



MINISTÉRIO DA EDUCAÇÃO  
UNIVERSIDADE FEDERAL DO PARANÁ  
SETOR DE CIÊNCIAS EXATAS  
DEPARTAMENTO 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**

Projeto de Pesquisa apresentado à disciplina Laboratório de Estatística do Curso de Graduação em Estatística da Universidade Federal do Paraná, como requisito para elaboração do Trabalho de Conclusão de Curso.

Orientadora: Profa. Dra. Suely Ruiz Giolo

**CURITIBA**  
**2013**

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	1
<b>2 OBJETIVOS</b> .....	3
2.1 OBJETIVOS GERAIS .....	3
2.2 OBJETIVOS ESPECÍFICOS .....	3
<b>3 MATERIAL E MÉTODOS</b> .....	4
3.1 MATERIAL .....	4
3.1.1 Banco de Dados .....	4
3.1.2 Recursos Computacionais .....	5
3.2 MÉTODOS .....	5
3.2.1 Estimador Não-Paramétrico de Kaplan-Meier .....	5
3.2.1 Estimador Não-Paramétrico de Turnbull.....	6
3.2.1 Modelo de Mistura com Fração de Cura .....	6
<b>4 CRONOGRAMA DE ATIVIDADES</b> .....	7
<b>REFERÊNCIAS</b> .....	8

## 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, geram desperdícios sociais, acadêmicos e econômicos. No setor público, têm-se recursos públicos investidos sem o devido retorno. No setor privado, tem-se 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 - INEP (2006), muitas instituições criaram cursos de nivelamento, com o intuito de amenizar a taxa de evasão dos cursos de graduação. 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 aspecto de 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 projeto é a de dar continuidade ao estudo publicado por Martins e Rocha (2011), realizando análises estatísticas adicionais e complementares 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, serão 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 já com o PSE). O *status* desses alunos (formado, evadido, cursando etc.), será o que constar na data de 31 de dezembro de 2013 no Sistema de Informações para o Ensino (SIE) da UFPR.

## 2 OBJETIVOS

### 2.1 OBJETIVOS GERAIS

Efetuar análises estatísticas a fim de auxiliar na avaliação do Processo Seletivo Estendido (PSE), implantado desde 2006 no Curso de Estatística da UFPR com o objetivo de reduzir a taxa de evasão. Adicionalmente, comparar o tempo até a formatura dos alunos que ingressaram no período de 1991 a 2005 (anterior ao PSE) com os que ingressaram entre 2006 e 2010 (com o PSE).

### 2.2 OBJETIVOS ESPECÍFICOS

- Realizar uma análise descritiva dos dados para melhor entendimento das variáveis disponíveis para o estudo;
- Fazer uso do modelo de mistura com fração de cura apresentado por Corbière e Joly (2007), a fim de modelar o tempo até a formatura (em semestres) dos alunos que ingressaram entre 1991 e 2010. A data final de acompanhamento do *status* (formado, evadido, cursando etc.) desses alunos será 31 de dezembro de 2013;
- Comparar a taxa de evasão e o tempo até a formatura dos alunos que ingressaram no Curso de Estatística da UFPR nos currículos de 1991 (período de 1991 a 2005) e 2006 (período de 2006 a 2010), tendo em vista que a partir de 2006 o Curso de Estatística passou a adotar o Processo Seletivo Estendido (PSE);
- Obter conclusões a partir das análises realizadas e repassá-las à Coordenação do Curso de Estatística da UFPR.

## 3 MATERIAL E MÉTODOS

### 3.1 MATERIAL

#### 3.1.1 Banco de Dados

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: o primeiro vigorou de 1991 a 2005 e, o segundo, de 2006 a 2010 (este segundo coincidindo com o início do PSE). Quanto ao tempo de integralização previsto para esses dois currículos, este era de 4 anos e meio (9 semestres) para o primeiro currículo e de 4 anos (8 semestres) para o segundo.

A decisão em considerar 2010 como o último ano de ingresso para o estudo proposto neste projeto, 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 integralização do curso dentre os que ingressaram após 2010.

No geral, o banco de dados contém informações de aproximadamente 1600 alunos, sendo que 1300 ingressaram sob o currículo de 1991 (1991 a 2005) e 300 sob o currículo de 2006 (2006 a 2010). Durante o desenvolvimento do projeto serão verificadas algumas inconsistências no banco de dados; porém o total de alunos deve ser muito próximo do mencionado.

Dentre as informações (variáveis) disponíveis no banco de dados, tem-se: data de ingresso no curso; forma de ingresso no curso (vestibular, transferência, entre 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 integralização do curso (em semestres) será considerado como variável resposta no modelo de análise de sobrevivência. Para aqueles alunos em que a integralização do curso não tenha ocorrido até 31/12/2013 (evadidos, cursando etc.), o tempo considerado será o decorrido desde a data de ingresso até 31/12/2013. Em análise de sobrevivência, esses tempos são

denominados tempos censurados (COLOSIMO; GIOLO, 2006).

### 3.1.2 Recursos Computacionais

Para as análises e tratamento do banco de dados serão utilizados os recursos dos *softwares*: *R* versão 3.0 (R DEVELOPMENT CORE TEAM, 2013) e *Statistical Analysis System* versão 4.3 *SAS Enterprise Guide* (SAS INSTITUTE INC, 2012).

## 3.2 MÉTODOS

Em análise de sobrevivência a variável resposta é o tempo até ocorrência de um evento de interesse, usualmente denominado 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 que não apresentam o 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 à integralização 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, dentre outros motivos.

A seguir, são descritas brevemente as metodologias estatísticas as quais pretende-se utilizar para o desenvolvimento deste projeto.

### 3.2.1 Estimador Não-Paramétrico de Kaplan-Meier

Em um contexto exploratório, o estimador não-paramétrico de Kaplan-Meier (KAPLAN; MEIER, 1958) será utilizado para estimar a função de sobrevivência,  $S(t) = P(T > t)$ , com  $T$  denotando a variável tempo até a integralização do Curso de Estatística (formatura), bem como para obter estatísticas básicas de interesse tais como o tempo médio e o tempo mediano. Tal estimador também será utilizado para

estimar as curvas de sobrevivência associadas às categorias de cada covariável disponível, com o intuito de avaliar possíveis associações das covariáveis com a variável resposta. Para testar se as curvas diferem entre as categorias de cada covariável avaliada será utilizado o teste *logrank* (MANTEL, 1966).

### 3.2.2 Estimador Não-Paramétrico de Turnbull

O estimador de Turnbull (TURNBULL, 1976) é uma alternativa ao estimador de Kaplan-Meier na presença de dados de sobrevivência grupados. Tais dados decorrem de situações em que todas as unidades amostrais são avaliadas nos mesmos instantes de tempo (COLOSIMO; GIOLO, 2006), o que pode ocasionar um percentual considerável de tempos empatados.

Tendo em vista que as informações dos alunos sob estudo são atualizadas semestralmente, tal estimador poderá ser utilizado como uma alternativa ao estimador de Kaplan-Meier em função do percentual de tempos empatados que vier a ser observado nos dados objeto desse projeto.

### 3.2.3 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 e nem irá experimentar o evento de interesse. Tais indivíduos são denominados sobreviventes de longa duração, ou curados, ou simplesmente indivíduos imunes ao evento.

No contexto dos dados desse projeto, há indivíduos nessa condição; os alunos evadidos. Sendo assim, pretende-se utilizar para a análise desses dados, o modelo de mistura com fração de cura apresentado por Corbière e Joly (2007), levando-se em consideração a presença de dados de sobrevivência grupados.

## 4. CRONOGRAMA DE ATIVIDADES

ATIVIDADES	AGO/2013	SET/2013	OUT/2013	NOV/2013	DEZ/2013
<b>1 Projeto de Pesquisa</b>					
Definição do tema de estudo	■				
Definição do conjunto de dados e dos métodos estatísticos	■				
Elaboração e entrega do projeto de pesquisa ao orientador		■			
<b>2 Elaboração do Trabalho de Conclusão de Curso</b>					
Revisão de literatura sobre o tema	■	■			
Análise dos dados e discussão dos resultados obtidos		■	■	■	
Redação do trabalho de conclusão de curso			■	■	
Leitura do trabalho pelo orientador e correções				■	■
Entrega do trabalho redigido aos membros da banca					■
<b>3 Defesa do Trabalho de Conclusão de Curso</b>					
Preparação e apresentação do trabalho de conclusão de curso					■
<b>4 Elaboração da Versão Final do Trabalho de Conclusão de Curso</b>					
Elaboração da versão final do TCC					■
Entrega da versão final do trabalho ao orientador					■

## 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.

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, SAS Institute Inc., 2012.

TURNBULL, B. W. Nonparametric estimation of a survivorship function with doubly censored data. **Journal Royal Statistical Society, B**, v. 38, p. 290-295, 1976.