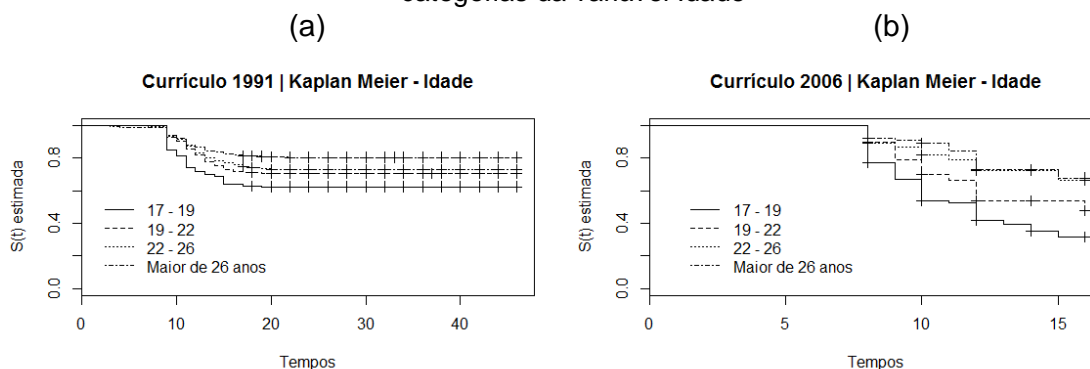
Figura 7 - Curvas de sobrevivência obtidas pelo estimador de Kaplan-Meier para as categorias da variável Índice de Rendimento Acadêmico (IRA)



Fonte: Os autores (2013).

Quanto à idade, observa-se, de acordo com a Figura 8 (gráficos a e b), que alunos com idade entre 17 e 19 anos no momento de ingresso no curso apresentam probabilidade menor de sobrevida, indicando que alunos nesta faixa etária que finalizam o curso, o fazem em menos semestres que os demais. Já entre os alunos nas faixas etárias 19-22 e 22-26 anos, observam-se probabilidades de sobrevida similares, porém superiores às dos alunos na faixa etária 17-19 anos. Ainda, alunos com idade superior a 26 anos apresentam as maiores probabilidades de sobrevida, indicando que alunos nesta faixa etária que finalizam o curso são os que demandam um número maior de semestres até a formatura. De acordo com o teste *logrank* há indicativos de diferenças entre as curvas de sobrevivência associadas às categorias desta variável nos currículos 1991 e 2006 (LR = 29,9 e 28,5, p-valores < 0,0001, respectivamente).

Figura 8 - Curvas de sobrevivência obtidas pelo estimador de Kaplan-Meier para as categorias da variável Idade



Fonte: Os autores (2013).

4.2.2 Modelo de Mistura com Fração de Evasão para o Currículo 1991

Observando as curvas de sobrevivência apresentadas na Seção 4.2, pode-se notar a existência de um número elevado de censuras à direita (i.e., alunos que se evadiram). No entanto, esse fato já era conhecido e, por isso, o modelo de mistura com fração de evasão descrito na Seção X foi escolhido para modelar o tempo até o término do curso de graduação em Estatística do currículo 1991.

Inicialmente, todas as variáveis foram incluídas simultaneamente no componente logístico $\pi(\mathbf{z})$ e no componente de sobrevivência $S(t | U = 1, \mathbf{x})$. Os critérios estabelecidos para permanência ou inclusão de variáveis foram $\alpha = 0,10$

para o componente logístico e 0,20 para o componente de sobrevivência. Logo, variáveis com p-valores superiores aos critérios estabelecidos foram excluídas. As variáveis gênero e tipo de ingresso foram removidas do componente logístico por terem apresentado efeito não significativo (p-valores iguais a 0,9323 e 0,1761, respectivamente). Já no componente de sobrevivência, as variáveis gênero e idade não apresentaram efeitos significativos (p-valores iguais a 0,2256 e 0,6871, respectivamente). Portanto, no componente logístico do modelo final permaneceram as variáveis IRA e idade e, no componente de sobrevivência, as variáveis IRA e forma de ingresso na universidade. A variável IRA, inicialmente considerada com 4 categorias, foi recategorizada (foram agrupadas as categorias 70-85 e 85-100), permanecendo, assim, 3 categorias (0-50, 50-70 e 70-100). Também a variável idade, que na análise exploratória foi considerada como categórica com 4 classes, foi incluída no modelo como contínua.

Nas Tabelas 10 e 11 têm-se, respectivamente, as estimativas e algumas estatísticas associadas ao componente logístico do modelo selecionado.

Tabela 10 - Estimativas associadas ao componente logístico $\pi(\mathbf{z})$ – Currículo 1991

Variável	Graus de Liberdade	Estimativas	Erro Padrão	Wald Qui-Quadrado	P-valor
Intercepto	1	-2,8556	0,4143	47,518	<0,0001
IRA (50-70)	1	4,3612	0,2516	300,57	<0,0001
IRA (70-100)	1	5,5373	0,3044	330,89	<0,0001
IDADE	1	-0,0255	0,0147	2,9880	0,0839

Obs: IRA 0-50 foi considerada como a categoria de referência.

Fonte: Os autores (2013).

Tabela 11 - Estatísticas associadas ao componente logístico – Currículo 1991

Critério	Somente Intercepto	Intercepto e variáveis
AIC	1540,968	659,229
-2 Log L	1538,968	651,229

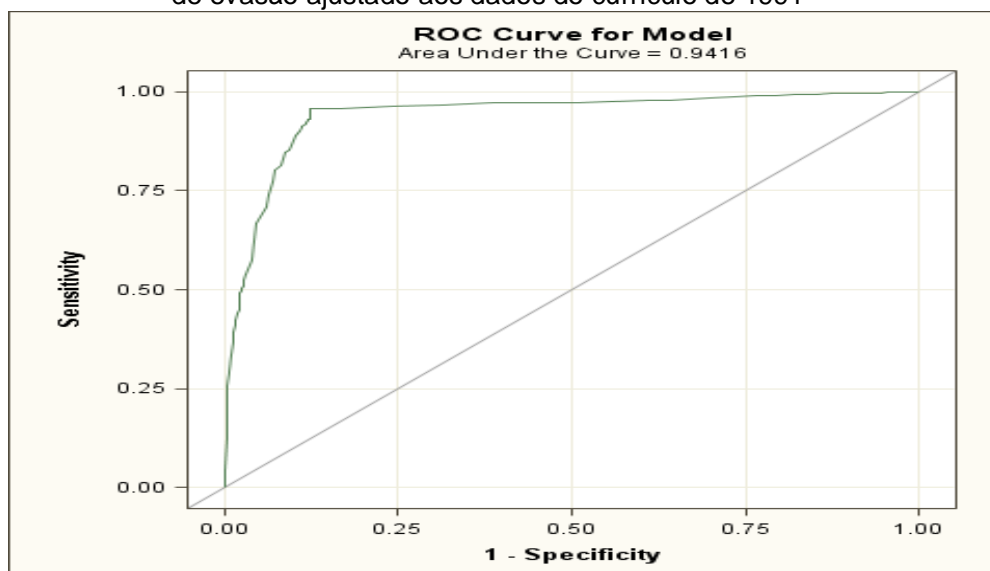
Fonte: Os autores (2013).

A partir da Tabela 10, pode-se notar que as variáveis remanescentes no componente logístico apresentaram significância ao nível de 0,10, obedecendo ao critério previamente estabelecido. Quanto aos valores do AIC e -2LogL (com LogL

denotando o logaritmo da função de verossimilhança) apresentados na Tabela 11, pode-se observar um decréscimo acentuado dos mesmos quando da comparação entre os modelos com somente o intercepto e com as variáveis IRA e idade, evidenciando a contribuição destas variáveis.

A curva ROC também foi utilizada para verificar a adequação do modelo de mistura ajustado. Na Figura 9 é possível visualizar tal curva associada ao componente logístico, cuja área abaixo da curva (AUC) resultou em 0,9416, evidenciando a adequação deste componente.

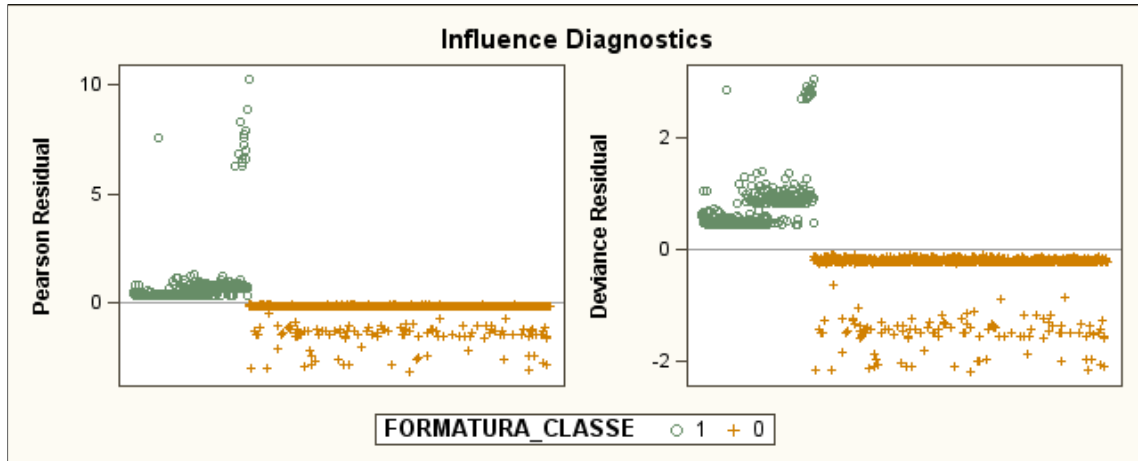
Figura 9 - Curva ROC associada ao componente logístico do modelo de mistura com fração de evasão ajustado aos dados do currículo de 1991



Fonte: Os autores (2013).

Além da curva ROC, análise dos resíduos *deviance* e de Pearson foi também realizada a fim de avaliar a presença de pontos influentes os quais poderiam estar afetando a qualidade do modelo ajustado. Tais resíduos, mostrados na Figura 10, apresentam valores em torno de zero e dentro de um intervalo de variação satisfatório, em particular os resíduos *deviance* (que são mais simétricos), o que corrobora com a adequação do componente logístico do modelo ajustado.

Figura 10 - Análise dos resíduos de Pearson e *deviance* associados ao componente logístico do modelo de mistura ajustado aos dados do currículo de 1991



Fonte: Os autores (2013).

Com base no que foi apresentado, o componente logístico ficou expresso por:

$$\hat{\pi}(z_i) = \frac{e^{\beta_0 + \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_3}}{1 + e^{\beta_0 + \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_3}} = \frac{e^{2,8556 + 4,3612z_1 + 5,5373z_2 - 0,0255z_3}}{1 + e^{2,8556 + 4,3612z_1 + 5,5373z_2 - 0,0255z_3}}$$

bem como o correspondente componente em termos do logito por:

$$\text{logit}(\hat{\pi}(z_i)) = 2,8556 + 4,3612z_1 + 5,5373z_2 - 0,0255z_3,$$

com $(z_1 = 0, z_2 = 0)$ se o aluno tem IRA entre 0-50; $(z_1 = 1, z_2 = 0)$ se o aluno tem IRA entre 50 e 70 e $(z_1 = 0, z_2 = 1)$ se o aluno tem IRA entre 70 e 100; e z_3 a idade do aluno (em anos).

Logo, *odds ratios* (razões de chances) podem ser obtidas utilizando as estimativas dos parâmetros, no intuito de encontrar resultados sobre o perfil dos alunos em função das variáveis. Nos casos em que não há interação entre as variáveis, a *odds* pode ser calculada por $\exp\{\hat{\beta}_i\}$. Por exemplo, para alunos que apresentaram IRA entre 50 e 70 em relação aos alunos que apresentaram IRA menor que 50, tem-se:

$$\widehat{OR}_{IRA(50-70|<50)} = \exp\{\widehat{\beta}_3\},$$

o que resulta em:

$$\widehat{OR}_{IRA(50-70|<50)} = \exp\{4,3612\} = 78,3511.$$

Com isso, a chance de um aluno se formar dado que ele tem IRA entre 50 e 70 foi estimada em, aproximadamente, 78 vezes a de um aluno que tem IRA abaixo de 50, mantendo fixos os valores das demais variáveis.

Utilizando os valores das estimativas dos parâmetros mostrados na Tabela 10 é possível calcular outras razões de chances. Por exemplo,

$$\widehat{OR}_{20|30} = \frac{odds_{(20)}}{odds_{(30)}} = \frac{\exp\{-0,0255*20\}}{\exp\{-0,0255*30\}} = \exp\{\beta_3(20 - 30)\} = 1,2904;$$

ou seja, alunos que possuem idade 20 anos apresentam chance de se formarem 1,29 vezes a dos que possuem idade 30 anos.

Em termos gerais, o perfil dos alunos que apresentaram maior probabilidade de formatura foi o com as seguintes características: IRA maior que 70, combinado com pouca idade. Por outro lado, o perfil dos com menor probabilidade de formatura foi: IRA abaixo de 50, combinado com idade mais avançada.

Quanto ao componente de sobrevivência $S(t | U=1, \mathbf{x})$, tem-se na Tabela 12 as correspondentes estimativas associadas ao modelo de mistura ajustado. A partir desta tabela, pode-se notar que as variáveis remanescentes no componente de sobrevivência apresentaram significância ao nível de 0,20, obedecendo ao critério previamente estabelecido.

Tabela 12 - Estimativas associadas ao componente de sobrevivência do modelo de mistura ajustado aos dados do Currículo de 1991

Variável	Graus de liberdade	Estimativas	Erro Padrão	Wald Qui-Quadrado	P-valor
IRA (50-70)	1	1,0686	0,2748	15,1187	0,0001
IRA (70-100)	1	2,3031	0,2805	67,4289	< 0,0001
VEST_1	1	1,4456	0,2172	44,2968	< 0,0001

Fonte: Os autores (2013).

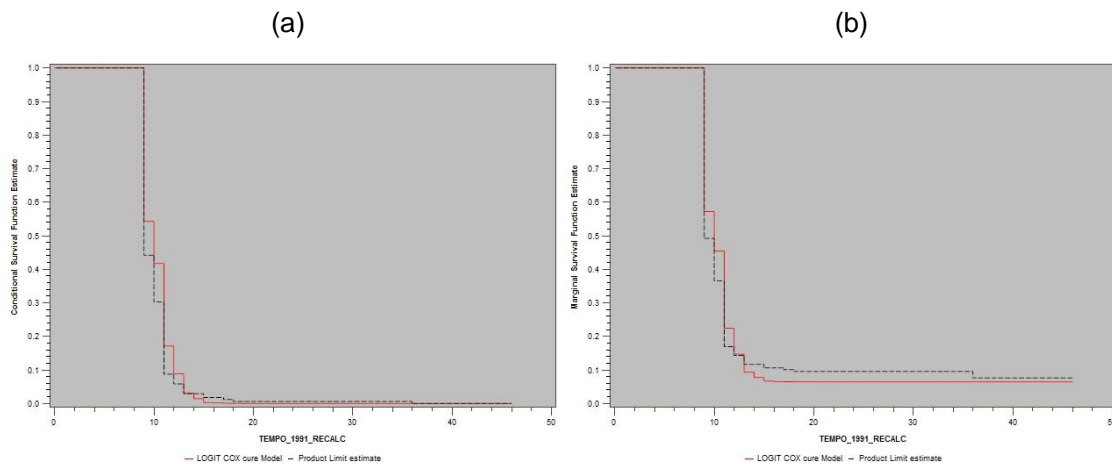
Tabela 13 - Estatísticas associadas ao modelo escolhido (componente de sobrevivência) – Currículo de 1991

Critério	Somente Intercepto	Intercepto e variáveis
-2 LOG L	3.663,49	3.474,95
AIC	3.663,49	3.480,95

Fonte: Os autores (2013).

Para verificar a qualidade do ajuste do modelo final, foram comparadas as curvas de sobrevivência observadas (representadas pelas curvas obtidas pelo estimador de Kaplan-Meier) e as curvas de sobrevivência estimadas pelo modelo, tanto para a sobrevivência populacional $S(t | \mathbf{x}, \mathbf{z})$ quanto para a sobrevivência condicional $S(t | U=1, \mathbf{x})$. Estas curvas, para uma das combinações de \mathbf{x} e \mathbf{z} , podem ser visualizadas na Figura 11 e mostram que as estimativas produzidas pelo modelo são bastante próximas das obtidas por Kaplan-Meier, evidenciando a adequação do modelo aos dados.

Figura 11 - Curvas de sobrevivência observada e estimada pelo modelo de fração de evasão ajustado para (a) $S(t | U=1, \mathbf{x})$ e (b) $S(t | \mathbf{x}, \mathbf{z})$ para uma das combinações das variáveis IRA e forma de ingresso na universidade para currículo 1991



Fonte: Os autores (2013).

Para avaliar a adequação do modelo final foram também obtidos os coeficientes de correlação de Pearson e seus correspondentes valores ao quadrado (R^2) entre as probabilidades de sobrevivência observadas (obtidas pelo método proposto por Kaplan-Meier) e as estimadas pelo modelo para cada uma das combinações das variáveis IRA e forma de ingresso na universidade. Ambas as estatísticas indicaram bom ajuste do modelo, produzindo correlações satisfatórias (acima de 0,7) para 4 das 6 combinações, conforme pode ser visto na Tabela 14.

Tabela 14 - Correlação estatística entre as funções de sobrevivência estimadas e observadas (Kaplan-Meier) para cada uma das combinações das variáveis IRA e forma de ingresso na universidade para o currículo 1991

Estrato	IRA	Forma de Ingresso	R^2	Coefficiente de Correlação de Pearson
1	0 – 50	Outros	0,00859	0,09268
2	0 – 50	Vestibular	0,11939	0,34553
3	50 – 70	Outros	0,50285	0,70912
4	50 – 70	Vestibular	0,98771	0,99384
5	70 – 100	Outros	0,91145	0,9547
6	70 – 100	Vestibular	0,98007	0,98998

Fonte: Os autores (2013).

Com isso, tem-se a expressão do modelo final dada por:

$$S(t | \mathbf{x}, \mathbf{z}) = \pi(\mathbf{z})S(t | U = 1, \mathbf{x}) + 1 - \pi(\mathbf{z}),$$

sendo $\pi(\mathbf{z})$ o componente logístico e $S(t | U = 1, \mathbf{x})$ o componente de sobrevivência condicional expresso por:

$$S(t | \mathbf{x}, \mathbf{z}) = \frac{\exp(\mathbf{z}'\boldsymbol{\beta})}{1 + \exp(\mathbf{z}'\boldsymbol{\beta})} S_0(t|U = 1)^{\exp(\boldsymbol{\gamma}'\mathbf{x})} + 1 - \frac{\exp(\mathbf{z}'\boldsymbol{\beta})}{1 + \exp(\mathbf{z}'\boldsymbol{\beta})}.$$

As estimativas dos vetores de parâmetros $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ estão apresentadas nas Tabelas 10 e 12, respectivamente.

A Tabela 15 mostra os resultados do modelo ajustado para todas as combinações das variáveis. Considerando as estimativas do modelo, pode-se notar que alunos que apresentaram IRA entre 0 e 50 e idade igual a 40 anos apresentaram uma taxa de evasão de 97,97%, a maior entre todas as combinações. Por outro lado, a menor taxa de evasão (26,98%) foi a dos alunos com IRA entre 70 e 100 e idade igual a 20 anos.

A partir da Tabela 15, que apresenta a probabilidade de um aluno não se formar para os tempos 9, 12 e 15 semestres, pode-se notar para alunos que apresentaram IRA entre 0 e 50 e forma de ingresso igual a outros, que a probabilidade de sobrevivência em $t = 9$ foi de 98,57%. Ou seja, de acordo com as estimativas obtidas a partir do modelo, 98,57% dos alunos com IRA entre 0 e 50 e que ingressaram por outras formas que não o vestibular, não se formaram no tempo

mínimo estabelecido para o currículo 1991. Por outro lado, alunos com IRA entre 70 e 100 e que ingressaram via vestibular apresentaram probabilidade de sobrevivência em $t = 9$ e $t = 15$ de 54,26% e 0,27%, respectivamente. Ou seja, em $t = 15$ apenas 0,27% dos alunos que apresentaram IRA entre 70 e 100 e forma de ingresso igual a vestibular não tinham se formado.

Tabela 15 - Estimativas obtidas com o modelo de mistura com fração de evasão ao longo do tempo, para cada combinação das variáveis do currículo 1991.

Combinações			<i>Taxa de evasão</i>			$S(t U = 1, \mathbf{x})$			$S(t \mathbf{x}, \mathbf{z})$		
			$1 - \pi(\mathbf{z})$			t= 9 meses	t= 12 meses	t= 15 meses	t= 9 meses	t= 12 meses	t= 15 meses
Estrato	IRA	Forma de Ingresso	Idade 20 anos	Idade 30 anos	Idade 40 anos	t= 9 meses	t= 12 meses	t= 15 meses	t= 9 meses	t= 12 meses	t= 15 meses
1	0 - 50	Outros	96,66%	97,39%	97,97%	98,57%	94,46%	86,97%	99,95%	99,86%	99,74%
2	0 - 50	Vestibular	96,66%	97,39%	97,97%	94,07%	78,50%	55,28%	97,00%	89,64%	79,32%
3	50 - 70	Outros	26,98%	32,29%	38,09%	95,90%	84,70%	66,59%	87,96%	62,09%	36,74%
4	50 - 70	Vestibular	26,98%	32,29%	38,09%	83,70%	49,42%	17,81%	99,95%	99,86%	99,74%
5	70 - 100	Outros	10,23%	12,82%	15,95%	86,59%	56,52%	24,73%	97,00%	89,64%	79,32%
6	70 - 100	Vestibular	10,23%	12,82%	15,95%	54,26%	8,87%	0,27%	87,96%	62,09%	36,74%

Fonte: Os autores (2013).

4.2.3 Modelo de Mistura com Fração de Evasão para o Currículo 2006

Para o currículo 2006 as curvas de sobrevivência mostradas na Seção 4.2 apresentam também um número elevado de censuras à direita. Contudo, diferente do currículo 1991, em que há somente alunos evadidos e formados, no currículo 2006 há alunos que ainda estão cursando. Com isso, o modelo de mistura com fração de evasão descrito na Seção 3.2.4 foi escolhido para modelar o tempo até o término do Curso de graduação em Estatística do currículo 2006.

Assim como no currículo 1991, todas as variáveis foram incluídas simultaneamente no componente logístico $\pi(\mathbf{z})$ e no componente de sobrevivência $S(t | U = 1, \mathbf{x})$. Os critérios estabelecidos para permanência ou inclusão de variáveis foram os mesmos ($\alpha = 0,1$ para o componente logístico e $0,20$ para o componente de sobrevivência). As variáveis gênero e tipo de ingresso foram removidas do componente logístico por terem apresentado efeito não significativo (p-valores iguais a $0,7325$ e $0,4592$, respectivamente). Já para o componente de sobrevivência, apenas a variável idade não apresentou efeito significativo (p-valor igual a $0,3483$). Com isso, no componente logístico do modelo final permaneceram as variáveis IRA e idade e, no componente de sobrevivência, as variáveis: gênero, IRA e forma de ingresso na universidade. A variável IRA, inicialmente considerada com 4 categorias, também foi recategorizada (foram agrupadas as categorias 70-85 e 85-100), permanecendo, assim, 3 categorias (0-50, 50-70 e 70-100). Também a variável idade, que na análise exploratória foi considerada como categórica com 4 classes, foi incluída no modelo como contínua.

Nas Tabelas 16 e 17 têm-se, respectivamente, as estimativas e algumas estatísticas associadas ao componente logístico do modelo selecionado para os dados do currículo 2006.

Tabela 16 - Estimativas associadas ao componente logístico $\pi(\mathbf{z})$ – currículo 2006

Variável	GL	Estimativas	Erro Padrão	Wald Qui-Quadrado	P-valor
Intercepto	1	-0,4881	0,6659	0,5373	0,4635
IRA (50-70)	1	3,9089	0,4368	80,1026	<0,0001
IRA (70-100)	1	4,7221	0,5458	74,8555	<0,0001
IDADE	1	-0,1201	0,0287	17,5612	<0,0001

Obs: IRA 0-50 foi considerada como a categoria de referência

Fonte: Os autores (2013).

Tabela 17 - Estatísticas associadas ao componente logístico – currículo 2006

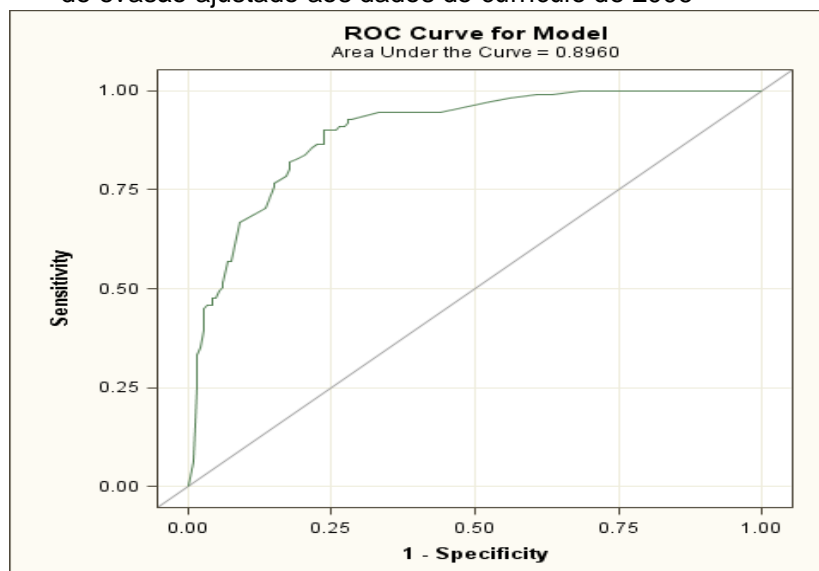
Critério	Somente Intercepto	Intercepto e variáveis
AIC	413,703	204,266
-2 Log L	411,703	196,266

Fonte: Os autores (2013).

A partir da Tabela 16, pode-se notar que todas as variáveis remanescentes no componente logístico apresentaram significância ao nível de 0,0001. Quanto aos valores do AIC e -2LogL apresentados na Tabela 17, observa-se uma diminuição acentuada dos mesmos quando da comparação entre os modelos com somente o intercepto e com as variáveis IRA e idade, evidenciando a contribuição destas variáveis.

Assim como no modelo do currículo 1991, a curva ROC também foi utilizada para verificar a adequação do modelo de mistura ajustado. Na Figura 12 é possível visualizar a curva associada ao componente logístico, cuja área abaixo da curva (AUC) resultou em 0,8960, evidenciando a adequação deste componente.

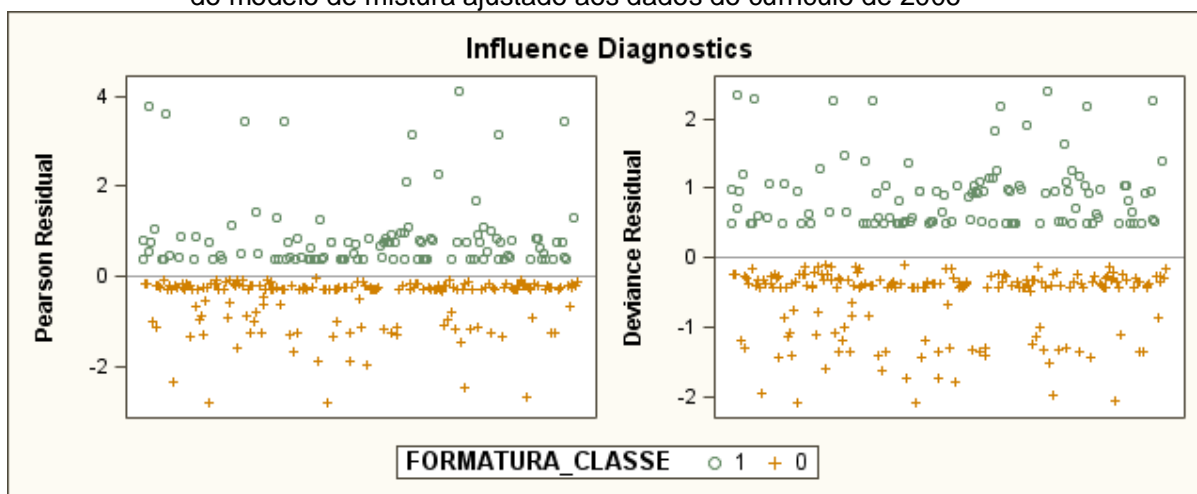
Figura 12 - Curva ROC associada ao componente logístico do modelo de mistura com fração de evasão ajustado aos dados do currículo de 2006



Fonte: Os autores (2013).

Além da curva ROC, análise dos resíduos *deviance* e de Pearson também foi realizada a fim de avaliar a presença de pontos influentes os quais poderiam estar afetando a qualidade do modelo ajustado. Na Figura 13, tais resíduos apresentam valores em torno de zero dentro de um intervalo de variação aceitável, o que reitera a adequação do componente logístico do modelo ajustado.

Figura 13 - Análise dos resíduos de Pearson e *deviance* associados ao componente logístico do modelo de mistura ajustado aos dados do currículo de 2006



Fonte: Os autores (2013).

Com base no que foi apresentado, o componente logístico ficou expresso por:

$$\hat{\pi}(\mathbf{z}_i) = \frac{e^{\beta_0 + \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_3}}{1 + e^{\beta_0 + \beta_1 z_1 + \beta_2 z_2 + \beta_3 z_3}} = \frac{e^{0,4881 + 3,9089z_1 + 4,7221z_2 - 0,1201z_3}}{1 + e^{0,4881 + 3,9089z_1 + 4,7221z_2 - 0,1201z_3}}$$

bem como o correspondente componente em termos do logito por:

$$\text{logit}(\hat{\pi}(\mathbf{z}_i)) = 0,4881 + 3,9089z_1 + 4,7221z_2 - 0,1201z_3,$$

com $(z_1 = 0, z_2 = 0)$ se o aluno tem IRA entre 0-50; $(z_1 = 1, z_2 = 0)$ se o aluno tem IRA entre 50 e 70 e $(z_1 = 0, z_2 = 1)$ se o aluno tem IRA entre 70 e 100; e z_3 a idade do aluno (em anos).

Com isso, *odds ratios* (razões de chances) podem ser obtidas utilizando as estimativas dos parâmetros com o intuito de encontrar resultados sobre o perfil dos alunos em função das variáveis. Por exemplo, para alunos que apresentaram IRA entre 50 e 70 em relação aos alunos que apresentaram IRA menor que 50, tem-se:

$$\widehat{OR}_{IRA(50-70|<50)} = \exp\{\widehat{\beta}_3\},$$

o que resulta em:

$$\widehat{OR}_{IRA(50-70|<50)} = \exp\{3,9089\} = 49,8440.$$

Com isso, a chance de um aluno se formar, dado que ele tem IRA entre 50 e 70 foi estimada em, aproximadamente, 50 vezes a de um aluno que tem IRA abaixo de 50, mantendo fixos os valores das demais covariáveis.

Utilizando os valores das estimativas dos parâmetros mostrados na Tabela 16 é possível calcular outras razões de chances. Por exemplo,

$$\widehat{OR}_{20|30} = \frac{\text{odds}_{(20)}}{\text{odds}_{(30)}} = \frac{\exp\{-0,1201 \cdot 20\}}{\exp\{-0,1201 \cdot 30\}} = \exp\{\beta_3(20 - 30)\} = 3,3234;$$

ou seja, alunos que possuem idade 20 anos apresentaram chance de se formarem de, aproximadamente, 3 vezes a dos que possuem idade 30 anos.

De forma geral, o perfil dos alunos que apresentaram maior probabilidade de formatura foi o com as seguintes características: IRA maior que 70, combinado com pouca idade. Por outro lado, o perfil dos com menor probabilidade de formatura foi: IRA abaixo de 50, combinado com idade mais avançada.

Quanto ao componente de sobrevivência $S(t | U = 1, \mathbf{x})$, tem-se na Tabela 18 as correspondentes estimativas associadas ao modelo de mistura ajustado. A partir

desta tabela, é possível observar que as variáveis remanescentes no componente de sobrevivência apresentaram significância ao nível de 0,20, obedecendo ao critério previamente estabelecido de 0,20.

Tabela 18 - Estimativas associadas ao componente de sobrevivência do modelo de mistura ajustado aos dados do Currículo de 2006

Variável	GL	Estimativas	Erro Padrão	Wald Qui-Quadrado	P-valor
Gênero	1	0,3967	0,1950	4,1380	0,0419
IRA (50-70)	1	0,5186	0,4110	1,6570	0,1980
IRA (70-100)	1	1,5864	0,4145	14,6472	0,0001
VEST_1	1	0,8945	0,3748	5,6963	0,0170

Obs: IRA 0-50 e gênero masculino foram considerados como categorias de referência.

Fonte: Os autores (2013).

Tabela 19 - Estatísticas associadas ao modelo escolhido (componente de sobrevivência) – currículo de 2006

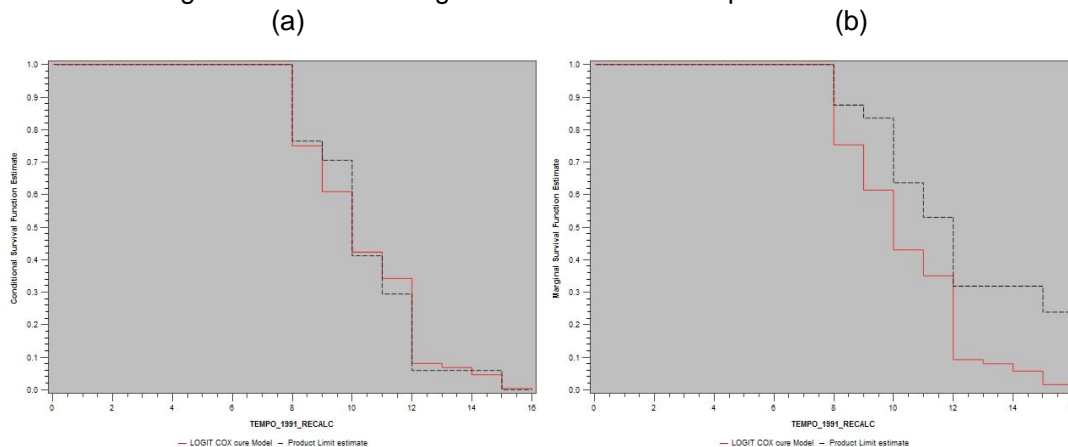
Critério	Somente Intercepto	Intercepto e variáveis
-2 LOG L	939,237	895,578
AIC	939,237	903,578

Fonte: Os autores (2013).

Quanto aos valores do AIC e -2LogL apresentados na Tabela 19, pode-se observar um decréscimo dos mesmos quando da comparação entre os modelos com somente o intercepto e com as variáveis IRA e forma de ingresso na universidade, evidenciando a contribuição destas variáveis.

Para verificar a qualidade do ajuste do modelo final, foram comparadas as curvas de sobrevivência observadas (representadas pelas curvas obtidas pelo estimador de Kaplan-Meier) e as curvas de sobrevivência estimadas pelo modelo, tanto para a sobrevivência populacional $S(t | \mathbf{x}, \mathbf{z})$ quanto para a sobrevivência condicional $S(t | U=1, \mathbf{x})$. Estas curvas, para uma das combinações de \mathbf{x} e \mathbf{z} , podem ser visualizadas na Figura 14 e mostram que as estimativas produzidas pelo modelo são bastante próximas das obtidas por Kaplan-Meier, evidenciando a adequação do modelo aos dados.

Figura 14 - Curvas de sobrevivência observada e estimada pelo modelo de fração de evasão ajustado para (a) $S(t | U = 1, \mathbf{x})$ e (b) $S(t | \mathbf{x}, \mathbf{z})$ para uma das combinações das variáveis IRA, gênero e forma de ingresso na universidade para currículo 2006



Fonte: Os autores (2013).

Para avaliar a adequação do modelo final foi também calculado o coeficiente de correlação de Pearson (e seu correspondente valor ao quadrado R^2) entre as probabilidades de sobrevivência observadas, obtidas pelo método proposto por Kaplan-Meier, e as estimadas pelo modelo para cada uma das combinações das variáveis IRA e forma de ingresso na universidade. Ambas as estatísticas mostram o bom ajuste do modelo, tendo em vista as correlações elevadas obtidas para todas as combinações, conforme pode ser observado na Tabela 20. O menor valor para o coeficiente de Pearson foi 0,8766 e para o R^2 , 0,7685. Para as combinações que apresentaram um número reduzido de alunos (indicados com asterisco na Tabela 20) ficou inviável o cálculo desses valores.

Tabela 20 - Correlação estatística entre as funções de sobrevivência estimadas e observadas (Kaplan-Meier) para cada uma das combinações das variáveis IRA e forma de ingresso na universidade para o currículo 2006

Estrato	Gênero	IRA	Forma de Ingresso	R^2	Coefficiente de Correlação de Pearson
1	Masculino	0 - 50	Outros	*	*
2	Masculino	0 - 50	Vestibular	0,85596	0,92518
3	Masculino	50 - 70	Outros	*	*
4	Masculino	50 - 70	Vestibular	0,97656	0,98821
5	Masculino	70 - 100	Outros	0,95989	0,97974
6	Masculino	70 - 100	Vestibular	0,98059	0,99025
7	Feminino	0 - 50	Outros	0,92825	0,96346
8	Feminino	0 - 50	Vestibular	0,90082	0,94912
9	Feminino	50 - 70	Outros	0,76859	0,87669
10	Feminino	50 - 70	Vestibular	0,98185	0,99088
11	Feminino	70 - 100	Outros	*	*
12	Feminino	70 - 100	Vestibular	0,97219	0,986

Fonte: Os autores (2013).

* combinações com número reduzido de alunos, o que tornou inviável o cálculo do coeficiente de Pearson e seu correspondente valor ao quadrado R^2 .

Com isso, tem-se a expressão do modelo final dada por:

$$S(t | \mathbf{x}, \mathbf{z}) = \pi(\mathbf{z}) S(t | U = 1, \mathbf{x}) + 1 - \pi(\mathbf{z})$$

sendo $\pi(\mathbf{z})$ o componente logístico e $S(t | U = 1, \mathbf{x})$ o componente de sobrevivência condicional dado por:

$$S(t | \mathbf{x}, \mathbf{z}) = \frac{\exp(\mathbf{z}'\boldsymbol{\beta})}{1 + \exp(\mathbf{z}'\boldsymbol{\beta})} S_0(t | U = 1)^{\exp(\boldsymbol{\gamma}'\mathbf{x})} + 1 - \frac{\exp(\mathbf{z}'\boldsymbol{\beta})}{1 + \exp(\mathbf{z}'\boldsymbol{\beta})}$$

As estimativas dos parâmetros $\boldsymbol{\beta}$ e $\boldsymbol{\gamma}$ estão nas Tabelas 16 e 18, respectivamente. A Tabela 21 apresenta as estimativas do modelo para todas as combinações das variáveis: gênero, IRA e forma de ingresso. Pode-se observar que alunos com IRA entre 0 e 50 e idade igual a 20 anos apresentaram uma taxa de evasão 87,14%, a maior entre todas combinações. Por outro lado, alunos com IRA entre 70 e 100 e idade igual a 20 anos apresentaram uma taxa de evasão de 11,97%. Com isso, nota-se, como esperado, que à medida que o IRA aumenta, a taxa de evasão diminui.

Entre os alunos que se formaram, a Tabela 21 apresenta as probabilidades de sobrevivência para $t = 8, 10$ e 12 semestres. Observa-se que para os alunos do gênero masculino com IRA entre 0 e 50 e forma de ingresso igual a outros a probabilidade de sobrevivência (i.e., probabilidade do aluno não se formar) em $t = 9$ foi estimada em 95,41%. Ou seja, de acordo com o modelo, 95,41% dos alunos do gênero masculino com IRA entre 0 e 50 e que ingressaram por outras formas que não o vestibular não se formaram no tempo mínimo estabelecido para o currículo 2006. Por outro lado, para alunos do gênero feminino com IRA entre 70 e 100 e que ingressaram por vestibular, a probabilidade de sobrevivência em $t = 8$ foi estimada em 43,35%. Ainda, observa-se a partir da Tabela 21, que os alunos do gênero masculino com IRA entre 70 e 100 e que ingressaram por vestibular apresentaram uma probabilidade de sobrevivência de 57%.

Com isso, nota-se que os alunos do gênero feminino se formam em maior proporção que os alunos do gênero masculino. Nota-se, também, que para todas as combinações das variáveis, o gênero feminino apresenta tempos até a graduação menores que o dos alunos do gênero masculino.

Tabela 21 - Estimativas obtidas com o modelo de mistura com fração de evasão ao longo do tempo, para cada combinação das variáveis do currículo 2006

Combinações				<i>Taxa de evasão</i> $1 - \pi(z)$			$S(t U = 1, x)$			$S(t x, z)$		
Estrato	Gênero	IRA	Forma de Ingresso	Idade 20 anos	Idade 30 anos	Idade 40 anos	t= 8 meses	t= 10 meses	t= 12 meses	t= 8 meses	t= 10 meses	t= 12 meses
1	Masculino	0 - 50	Outros	87,14%	95,75%	98,68%	95,41%	86,86%	66,29%	99,41%	99,44%	99,56%
2	Masculino	0 - 50	Vestibular	87,14%	95,75%	98,68%	89,13%	70,85%	36,58%	99,41%	99,44%	99,56%
3	Masculino	50 - 70	Outros	11,97%	31,11%	60,01%	92,40%	78,93%	50,13%	93,31%	85,48%	80,05%
4	Masculino	50 - 70	Vestibular	11,97%	31,11%	60,01%	82,43%	56,05%	18,46%	93,31%	85,48%	80,05%
5	Masculino	70 - 100	Outros	5,68%	16,68%	39,95%	79,47%	50,24%	13,41%	80,64%	58,54%	48,00%
6	Masculino	70 - 100	Vestibular	5,68%	16,68%	39,95%	57,00%	18,56%	0,73%	80,64%	58,54%	48,00%
7	Feminino	0 - 50	Outros	87,14%	95,75%	98,68%	93,25%	81,10%	54,26%	99,41%	99,44%	99,56%
8	Feminino	0 - 50	Vestibular	87,14%	95,75%	98,68%	84,28%	81,10%	54,26%	99,41%	99,44%	99,56%
9	Feminino	50 - 70	Outros	11,97%	31,11%	60,01%	88,92%	70,34%	35,81%	93,31%	85,48%	80,05%
10	Feminino	50 - 70	Vestibular	11,97%	31,11%	60,01%	75,02%	42,28%	8,11%	93,31%	85,48%	80,05%
11	Feminino	70 - 100	Outros	5,68%	16,68%	39,95%	71,05%	35,93%	5,04%	80,64%	58,54%	48,00%
12	Feminino	70 - 100	Vestibular	5,68%	16,68%	39,95%	43,35%	8,17%	0,07%	80,64%	58,54%	48,00%

Fonte: Os autores (2013).

4.2.4 Comparação entre currículos

É importante ressaltar que o tempo mínimo até a formatura difere nos dois currículos analisados. No currículo 1991 eram necessários pelo menos 9 semestres (4,5 anos) para a conclusão do Curso de Estatística na UFPR, enquanto no currículo 2006, pelo menos 8 semestres (4 anos).

Com isso, a Tabela 22 foi construída considerando para o componente de sobrevivência os tempos mínimos para cada currículo (8 semestres para o currículo de 1991 e 9 semestres para o de 2006). Posteriormente, foram adicionados 2 e 4 semestres para que se tenha uma visão comparável dos currículos.

De modo geral, o modelo de fração de evasão se ajustou bem aos dados de ambos os currículos. Quando observados separadamente, os componentes logístico e o de sobrevivência apresentam resultados bastante similares em ambos os currículos, com leve destaque para o currículo 2006, que apresentou valores maiores de correlação de Pearson e R^2 (correlação de Pearson elevada ao quadrado) no componente de sobrevivência, apesar do baixo número de alunos em algumas combinações.

A partir da Tabela 22 pode-se notar que a probabilidade de formatura dos alunos do currículo 2006 é maior do que a dos alunos do currículo 1991, para a maioria das combinações. Por exemplo, alunos com IRA entre 0 e 50 e idade igual a 20 anos apresentaram taxas de evasão de 96,66% e 87,14% para os currículos 1991 e 2006, respectivamente. Neste caso, estão sendo comparados alunos de mesmo perfil em ambos os currículos. No geral, alunos do currículo 2006 apresentaram uma taxa de evasão menor que a dos alunos do currículo 1991.

Quanto ao tempo até a formatura, o mesmo acontece. Por exemplo, alunos de IRA entre 70 e 100, gênero feminino e que ingressaram por vestibular apresentaram probabilidade de sobrevivência em $t = 9$ e $t = 8$ de 54,23% e 43,35% para os currículos 1991 e 2006, respectivamente. Com isso, verifica-se que alunos do currículo 2006 apresentaram probabilidade de sobrevivência menor (consequentemente, probabilidade de formatura maior) do que os alunos do currículo 1991.

Tabela 22 - Estimativas obtidas com o modelo de mistura com fração de evasão ao longo do tempo, para cada combinação das variáveis do currículos 1991 e 2006

Combinações				<i>Taxa de evasão</i>			$S(t U = 1, x)$			$S(t x, z)$			
Estrato	Gênero	IRA	Forma de Ingresso	Idade 20 anos	Idade 30 anos	Idade 40 anos	t= 9 meses	t= 11 meses	t= 13 meses	t= 9 meses	t= 11 meses	t= 13 meses	
Currículo 2006	1	Masculino	0 - 50	Outros	87,14%	95,75%	98,68%	95,41%	86,86%	66,29%	99,41%	99,44%	99,56%
	2	Masculino	0 - 50	Vestibular	87,14%	95,75%	98,68%	89,13%	70,85%	36,58%	99,41%	99,44%	99,56%
	5	Masculino	50 - 70	Outros	11,97%	31,11%	60,01%	92,40%	78,93%	50,13%	93,31%	85,48%	80,05%
	6	Masculino	50 - 70	Vestibular	11,97%	31,11%	60,01%	82,43%	56,05%	18,46%	93,31%	85,48%	80,05%
	5	Masculino	70 - 100	Outros	5,68%	16,68%	39,95%	79,47%	50,24%	13,41%	80,64%	58,54%	48,00%
	6	Masculino	70 - 100	Vestibular	5,68%	16,68%	39,95%	57,00%	18,56%	0,73%	80,64%	58,54%	48,00%
	7	Feminino	0 - 50	Outros	87,14%	95,75%	98,68%	93,25%	81,10%	54,26%	99,41%	99,44%	99,56%
	8	Feminino	0 - 50	Vestibular	87,14%	95,75%	98,68%	84,28%	81,10%	54,26%	99,41%	99,44%	99,56%
	9	Feminino	50 - 70	Outros	11,97%	31,11%	60,01%	88,92%	70,34%	35,81%	93,31%	85,48%	80,05%
	10	Feminino	50 - 70	Vestibular	11,97%	31,11%	60,01%	75,02%	42,28%	8,11%	93,31%	85,48%	80,05%
	11	Feminino	70 - 100	Outros	5,68%	16,68%	39,95%	71,05%	35,93%	5,04%	80,64%	58,54%	48,00%
	12	Feminino	70 - 100	Vestibular	5,68%	16,68%	39,95%	43,35%	8,17%	0,07%	80,64%	58,54%	48,00%
Currículo 1991	Estrato	Gênero	IRA	Forma de Ingresso	Idade 20 anos	Idade 30 anos	Idade 40 anos	t= 9 meses	t= 11 meses	t= 13 meses	t= 9 meses	t= 11 meses	t= 13 meses
	1	*	0 - 50	Outros	96,66%	97,39%	97,97%	98,57%	95,93%	92,14%	99,95%	99,89%	99,84%
	2	*	0 - 50	Vestibular	96,66%	97,39%	97,97%	94,07%	83,82%	70,66%	99,95%	99,89%	99,84%
	3	*	50 - 70	Outros	26,98%	32,29%	38,09%	95,90%	88,60%	78,80%	97,00%	92,28%	86,88%
	4	*	50 - 70	Vestibular	26,98%	32,29%	38,09%	83,70%	59,82%	36,38%	97,00%	92,28%	86,88%
	5	*	70 - 100	Outros	10,23%	12,82%	15,95%	86,59%	65,97%	44,10%	87,96%	70,33%	53,02%
6	*	70 - 100	Vestibular	10,23%	12,82%	15,95%	54,26%	17,10%	3,10%	87,96%	70,33%	53,02%	

Fonte: Os autores (2013).

* variável não significativa no currículo 1991.

5 CONCLUSÃO

Para os currículos 1991 e 2006, os modelos ajustados apresentaram ajustes satisfatórios tanto no componente logístico quanto no componente de sobrevivência.

No que diz respeito ao componente logístico, foi possível obter, a partir das estimativas do modelo, a probabilidade de cada aluno se formar ou não se formar, o que contribuiu na definição de alguns perfis de alunos que têm maior e menor probabilidade de se formar. De forma geral, alunos com menos idade e que possuem IRA maior que 70 apresentaram maior probabilidade de falharem (i.e., de se formarem). E, conseqüentemente, alunos com mais idade e com IRA até 50 apresentaram menor probabilidade de falha. Em ambos os currículos, as variáveis que permaneceram no componente logístico foram: idade (contínua) e a variável Índice de Rendimento Acadêmico (IRA) categorizada em três classes (0-50, 50-70 e maior que 70). A variável Idade se refere à idade do aluno no momento do ingresso no curso, Já o IRA, corresponde ao valor obtido no último semestre para os que se formaram e, para os que se evadiram ou que ainda estão cursando, ao valor mais recente até o ano de 2013.

Quanto ao componente de sobrevivência, que forneceu informação referente ao risco dos alunos se formarem (falha) ou não se formarem (não falha) ao longo do período todo de acompanhamento, foi possível obter, a partir das estimativas fornecidas pelo modelo, a probabilidade de cada aluno sobreviver (i.e., não se formar) a cada instante de tempo após o nono e oitavo períodos, para os currículos 1991 e 2006, respectivamente. Em termos gerais, alunos que ingressaram por vestibular e obtiveram IRA maior que 70 foram os que apresentaram maior probabilidade de formatura para todos os instantes a partir do nono e oitavo semestres (currículos 1991 e 2006, respectivamente). Para o currículo 1991 as variáveis que permaneceram no componente de sobrevivência foram: forma de ingresso na universidade e o IRA. E no currículo 2006, além de forma de ingresso na universidade e IRA, a variável gênero também apresentou efeito significativo.

Observa-se que o conhecimento desses perfis é relevante para a Coordenação do Curso de Estatística da UFPR no sentido de direcionar ao estabelecimento de estratégias adequadas aos alunos.

O PSE é o ponto de maior atenção na mudança dos currículos. Em um primeiro momento, avaliando a diminuição na taxa de evasão, há indícios que o PSE

esteja sendo efetivo. Porém, deve-se levar em consideração o PSE como um todo, ou seja, todo o esforço que é feito pela UFPR e pelos professores, no qual são disponibilizadas duas salas com as mesmas disciplinas, além de todo o processo de logística para o cadastro provisório destes alunos. E é claro, o custo envolvido em cada etapa do processo.

Historicamente, a quantidade de alunos que entra no PSE (2,5 vezes o número de vagas disponível, que é de 66) é muito maior que a quantidade de alunos que conseguem entrar no curso. Contudo, alguns dos que entram não chegam até o final do processo, de modo que o PSE funciona como uma espécie de filtro, no qual grande parte dos alunos acaba se evadindo seja pela dificuldade de acompanhar o ritmo do curso ou por não se identificar com o mesmo. Com isso, a taxa de evasão para o currículo 2006 se torna menor que a do currículo 1991.

Para trabalhos futuros, sugere-se o acompanhamento dos alunos do currículo 2006 por um período de tempo mais extenso para que possam ser observados os alunos que ainda estão cursando. Sugere-se, também, que quando forem disponibilizadas pelo NC (Núcleo de Concurso) da UFPR as notas das três fases do PSE, seja realizado um estudo comparativo entre os 66 alunos que entrariam se fossem consideradas somente as fases 1 e 2 (vestibular padrão) e aqueles que efetivamente entraram com base nas três etapas do PSE (estudo de correlação entre a classificação de ambos os processos). Com isso, caso seja verificado que os alunos que efetivamente entram no curso por meio do PSE são os melhores classificados no vestibular composto de 2 fases, não haveria razões muito efetivas para a manutenção do PSE.

REFERÊNCIAS

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

CORBIÈRE, F.; JOLY, P. A SAS macro for parametric and semiparametric mixture cure models. **Computer Methods and Programs in Biomedicine**, v. 83, n.2, p. 173-180, 2007.

EUDES, A.M.; TOMAZELLA, V.L.D.; CALSAVARA, V.F. Modelagem de sobrevivência com fração de cura para dados de tempo de vida weibull modificada. **Rev. Bras. Biom.**, São Paulo, v.30, n.3, p.326-342, 2012.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS ANÍSIO TEIXEIRA (INEP). **Resumo Técnico do Censo da Educação Superior 2011**. Disponível em: <<http://portal.inep.gov.br/web/centro-da-educacao-superior/resumos-tecnicos>> Acesso em: 14 set. 2013.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS ANÍSIO TEIXEIRA (INEP). **Educação Superior Brasileira 1991-2004: Paraná**. Brasília: Inep/MEC, 2006.

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

LOBO, M. B. C.; HIPÓLITO, O.; MOTEJUNAS, P. R; SILVA, R. L. L. **Evasão no Ensino Superior Brasileiro**. Disponível em: <<http://www.scielo.br/pdf/cp/v37n132/a0737132.pdf>>. Acesso em: 14 set. 2013.

MANTEL, N. Evaluation of survival data and two new rank order statistics arising in its consideration. **Cancer Chemotherapy Reports**, v. 50, p.163-170, 1966.

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**. 2011. 79p. Monografia (Trabalho de Conclusão do Curso de Estatística) – 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/>.

SAS INSTITUTE INC. **The SAS Enterprise Guide System, release 4.3**. Cary: NC, USA, 2006.

APÊNDICE A

APÊNDICE A1 – Comandos R para obtenção das curvas de sobrevivência por meio do estimador não-paramétrico de Kaplan-Meier.

```
base_1991 = read.table("Base_1991.txt",sep="\t",header=TRUE)
base_2006 = read.table("Base_2006.txt",sep="\t",header=TRUE)

attach(base_1991)
attach(base_2006)

# KAPLAN-MEIER GERAL

ekm = survfit(Surv(TEMPO_1991_RECASC,FORMATURA_CLASSE)~1)
plot(ekm,conf.int = F,xlab="Tempos",ylab="S(t) estimada",main="Kaplan-Meier |
Função de Sobrevivência")

# IDADE_CLASSE
ekm =
survfit(Surv(TEMPO_1991_RECASC,FORMATURA_CLASSE)~IDADE_CLASSE)
plot(ekm,lty=c(1,2,3,4),xlab="Tempos",ylab="S(t) estimada",main="Kaplan
Meier - Idade")
legend(1,0.6,lty=c(1,2,3,4),c("17 - 19","19 - 22","22 - 26","Maior de 26 anos"))

# IRA_CLASSE
ekm = survfit(Surv(TEMPO_1991_RECASC,FORMATURA_CLASSE)~IRA.CLASSE)
plot(ekm,lty=c(1,2,3,4),xlab="Tempos",ylab="S(t) estimada",main="Kaplan
Meier - IRA")
legend(1,0.8,lty=c(1,2,3,4),c("0 - 50","50 - 70","70 - 85","85 - 100"))

# SEXO_CLASSE
ekm =
survfit(Surv(TEMPO_1991_RECASC,FORMATURA_CLASSE)~SEXO_CLASSE)
plot(ekm,lty=c(1,2),xlab="Tempos",ylab="S(t) estimada",main="Kaplan Meier -
Gênero")
legend(1,0.6,lty=c(1,2),c("F","M"))

# FORMA_DE_INGRESSO_CLASSE
ekm =
survfit(Surv(TEMPO_1991_RECASC,FORMATURA_CLASSE)~FORMA_DE_INGRE
SSO...CLASSE)
plot(ekm,lty=c(1,2),xlab="Tempos",ylab="S(t) estimada",main="Kaplan Meier -
Forma de Ingresso")
legend(1,0.6,lty=c(1,2),c("Outros","Vestibular"))
```

APÊNDICE A2 – Comandos SAS para ajuste do modelo de mistura com evasão.

Obs: inicialmente é necessário carregar a macro PSPMCM.

<http://www.isped.u-bordeaux2.fr/recherche/biostats/FR-biostats-accueil.htm>

```
%pspmcm(DATA=DADOS_1991,ID=ID_PESSOA,CENSCOD=FORMATURA_CLAS
SE,TIME=TEMPO_1991_RECALC,
    VAR=
        IRA_1(IS,0)
        IRA_2(IS,1)
        IDADE_CONTINUA(I,..)
        VEST_1(S,1)
    ,INCPART= logit,
    SURVPART= Cox,
    PLOTFIT = Y,
    TAIL= zero, SUOMET= pl,
    MAXITER= 200, CONVCRIT= 1e-5, ALPHA= 0.05,
    FAST= Y,BOOTSTRAP= N, NSAMPLE= 2000, BOOTMET= ALL,
    GESTIMATE= Y,
    BASELINE=Y,
    SPLOT=Y);
```

run;

```
%pspmcm(DATA=DADOS_2006,ID=ID_PESSOA,CENSCOD=FORMATURA_CLAS
SE,TIME=TEMPO_1991_RECALC,
    VAR=
        SEXO_F(S, 1)
        IDADE_CONTINUA(I,..)
        IRA_1(IS,0)
        IRA_2(IS,0)
        VEST_1(S,1)
    ,INCPART= logit,
    SURVPART= Cox,
    PLOTFIT = Y,
    TAIL= zero, SUOMET= pl,
    MAXITER= 200, CONVCRIT= 1e-5, ALPHA= 0.05,
    FAST= Y,BOOTSTRAP= N, NSAMPLE= 2000, BOOTMET= ALL,
    GESTIMATE= Y,
    BASELINE=Y,
    SPLOT=Y);
```

run;