



2º EnGOPE

2º Encontro Goiano de Probabilidade e Estatística

18 a 20 de Novembro de 2015

<http://engope.mat.ufg.br>

Anais do 2º Encontro Goiano de Probabilidade e Estatística do IME/UFG

Goiânia/GO, 18 a 20 de Novembro

Apoio



Centro Editorial e Gráfico
Universidade
Federal de Goiás

Anais do 2º Encontro Goiano de Probabilidade e Estatística do IME/UFG

Goiânia/GO, 18 a 20 de Novembro

É com entusiasmo que anunciamos a 2 edição do EnGOPE (Encontro Goiano de Probabilidade e Estatística), promovido pelo Instituto de Matemática e Estatística da Universidade Federal de Goiás - IME/UFG, Campus Samambaia - Goiânia-GO .

Os avanços metodológicos da Estatística são frequentes, o que exige uma constante atualização do conhecimento científico de profissionais da área. Baseado nessa premissa, II Encontro Goiano de Probabilidade e Estatística é um evento científico motivado pela necessidade de propiciar aos alunos de graduação, pós-graduação e docentes do IME/UFG um contato com demais pesquisadores, docentes e profissionais liberais renomados da área de Probabilidade e Estatística.

Assim temos como objetivo, reunir a comunidade estatística do Estado de Goiás e do Brasil para discutir e difundir a pesquisa, novas ideias e técnicas estatísticas junto a pesquisadores, alunos e profissionais da Estatística. Sejam bem-vindos e aproveitem o evento!

Anais do 2º Encontro Goiano de Probabilidade e Estatística do IME/UFG

Goiânia/GO, 18 a 20 de Novembro de 2015

Comissão Científica

- Prof. Dr. Everton Batista da Rocha;
- Prof. Dr. Luis Rodrigo Fernandes Baumann;
- Prof. Dr. Renato Rodrigues Silva;
- Prof. Dr. Fabiano Fortunato Teixeira dos Santos;
- Prof. Dr. Helton Saulo Bezerra dos Santos;
- Prof. Ms. Márcio Augusto Ferreira Rodrigues;
- Prof. Ms. Marta Cristina Colozza Bianchi.

Comissão organizadora

- Prof. Dr. Everton Batista da Rocha;
- Prof. Dr. Luis Rodrigo Fernandes Baumann;
- Prof. Dr. Renato Rodrigues Silva;
- Prof. Dr. Fabiano Fortunato Teixeira dos Santos;
- Prof. Dr. Helton Saulo Bezerra dos Santos;
- Prof. Dr(a). Kélem Gomes Lourenço;
- Prof. Ms. Márcio Augusto Ferreira Rodrigues;
- Prof. Ms. Marta Cristina Colozza Bianchi.

Corpo Editorial

- Prof. Dr. Everton Batista da Rocha;
- Prof. Dr. Renato Rodrigues Silva;
- Prof. Dr(a). Kélem Gomes Lourenço;
- Prof. Ms. Márcio Augusto Ferreira Rodrigues.

Os artigos assinados são da responsabilidade dos autores.

É permitida a reprodução, desde que seja citada a fonte.

Sumário

Resumo Expandido

Ajuste Sazonal da Produção Industrial em Goiás: Uma Análise da Indústria de Transformação e suas Desagregações	2
Aplicação da Teoria das Filas no Restaurante cidadão	7
Teoria das Filas Aplicada ao Transplante de Córneas em Goiás	11
Inferência e Correção de Vício em Filas Markovianas de Servidor Único	14
Desempenho dos times que disputaram o campeonato brasileiro de futebol (Série A) de 2003 a 2014	19
Análise de Sobrevivência: Uma aplicação para a evasão universitária na UFG 2009	22
Casualidade do programa UFGInclui sobre o desempenho acadêmico usando propensity scoring matching	26
Distribuição Birnbaum-Saunders baseada no núcleo logístico e alguns problemas de inferência	29

Resumo Simples

Ajuste Sazonal da Produção Industrial em Goiás: Uma Análise da Indústria de Transformação e suas Desagregações	34
Aplicação da Teoria das Filas no Restaurante cidadão	35
Teoria das Filas Aplicada ao Transplante de Córneas em Goiás	36
Inferência e Correção de Vício em Filas Markovianas de Servidor Único	37
Desempenho dos times que disputaram o campeonato brasileiro de futebol (Série A) de 2003 a 2014	38
Análise de Sobrevivência: Uma aplicação para a evasão universitária na UFG 2009	39
Casualidade do programa UFGInclui sobre o desempenho acadêmico usando propensity scoring matching	40
Distribuição Birnbaum-Saunders baseada no núcleo logístico e alguns problemas de inferência	41

Apresentação Oral

Estatística e Big Data na indústria e agronegócio	42
Introdução ao mapeamento de QTL's.	43
Likelihood-based Inference for Mixed-Effects Models with Censored Response Using the Multivariate-t Distribution	44

AJUSTE SAZONAL DA PRODUÇÃO INDUSTRIAL EM GOIÁS: UMA ANÁLISE DA INDÚSTRIA DE TRANSFORMAÇÃO E SUAS DESAGREGAÇÕES

Alex Felipe Rodrigues Lima - *alexfrl@segplan.go.gov.br* - Instituto Mauro Borges - SEGPLAN

Sergio Borges Fonseca Júnior - *seergio-bj@segplan.go.gov.br* - Instituto Mauro Borges - SEGPLAN

Dinamar Maria Ferreira Marques - *dinamar-mfm@segplan.go.gov.br* - Instituto Mauro Borges - SEGPLAN

Mario Ernesto Piscoya Díaz - *mario.piscoya@gmail.com* - IME - UFG

O objetivo deste artigo é realizar o ajuste sazonal e obter o melhor modelo de ajuste para a indústria de transformação e em suas desagregações - indústria alimentícia, de minerais não-metálicos e de metalurgia - no estado de Goiás. O ajuste sazonal é importante, pois permite a comparação entre meses subsequentes, o que é imprescindível, para o melhor acompanhamento da evolução dos indicadores da indústria, sobretudo, em anos que ocorrem profundas mudanças na atividade econômica. Os resultados foram obtidos através do software R em conjunto com o X-13-ARIMA-SEATS. Verificou-se que as variáveis relacionadas ao calendário (carnaval, anos com maior números de dias trabalhados) e outliers exercem influencia significativa para o melhor ajuste do modelo.

1 Introdução

A análise de dados em séries temporais é um campo de vastas técnicas estatísticas. Uma das formas mais corriqueiras de se trabalhar com séries temporais é a sua decomposição em quatro componentes: tendência, ciclo, sazonalidade e resíduo. Esta decomposição permite verificar com clareza quais componentes são responsáveis por determinado comportamento de uma série temporal. As técnicas de decomposição de uma série temporal vêm se aperfeiçoando ao longo dos anos, permitindo a obtenção de informações cada vez mais fidedignas por meio de softwares estatísticos.

O ajuste sazonal é amplamente utilizado por diversos órgãos de estatística, tais como: IBGE, United States Bureau (U.S. Bureau), Eurostat, entre outros. Em referência as técnicas de ajuste sazonal merecem destaque o trabalho seminal de Shiskin et. al. (1967), onde o autor apresenta uma das primeiras incorporações entre as técnicas de dessazonalização e interface de um software, que foi responsável pelo surgimento da nomeada técnica X-11, amplamente utilizada no mundo inteiro. Dando sequência a estes estudos merecem destaque também Zellner (1979) e Plosser (1979), ambos autores aprofundaram na aderência entre softwares à técnicas para dessazonalização.

O presente trabalho tem como objetivo realizar uma aplicação dos modelos estatísticos para séries temporais sazonais, nas séries de dados econômicos do estado de Goiás. Analisou-se a série temporal da produção industrial do estado de Goiás, cujos dados foram obtidos da Pesquisa Industrial Mensal elaborada pelo Instituto Brasileiro de Geografia Estatística (PIM/IBGE), no período de janeiro de 2002 a dezembro de 2014. Os resultados indicam a presença da componente sazonal nos setores de indústria e alimentos, minerais não metálicos e transformação. Além disso, os resultados deste trabalho também sugerem um impacto negativo de datas como o *carnaval* na produção industrial daquele estado.

2 Metodologia

2.1 Dados

Os dados correspondem à Pesquisa Industrial Mensal Produção Física (PIM-PF) realizada no estado de Goiás no período 01.2002 - 12.2014 pelo IBGE. A variável em análise é o índice, de base fixa sem ajuste sazonal, da produção física industrial, focalizando esse indicador nos setores da indústria de transformação agregada e as suas desagregações relativas à fabricação de produtos alimentícios, metalurgia e minerais não-metálicos.

2.2 Métodos

A estratégia de análise inicia-se com uma análise exploratória da série, para verificar a presença de eventos atípicos (outliers). Seguidamente, foi analisada a presença do efeito calendário nos dados. Esse efeito está relacionado com o número de dias que podem afetar a componente sazonal tais como feriados, greves, etc). Na identificação da componente sazonal, será considerando um modelo de decomposição aditivo. Essa identificação será realizada pelo uso de métodos exploratórios e, posteriormente, confirmada a partir dos teste de Friedman (1937), Kruskal-Wallis e F. Tais testes permitem detectar com precisão a presença do comportamento sazonal numa série de dados.

Nas séries de dados econômicas, principalmente as do setor de produção, é comum a presença de outliers, efeito calendário e sazonal. Por essa razão, seguindo Ferreira, Gondim e Matos (2015) o modelo proposto para a produção industrial e Goiás é o seguinte:

$$Y(t) = \sum_{i=1}^{n_{out}} \omega_i \lambda_i(B) d_i(t) + \sum_{i=1}^{n_c} \alpha_i cal_i(t) + \sum_i^{n_{reg}} \beta_i reg_i(t) + X(t) \quad (1)$$

Onde $d_i(t)$ é a posição do i -ésimo outlier, $\lambda_i(B)$ é o polinômio em B que reflete o padrão dinâmico do outlier, $cal_i(t)$ é a variável do tipo de calendário, $reg_i(t)$ é a variável de intervenção, $X(t)$ representa um processo ARIMA(p,d,q) nos erros, w_i é o instante dos efeitos do i -ésimo outlier, α_i e β_i são os coeficientes do calendário e da variável de intervenção, respectivamente e,

por final, n_{out} , n_c e n_{reg} é o número total de variáveis inseridas em cada somatório. Esse modelo é conhecido na literatura como regARIMA.

A estimação dos parâmetros foi realizada nos softwares R e X-13 ARIMA-SEATS. Este último, possibilita a incorporação do efeito de calendário e outliers. Nos resíduos estimados, as suposições de normalidade, autocorrelação e variância constante foram avaliadas por meio dos testes de estatísticos de Shapiro-Wilks, Box-Ljung e Box respectivamente. A escolha do modelo final ajustado foi realizada com base nos critérios Bayesiano de Schwarz (BIC) e de Akaike corrigido (AIC_c).

A verificação da existência da componente sazonal será realizada por meio do diagnóstico QS, que tem por objetivo detectar a presença de sazonalidade nos dados. Em seguida, realiza-se uma análise gráfica comparando-se o comportamento da série do componente sazonal com a série agregada dos componentes irregulares e sazonais (identificada como SI ratio). Por fim, a partir da constatação de um ajuste sazonal de qualidade, serão realizadas análises gráficas dos resultados das séries ajustadas vis-à-vis as originais, em índice.

3 Resultados e discussão

Uma primeira forma de identificação da presença do componente sazonal nas séries do índice de produção da indústria de alimentos, minerais não-metálicos, metalurgia e transformação total, é verificar se há manifestação de comportamento similar (de alta, baixa, ou estabilidade) em um mesmo mês em diferentes anos. Nem sempre é possível identificar a componente sazonal por análises exploratórias gráficas. Contudo, as análises realizadas neste trabalho sugerem a presença de sazonalidade, no período sob análise, para os setores de Indústria de Alimentos, Metalurgia e Transformação. Os testes estatísticos realizados se mostraram significantes a 1%. Somente não foi possível estabelecer a presença da componente sazonal para o setor de minerais não-metálicos. A Figura 1 resume os modelos finais para cada um dos setores considerados. Em cada modelo foram considerados os efeitos dos outliers e de calendário. Em relação com este último, observa-se na tabela 1 que o carnaval tende a reduzir o volume de produção industrial em todas as séries investigadas. Também é possível observar naquela tabela que o efeito dias de semana, tende a aumentar a produção industrial nos setores de indústria de alimentos e transformação. Variáveis como *Pascoa*, e *ano bissexto* que também compõem o efeito calendário não foram estatisticamente significativas (p-valor >0,05).

A figura 2 compara o componente sazonal (FS) e agregação dos componentes sazonais e irregulares (SI ratio) Observa-se que em todos os setores analisados existe aderência entre os valores de ambas as componentes mencionadas. Esse resultado sugere que a componente irregular não exerceu nenhum efeito sob o ajuste sazonal da produção industrial em cada um dos setores analisados.

A figura 3 mostra que as séries com ajuste sazonal apresentam comportamento menos volátil vis-à-vis as séries originais, uma decorrência natural da retirada do efeito da sazonalidade das

Modelo	Indústria de Alimentos		Minerais não-metálicos		Metalurgia		Transformação	
	SARIMA(1 0 1)(0 1 1)		SARIMA: (1 1 1)(1 0 1)		SARIMA: (0 1 1)(0 1 1)		SARIMA: (0 1 0)(0 1 1)	
	Variável	Estimado	Variável	Estimado	Variável	Estimado	Variável	Estimado
Carnaval	-4,6829**	Carnaval	-5,25993**	Carnaval	-5,25993**	Carnaval	-4,65938***	
Constante	2,92563***	Páscoa[15]	-4,32437*	Páscoa[15]	-4,32437*	Ano bissexto	2,57864	
Ano bissexto	2,82825	AO2012,Jul	-29,62154***	AO2012,Jul	-29,62154***	Dias trabalhados	0,34866***	
Dias trabalhados	0,30767**	AR- não sazonal-01	0,54869***	AR- não sazonal-01	0,54869***	Páscoa[8]	-2,6844**	
Páscoa [1]	-2,87932*	AR- sazonal-12	0,93672***	AR- sazonal-12	0,93672***	AO2011, Aug	-12,15008***	
AO2010, Jan	-18,20537***	MA- não sazonal-01	0,89391***	MA- não sazonal-01	0,89391***	LS2012, Jan	-30,86734***	
LS2012, Jan	-11,30461***	MA- sazonal-12	0,68978***	MA- sazonal-12	0,68978***	AO2013, Dec	17,07952***	
AR- não sazonal-01	0,56368***					MA- sazonal-12	0,48744***	
MA- sazonal-12	0,67877***							
Box-Ljung(Auto correlação)	27,46		21,48		21,81		33,24	
Shapiro (Normalidade)	0,9942		0,985		0,9783		0,9833	
AICc	908,8		1016		997,3		879,1	
BIC	936,9		1039		1014		904,4	

Nível de Significância: 0 **** 0,001 *** 0,01 ** 0,05 * , 0,1 *

Figura 1: Ajuste Sazonal para o índice da Indústria de Transformação do estado de Goiás

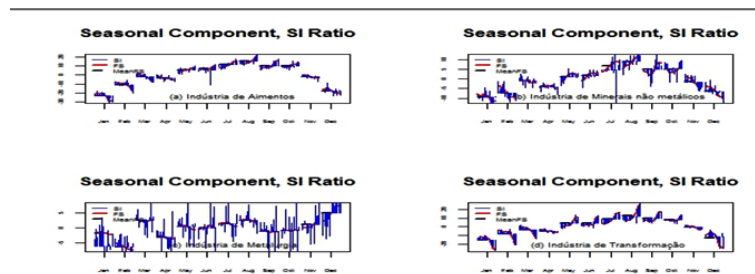


Figura 2: Comparação dos Componentes Sazonais (FS) e do SI Ratio para o modelo corrigido.

séries. Além disso, observa-se também uma tendência de crescimento da produção industrial a partir do ano de 2002 nos setores de indústria de alimentos, minerais não-metálicos e da indústria de transformação. No caso da indústria de metalurgia, observa-se na figura 3 a existência de uma mudança no nível médio da produção industrial nesse setor ocorrido aproximadamente no ano de 2006.

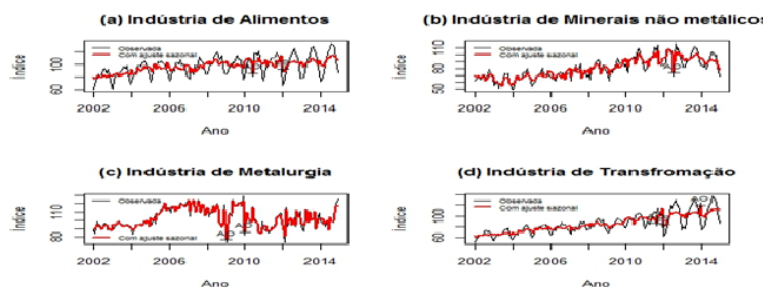


Figura 3: Comparação das séries originais (observada) com suas respectivas série com ajustes sazonal em índice.

4 Considerações finais

Este artigo teve o objetivo principal de criar condições para estabelecer comparações consistentes, em meses subsequentes, para dados que apresentam sazonalidade. Neste caso analisamos a indústria de transformação, alimentícia, de minerais não-metálicos e de metalurgia, no estado de Goiás. Vimos que esta comparação só é factível a partir do ajuste sazonal.

A comparação de meses subsequentes é importante, principalmente, nos momentos em que ocorrem profundas mudanças na atividade econômica, como, por exemplo, neste ano de 2015, em que a atividade industrial atravessa momento de claro descenso.

Assim, a comparação corriqueira, que consiste no confronto de um determinado mês com o mesmo do ano anterior, se feita no cenário vigente, tem mostrado quedas bastante acentuadas nas atividades industriais, pois no ano anterior as condições macroeconômicas eram amplamente mais favoráveis vis-à-vis este ano.

Dessa forma, em períodos de fortes mudanças econômicas a comparação dentro de um mesmo ano torna-se mais interessante, pois permite ao analista acompanhar o que de fato tem acontecido ao longo de um ano, sem o efeito do carregamento estatístico e sazonal do ano anterior.

Referências

- [1] ZELLNER, A. **Front matter to "seasonal analysis of economic time series**. The National Bureau of Economic Research, 1979.
- [2] BOX, G.;JENKINS G. **Time series analysis, forecasting and control** Editora Holgan Day, 1976 USA.
- [3] FERREIRA, P.;GONDIM, J.;MATTOS, D. Métodos de ajuste sazonal para séries de Business Tendency: um estudo de caso para a Sondagem da Indústria utilizando o método X13-ARIMA-SEATS **Fundação Getulio Vargas, Nota Técnica, Fevereiro de 2015**.
- [4] FRIEDMAN, M. The use of ranks to avoid the assumption of normality implicit in the analysis of variance **Journal of the American Statistical Association**. v.32, p. 675-701, 1937

APLICAÇÃO DA TEORIA DAS FILAS NO RESTAURANTE CIDADÃO

Ana Carolina do Couto Andrade - *carol.goiana@hotmail.com* - Instituto de Matemática e Estatística - IME

Fabiano Fortunato Teixeira dos Santos - *fortunato@ufg.br* - Instituto de Matemática e Estatística - IME

1 Introdução

À medida que a sociedade caminha para o imediatismo, aguardar em filas para ser atendido torna-se cada vez mais desgastante e incômodo. Ocasionalmente unicamente pelo fato de existir uma demanda maior do que a capacidade de atendimento do sistema, as filas estão presentes no cotidiano das pessoas, como em engarrafamentos, supermercados, bancos, hospitais e restaurantes ([6]).

Alguns estudos científicos têm sido realizados com o propósito de minorar esse contra-tempo, visto que as filas geram situações desfavoráveis para os usuários. Para isso é necessário explorar o arranjo, através de cálculos e análises que possibilitarão um planejamento adequado para o processo de atendimento. Dessa forma, a Teoria das Filas torna-se o pilar dessa pesquisa. Tal Teoria provê modelos que descrevem previamente o desempenho de um sistema que propõe serviços cuja procura cresce aleatoriamente, tornando possível encontrar um ponto de equilíbrio que satisfaça o cliente e seja viável economicamente para o provedor do serviço. Existem inúmeras aplicações da Teoria das Filas, em especial na área de prestação de serviços ([2],[3],[5],[8]).

Diante disso, este estudo tem por finalidade equacionar os indicadores do Sistema de Filas existente no Restaurante Cidadão, para que a posteriori seja possível dimensionar e planejar as instalações do local. Isto posto, busca-se mitigar o tempo de espera nas filas, levando em questão o interesse da empresa e de seus frequentadores.

2 Justificativa

O Restaurante Cidadão é um programa desenvolvido pela Organização das Voluntárias de Goiás (OVG), custeado com recursos do governo estadual. Tal programa, implantado em 2003, conta com oito unidades que servem diariamente 11.300 refeições ao preço de R\$ 2,00. Por mais rápido que o atendimento seja, a formação de filas é inevitável.

As unidades existentes estão localizadas em Goiânia (Centro e Campinas), Rio Verde, Luziânia (Jardim Ingá e Setor Estrela Dalva II), Anápolis (Centro e UEG) e Aparecida de Goiânia.

Paralelamente, em agosto deste ano, o Governador Marconi Perillo autorizou a abertura do Restaurante Comunitário de Valparaíso, que contará com o suporte da OVG ([4]).

Nesse contexto, é fácil ver a importância social do projeto, dado que garante segurança alimentar à população, com a oferta de um alimento de qualidade a baixo custo. “Em outro lugar eu não comeria por dois reais e ainda ganharia água, café e uma fruta. Além disso, cada dia é um cardápio. Como muito. Pensa se não aproveito!”, diz Maria Altina da Cruz, frequentadora assídua do restaurante ([7]).

3 Objetivos

De forma geral, o presente estudo tem por objetivo analisar a possibilidade de minimizar o tempo de espera na fila do Restaurante Cidadão de Goiânia, usando a Teoria das Filas como ferramenta. Além disso, estudar alguns conceitos importantes do mundo da Probabilidade/Estatística, como Cadeias de Markov, distribuições de probabilidade e Inferência.

4 Metodologia

Para entender performance de um sistema de filas foi efetuada uma ampla pesquisa bibliográfica em artigos, livros e periódicos. De início foram abordados diversos temas e tipos de filas ([5]). Posteriormente, houve um estudo fundamentado em trabalhos acerca de aplicações da Teoria das Filas em restaurantes ([3]).

Feito isso, através de observações in loco, serão coletadas as seguintes variáveis: quantidade de clientes na fila, no serviço e no sistema; tempo despendido pelos clientes na fila, no serviço e no sistema; capacidade do restaurante e da fila; número de servidores; disciplina da fila; horário de funcionamento, entre outras que se julgarem necessárias.

Após esse levantamento, os dados serão tabulados por meio da estatística descritiva. Por meio de planilhas eletrônicas, poderão ser feitos cálculos computacionais para a construção dos modelos ([1]). O modelo encontrado será capaz de simular, com algum grau de precisão, o comportamento da fila, e com isso serão feitas considerações relacionadas ao estudo.

5 Resultados

O principal resultado esperado é a modelagem eficaz do problema. Espera-se prever o comportamento da fila no Restaurante Cidadão de Goiânia, visando avaliar o modelo de filas e a conjuntura atual do sistema, viabilizando a elaboração de ações futuras, tais como o redimensionamento do *layout*, equipamentos e pessoal.

Embasado nos estudos feitos, propõe-se que as chegadas dos clientes se processam segundo uma distribuição de Poisson, e o atendimento segundo uma distribuição Exponencial. Portanto,

tem-se como hipótese o modelo M/M/c, no qual a fila apresenta regime FIFO (primeiro a entrar, primeiro a sair) e c canais de atendimento. Testes de aderência serão feitos para a validação do modelo sugerido.

6 Conclusões

A Teoria das Filas é um mecanismo útil para estimar a operacionalidade de um sistema. O tratamento dos dados e a análise dos resultados propicia concluir sobre o desempenho dos 3 segmentos do processo: logística-servidor-cliente, que tem tendências diferentes que devem ser agregadas em um resultado final. Este trabalho elucida a aplicação da Teoria das Filas em um sistema usual, salientando que é possível utilizar a Pesquisa Operacional no setor de serviços a fim de aperfeiçoar o sistema disposto pelo restaurante.

Referências

- [1] ANDRADE, E. L. **Introdução à Pesquisa Operacional: métodos e modelos para análise de decisões**. 3. ed. Rio de Janeiro: LTC, 2002.
- [2] BRUNS, R; SONCIM, S. P.; SINAY, M. C. F. Pesquisa Operacional: uma aplicação da Teoria de Filas a um sistema de atendimento. **Anais do ENEGEP**. 2001.
- [3] CARDOSO, N. C. J.; SILVA, J. U. B. Teoria das Filas: aplicação em um restaurante universitário. **Anais do CNMAC**. v.2, p. 594-595, 2009.
- [4] CARLOS, M. **Governador Marconi autoriza a abertura do Restaurante Comunitário de Valparaíso, graças a liderança de Lêda Borges**. Disponível em <<http://www.papodomarcelo.com.br/#!/Governador-Marconi-autoriza-a-abertura-do-Restaurante-Comunit%C3%A1rio-de-Valpara%C3%ADso-gra%C3%A7as-a-lideran%C3%A7a-d-L%C3%AAda-Borges/cu6k/55c38ff80cf26ee1088062da>>. Acesso em: 15 set. 2015.
- [5] CORDEIRO, J. C. A.; CORRÊA, V. A.; FILHO, J. M. A Teoria das Filas como ferramenta de apoio para área de planejamento da produção. **Anais do Congresso Internacional de Cooperação Universidade-Indústria**. 2012.
- [6] FOGLIATTI, M. C.; MATTOS, N. M. C. **Teoria de Filas**. Rio de Janeiro: Interciência, 2007.
- [7] SANTOS, E. **Restaurante Cidadão Completa 12 anos**. 2015. Disponível em <<http://www.ovg.org.br/post/ver/200269/restaurante-cidadao-completa-12-anos>>. Acesso em: 15 set. 2015.

- [8] SOARES, MAÍSA N. **Estudo de simulação para a minimização das filas em uma agência dos Correios do município de Viçosa.** Disponível em: <<http://www.saepto.ufv.br/Image/artigos/Artigo15.pdf>> Acesso: 20 set 2015.

Teoria das Filas Aplicada ao Transplante de Córneas em Goiás

Lucas Siqueira Cheim - *lucas_siqueira11@hotmail.com* - IME/UFG
Fabiano Fortunato Teixeira dos Santos - *fortunato@ufg.br* - IME/UFG

1 Introdução

A Teoria das Filas trata de problemas de congestionamento de sistemas nos quais a principal característica é a presença de “clientes” solicitando “serviços”. Assim, um sistema de filas é composto, basicamente, de elementos que querem ser atendidos em um posto de serviço e que devem esperar até que o posto esteja disponível [1].

Observa-se no campo da saúde uma intensa demanda por atendimento. A Teoria das Filas se mostra como uma ferramenta útil a fim de modelar as filas geradas nesse meio e reduzir os tempos de espera pelo atendimento.

O Brasil atualmente dispõe do maior programa público de transplantes do mundo, tendo realizado mais de 8500 transplantes em 2003 e tendo uma fila de espera de quase 60 mil pessoas no mesmo ano [4]. Segundo a Central de Notificação, Captação e Distribuição de Órgãos do Estado de Goiás (CNCDO/GO), “até o mês de agosto de 2013, foram realizados 656 transplantes no estado” e a fila de espera já conta com 3530 pessoas, dentre as quais, 2.976 esperam por um transplante de córnea [2].

Apesar da vasta gama de dados angariados pela CNCDO/GO, pouco se tem em produções acadêmicas a respeito de tais filas [4]. Tendo em vista tamanho desfalque, este estudo se mostra de suma importância para obtermos informações a respeito do tempo de espera na fila de transplante de córneas em Goiás, podendo em seguida ser aplicado às demais filas de espera por transplantes.

2 Objetivos

O principal objetivo deste estudo é modelar de forma eficiente a fila de espera para o transplante de córnea em Goiás, bem como analisar as variáveis que interferem no tempo de espera dos pacientes, contribuindo, assim, para a redução do mesmo. Desta forma, se faz necessário o estudo de certos conceitos da Probabilidade/Estatística tais como Cadeias de Markov, Teoria das Filas, Inferência Estatística etc ([1],[3]).

3 Métodos

Primeiramente houve a necessidade do estudo dos conceitos básicos para o entendimento dos sistemas de filas. Para tal, o uso de bibliografias específicas se fizeram indispensáveis ([1],[3]). Em seguida, foi feita a análise de artigos que mostraram a ampla aplicação da Teoria das Filas na modelagem de fenômenos físicos, biológicos, financeiros etc.

Em um segundo momento, será necessário encontrar um conjunto de dados que sirva de base para a modelagem. Os dados sobre as fila de espera por transplantes de córneas em Goiás serão coletados junto à CNCDO/GO, que se propôs a disponibilizá-los para este estudo. Desta forma, a leitura e estudo de artigos que correlacionam a Teoria das Filas ao caso das fila de espera por transplantes será indispensável ([4],[5],[6],[7]).

De posse do banco de dados, serão analisadas as informações por meio da estatística descritiva a fim de se obter, por exemplo, o tempo de espera dos pacientes pelo transplantes, o tamanho da fila, disciplina da fila etc. Para tal, tabelas no Excel se farão necessárias.

Assim, com a análise dos dados serão estimados valores para as seguintes variáveis:

λ - taxa média de chegada de pacientes para transplantes;

μ - taxa média de serviço, ou seja, o número de transplantes por unidade de tempo.

Obtendo tais variáveis será definido o modelo que melhor representa a fila de estudo. E, com as relações algébricas propostas nas bibliografias, poderemos estimar inúmeras variáveis interessantes no estudo de filas, tais quais:

$$N_s = \frac{\lambda}{(\lambda - \mu)} - \text{número esperado de pacientes no serviço de transplantes;}$$

$$W = \frac{N_s}{\lambda} - \text{tempo médio de espera total por transplantes;}$$

$$N_q = \left(\frac{\lambda}{\mu}\right) N_s - \text{número esperado de pessoas na fila de transplantes por dia.}$$

Encontrado o modelo e suas particularidades, deseja-se, também, realizar simulações e analisar como as variáveis (λ, μ) se relacionam ao tempo de espera na fila (por exemplo, aumentando a taxa de serviço ou chegada em 10%).

4 Resultados Esperados

Espera-se, de fato, que a fila seja bem representada. A exemplo de estudos anteriores, presume-se encontrar o modelo de fila M/M/s (“Markovian/Markovian/single model”), com o intervalo de chegada de pacientes e tempo de atendimento distribuídos exponencialmente, independente e identicamente distribuída (i.i.d). Então serão avaliadas simulações a fim de observar como as variáveis responderiam a melhorias ou sobrecargas no sistema ([4],[6]).

5 Conclusão

Com a sociedade em expansão, as filas têm se tornado cada vez mais comuns no dia a dia das pessoas. A Teoria das Filas tenta, por meio de análises matemáticas, equacionar os fenômenos que envolvem tais filas. No campo da saúde as aplicações são várias e, de posse das informações obtidas em estudos como este (que visam a redução do tempo de espera por parte dos pacientes), certamente muitas vidas são salvas. No contexto de transplante de órgãos a espera ainda é consideravelmente grande, podendo chegar a média de 11 anos por um transplante de rim [4]. Assim, considerando a carência de produções acadêmicas para fila de espera por transplantes, este estudo vem representar uma pequena parte do cenário de transplantes de órgãos em Goiás.

Referências

- [1] ANDRADE, E. L. **Introdução à pesquisa operacional: métodos e modelos para análise e decisões**. Rio de Janeiro: LTC. 2012. 202 p.
- [2] CENTRAL DE NOTIFICAÇÃO, CAPTAÇÃO E DISTRIBUIÇÃO DE ÓRGÃOS DO ESTADO DE GOIÁS. **Estatística**. Disponível em: <<http://www.saude.go.gov.br/index.php?idMateria=85308>>. Acesso em: 23 de out. 2015.
- [3] HILLIER, F. S.; LIEBERMAN, G. J. **Introdução à pesquisa operacional**. Porto Alegre: AMGH. 2012. 1028 p.
- [4] MARINHO, A. Um estudo sobre as filas para transplantes no Sistema Único de Saúde brasileiro. **Cad. Saúde Pública**, v. 22, n. 10, p.2229-2239, 2006.
- [5] MARINHO, A.; CARDOSO, S. S.; ALMEIDA, V. V. Disparidades nas filas para transplantes de órgãos nos estados brasileiros. **Cad. Saúde Pública**, v. 26, n. 4, p. 786-796, 2010.
- [6] MARINHO, A.; CARDOSO, S. S.; ALMEIDA, V. V. **Os transplantes de órgãos nos estados brasileiros**. 2007.
- [7] MARINHO, A. **A situação dos transplantes de órgãos no Brasil**. 2009. Disponível em:<<http://hdl.handle.net/11058/1564>>. Data de acesso 23.10.2015

Inferência e Correção de Vício em Filas Markovianas de Servidor Único

Luiz Carlos R. Pereira - *luizreal05@ufmg.br* - Universidade Federal de Minas Gerais

Márcio A. C. Almeida - *estatimarcio@gmail.com* - Universidade Federal do Pará

Frederico R. B. Cruz - *fcruz@est.ufmg.br* - Universidade Federal de Minas Gerais

Fernando L. P. Oliveira - *fernandoluz@iceb.ufop.br* - Universidade Federal de Ouro Preto

1 Introdução

As filas de espera estão muito presentes no nosso dia a dia, em situações tais como na lanchonete, no supermercado, no banco e no posto de gasolina. Comum a todos esses casos é que as filas ocorrem quando a procura por um determinado serviço é maior que a capacidade do sistema de prover este serviço. Conforme Wagner [6], você pode caracterizar um sistema de filas pelo seu processo de chegada, seu processo de saída, a disciplina da fila, seu mecanismo de serviço, entre outras formas. Essas características estão presentes em todos os sistemas de filas e são muito importantes para analisarmos cada tipo de fila.

Várias são as medidas que podem atestar o desempenho de um sistema de filas. Neste trabalho daremos foco à estimativa da intensidade do tráfego ρ (às vezes chamado de fator de utilização da estação de serviço), para o caso de filas $M/M/1$, que, na notação de Kendall [4], representa filas com tempos entre chegadas exponenciais (markoviano), tempos de serviço também exponenciais, servidor único e área de espera infinita.

Este artigo está organizado com se segue. Na Seção 2 é apresentada a metodologia utilizada. Os resultados computacionais e sua análise são discutidos na Seção 3. Finalmente, a Seção 4 encerra o artigo com conclusões finais e tópicos para futuras pesquisas na área.

2 Metodologia

2.1 Filas $M/M/1$

As filas $M/M/1$, apesar de serem as mais simples do ponto de vista matemático, estão entre aquelas com mais aplicações na prática. Elas aproximam diversas situações reais e são a base de modelos mais complexos. Segundo a notação de Kendall [4], são filas com:

$$\begin{aligned} \text{densidade exponencial para os tempos entre chegadas: } & f(t) = \lambda e^{-\lambda t}, \quad t \geq 0, \\ \text{densidade exponencial para os tempos de serviço: } & g(t) = \mu e^{-\mu t}, \quad t \geq 0. \end{aligned} \tag{1}$$

Assumindo que o sistema de filas atinge um regime estacionário (também denominado equilíbrio estatístico), isto é, que ele foi observado após um prolongado período de funcionamento, pode-se calcular a intensidade de tráfego ρ , definido como:

$$\rho \equiv \frac{\lambda}{\mu}, \quad (2)$$

que, para garantir que o sistema atinge o equilíbrio, deve obedecer a $\rho < 1$.

Associada a esta intensidade de tráfego, pode-se encontrar uma distribuição de probabilidade geométrica para o número de clientes no sistema no momento da partida, M , que, após o equilíbrio, pode ser escrita como:

$$P(M = m) = \begin{cases} \rho^m(1 - \rho), & m = 0, 1, 2, \dots, \\ 0, & \text{caso contrário.} \end{cases} \quad (3)$$

A partir da Eq. (3), importantes características de desempenho podem ser encontradas em função de ρ , tais como a probabilidade de servidor ocioso, $P(M = 0) = 1 - \rho$, a esperança do número de clientes no sistema, $L = \rho/(1 - \rho)$, e o tamanho médio da fila, $L_q = \rho^2/(1 - \rho)$.

2.2 Estimador de Máxima Verossimilhança para ρ

Para obtenção de uma estimativa para ρ , baseada em uma amostra \mathbf{X} de tamanho n , pode-se utilizar a Eq. (3) e definir-se a seguinte função de verossimilhança:

$$L(\rho, \mathbf{X}) = (1 - \rho)^n \rho^y, \quad (4)$$

em que $y = \sum_{i=1}^n x_i$.

Pode-se mostrar que o estimador de máxima verossimilhança (EMV), para ρ , é dado por:

$$\hat{\rho} = \frac{y}{n + y}. \quad (5)$$

2.3 Correção *Bootstrap* de Vício

Um método utilizado para correção de vício em estimadores é o *bootstrap* [3]. Na sua versão não paramétrica, para estimar o vício de um parâmetro de interesse, β , realizamos diversas reamostragens (com reposição) e estimamos o parâmetro para cada reamostragem. Tiramos então a média dessas estimativas, $\hat{\beta}_{(\bullet)}$ e o vício pode ser estimado da seguinte forma:

$$\text{vício} = \hat{\beta}_{(\bullet)} - \hat{\beta}. \quad (6)$$

A versão corrigida do estimador, pelo método *bootstrap* não paramétrico, é dada por:

$$\tilde{\beta}_B = 2\hat{\beta} - \hat{\beta}_{(\bullet)}. \quad (7)$$

Este método foi usado por diversos pesquisadores no passado, com resultados bastante satisfatórios (ver, por exemplo [1]). Trabalhos recentes mostram o uso do método *bootstrap* para estimação de intervalos de confiança (ver [2]). A seguir, são apresentados resultados computacionais e as principais conclusões obtidas.

3 Resultados e Discussão

Todos os algoritmos anteriormente descritos foram codificado em R [5] e estão disponíveis a pedido diretamente com os autores. Inicialmente foram realizadas simulações de amostras com intensidades de tráfego $\rho \in \{0,01; 0,10; 0,20; 0,50; 0,90; 0,99\}$ e tamanhos $n \in \{5; 10; 20; 50; 100; 200; 500\}$. Para todos esses cenários foram calculadas estimativas médias de ρ pelo EMV, Eq. (5), com os respectivos erros-padrão da média (EPM), para 10.000 replicações Monte Carlo. Os resultados podem ser vistos na Tabela 1 e nas Figuras 1 e 2.

Tabela 1: Estimativas médias de ρ com respectivo erro padrão entre parênteses

ρ	n						
	5	10	20	50	100	200	500
0.01	0,0080 (0,0004)	0,00910 (0,00029)	0,00968 (0,00021)	0,00987 (0,00014)	0,00994 (0,00010)	0,01000 (0,00007)	0,010025 (0,000044)
0.10	0,0849 (0,0011)	0,09279 (0,00084)	0,09668 (0,00062)	0,09820 (0,00039)	0,09940 (0,00028)	0,09953 (0,00020)	0,100027 (0,000127)
0.20	0,1725 (0,0015)	0,18718 (0,00109)	0,19372 (0,00079)	0,19710 (0,00050)	0,19892 (0,00035)	0,19923 (0,00025)	0,199806 (0,000160)
0.50	0,4528 (0,0017)	0,47665 (0,00116)	0,48838 (0,00081)	0,49509 (0,00050)	0,49801 (0,00035)	0,49868 (0,00025)	0,499603 (0,000158)
0.90	0,8784 (0,0006)	0,88979 (0,00036)	0,89500 (0,00023)	0,89801 (0,00014)	0,89895 (0,00010)	0,89940 (0,00007)	0,899811 (0,000043)
0.99	0,9875 (0,0001)	0,98887 (0,00004)	0,98945 (0,00002)	0,98978 (0,00001)	0,98990 (0,00001)	0,98994 (0,00001)	0,989976 (0,000004)

Podemos constatar que o EMV realmente apresenta vício, para amostras pequenas e moderadas. Para amostras grandes ($n \geq 200$), o comportamento é bastante satisfatório, conforme visto na Figura 1, que representa graficamente os valores apresentados na Tabela 1. Portanto, as medidas de desempenho da fila estimadas a partir de ρ também apresentarão vício, o que deve ser uma preocupação do pesquisador. Outra conclusão importante é que os EPM são maiores para intensidades de tráfego $\rho \approx 0,50$ e menores para intensidades extremas, $\rho \approx 0,99$ e $\rho \approx 0,01$, o que sugere que estes são os valores mais difíceis de estimar na prática. A Figura 2 mostra claramente este fenômeno.

Além disso, testes preliminares foram realizados com a versão corrigida do EMV para ρ , Eq. (5). Os valores médios obtidos, para diversos tamanhos de amostra, para 10.000 replicações Monte Carlo e $\rho = 0,5$, são apresentados na Figura 3. É possível notar pelos resultados que, com um esforço computacional extra, o pesquisador pode alcançar com amostras de tamanho $n = 5$ estimativas para ρ com o mesmo erro médio que para amostras de tamanho $n = 20$. Esta redução pode ser bastante importante do ponto de vista prático, pela diminuição que pode acarretar nos custos e no tempo para obtenção das amostras.

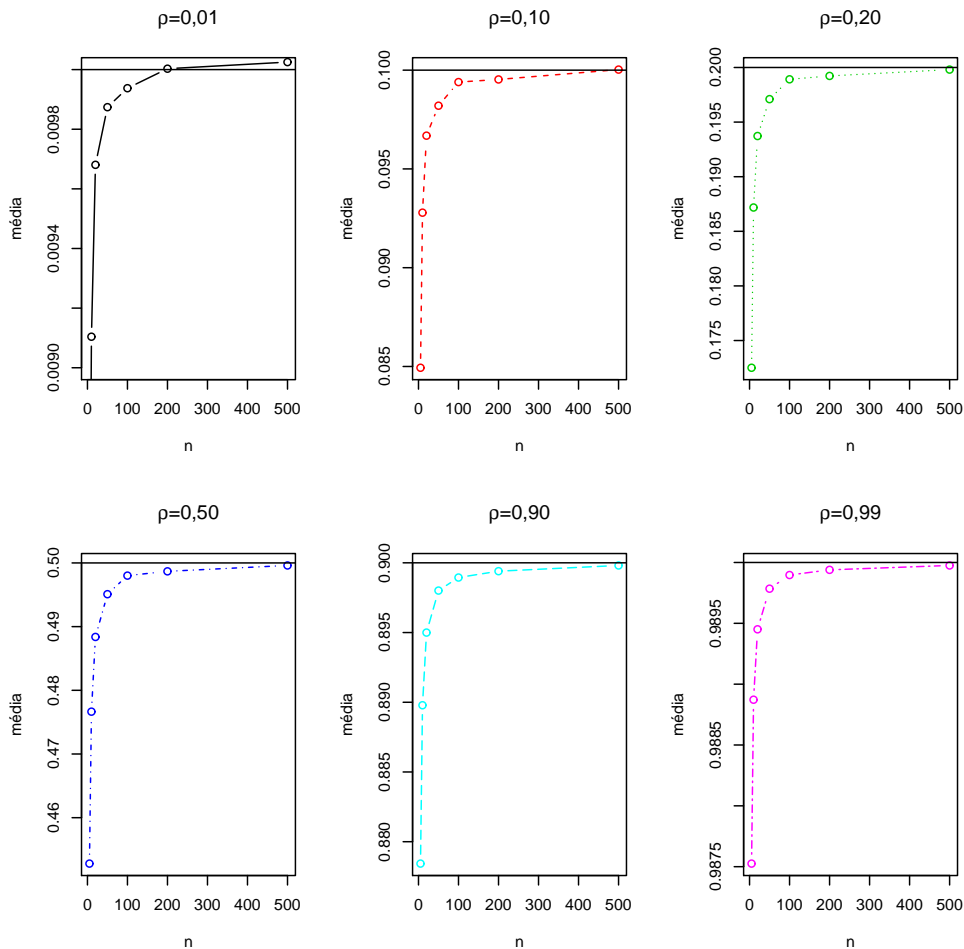


Figura 1: Estimativas médias de ρ em função do tamanho da amostra n

4 Conclusões e Observações Finais

Dada a importância das filas no dia a dia das pessoas, este trabalho abordou o problema de estimação de um dos seus parâmetros, a intensidade de tráfego ρ . O foco foi o tipo mais simples de filas, $M/M/1$, com chegadas e partidas markovianas e servidor único, que é uma aproximação acurada para vários problemas práticos e base para modelos mais completos e abrangentes. Foi constatada a presença de vício no estimador de máxima verossimilhança para ρ e investigada a utilização do conhecido método *bootstrap* não paramétrico para sua correção. Resultados preliminares promissores apontaram para uma correção efetiva do vício, possibilitando redução no tamanho de amostra, com conseqüente diminuição nos custos e tempo para sua obtenção.

Investigações futuras incluem a avaliação da qualidade da correção pelo *bootstrap* em outras medidas de desempenho da fila, tais como a probabilidade de sistema vazio, no tamanho médio da fila e número médio de usuários no sistema. Também outros tipos de filas devem ser investigados, tais como filas markovianas infinitas multi-servidoras, $M/M/c$, filas finitas, $M/M/c/K$, e assim por diante.

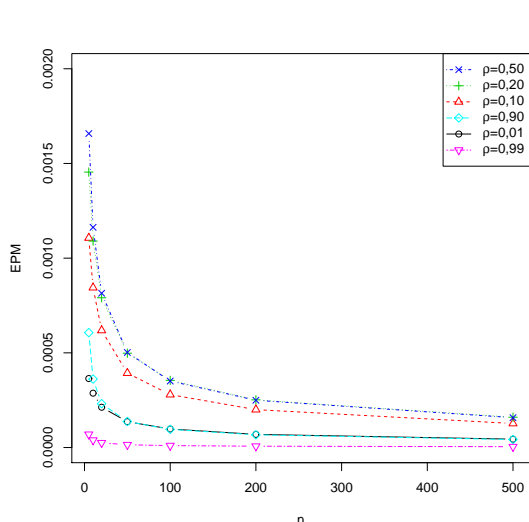


Figura 2: Erro-padrão das estimativas médias

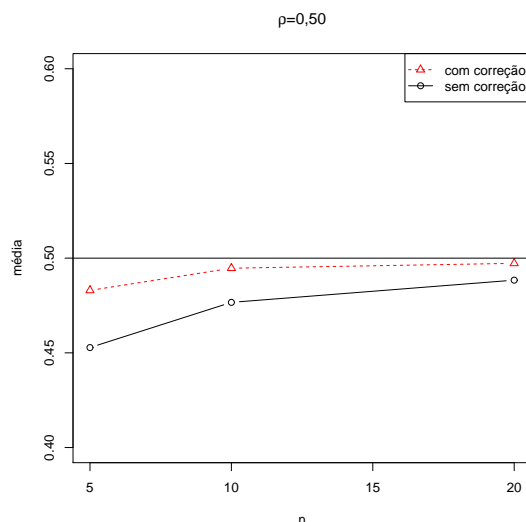


Figura 3: Correção das estimativas médias

Agradecimentos

Os autores agradecem ao CNPq e à Fundação de Amparo à Pesquisa do estado de Minas Gerais (FAPEMIG), pelo auxílio financeiro a esta pesquisa, bem como à UFOP e à UFPA.

Referências

- [1] CRUZ, F. R. B.; COLOSIMO, E. A.; MACGREGOR SMITH, J. Sample Size Corrections for the Maximum Partial Likelihood Estimator. **Communications in Statistics - Simulation and Computation**. Taylor & Francis, v. 33, n. 1, p. 35-47, 2004.
- [2] DOMINGUES, K. M.; OLIVEIRA, F. L. P.; CRUZ, F. R. B.; BESSEGATO, L. F. Estimação de Intervalos de Confiança via Reamostragem *Bootstrap*. **I Encontro Interacadêmico de Engenharia de Produção - EINEPro 2015**. São João da Barra, RJ, p. 1-12, 2015.
- [3] EFRON, B., TIBSHIRANI, R. **An introduction to the bootstrap**. London, UK: Chapman e Hall, 1993. 456 p.
- [4] KENDALL, D. G. Stochastic processes occurring in the theory of queues and their analysis by the method of imbedded Markov chains. **Annals of Mathematical Statistics**. Institute of Mathematical Statistics, v. 24, p. 338-354, 1953.
- [5] R CORE TEAM. **R: A Language & Environment for Statistical Computing**. R Foundation for Statistical Computing, Vienna, Austria, 2013. <http://www.R-project.org/>
- [6] WAGNER, H. M. **Pesquisa Operacional**. Prentice-Hall do Brasil Ltda., Rio de Janeiro, 2 ed., 1985. 851 p.

Desempenho dos times que disputaram o campeonato brasileiro de futebol (Série A) de 2003 a 2014

Nathalia Rodrigues Damasceno e-mail: *nathaliadamascenoufg2@gmail.com* IME-UFG

Valdivino Vargas Júnior e-mail: *vvjunior@gmail.com* IME-UFG

1 Introdução

O campeonato brasileiro de futebol vem sendo disputado anualmente desde 1959. Logo no início o campeonato teve um problema de padronização nas disputas, o que resultou em mudança de regras e número de clubes participantes em cada ano. De 1959 até 1968 era disputada a Taça Brasil, primeira competição nacional oficial disputada entre clubes, utilizando o sistema da copa. De 1967 a 1970 era disputado o chamado Torneio Roberto Gomes Pedrosa, sendo que entre 1968 e 1970 foi organizado pela CBD (precursora da atual CBF), que também utilizava o sistema da copa [1].

Em 1971 que o campeonato brasileiro propriamente dito começou a ser disputado, mas em 2010 fez-se uma unificação com a Taça Brasil e o Torneio Roberto Gomes Pedrosa comentados anteriormente [1].

Em 2003 o campeonato brasileiro começou a ser disputado segundo o sistema de pontos corridos, além desse formato já foi utilizado o sistema da copa e eliminatórias [1]. O interesse deste trabalho é analisar somente o sistema de pontos corridos (2003 a 2014). Cada clube enfrenta duas vezes os demais, sendo uma vez em seu estádio (em casa) e uma vez no estádio de seu adversário. Três pontos são atribuídos a cada vitória, um ponto a cada empate e a derrota não resulta em pontos, entretanto os clubes também podem perder pontos por irregularidades. Por fim esses pontos são somados e o clube que obtiver o melhor resultado recebe o título de campeão brasileiro do ano.

A maior parte dos brasileiros tem o seu carinho por um ou mais clubes em especial. Quando se discute o desempenho de clubes fica evidente o benefício agregado a um clube com maior preferência, ou fama que outros.

A análise proposta aqui é isenta de qualquer fanatismo por qualquer que seja o clube estudado, assim as análises serão baseadas somente nos resultados oficiais de cada campeonato, utilizando técnicas estatísticas para obter conclusões sobre os desempenhos dos clubes.

2 Dados e métodos

Vamos analisar o desempenho acumulado ano após ano dos clubes que já disputaram pelo menos uma vez o campeonato brasileiro na série A. Essa análise será feita através de cluster

(Análise de Agrupamentos). Assim se torna possível definir grupos segundo o desempenho de cada clube, os grupos formados se dividirão em clubes com melhores desempenhos, piores desempenhos e clubes com desempenho mediano. Verificando o desenvolvimento dos clubes ao longo dos anos.

Foram selecionadas oito variáveis para essa análise, sendo três delas referentes à classificação, três referentes à pontuação e duas referentes aos gols. As variáveis referentes à classificação são: posição média (soma das posições que o clube foi classificado dividido pelo número de campeonatos disputados até o ano analisado), número de títulos recebidos até o ano analisado e ranking de colocação (olha-se as classificações dos clubes no ano analisado, se o clube terminou o campeonato em primeiro lugar então ele recebe 10 pontos, se terminou em segundo recebe 9 pontos e assim sucessivamente até que o décimo colocado receba 1 ponto e os demais recebem zero).

Sobre a pontuação teremos: aproveitamento (pontos ganhos dividido por três vezes o número de jogos disputados até o ano analisado), número de campeonatos disputados até o ano analisado e número de vitórias por partida (soma dos números de vitórias dividido pelo número de jogos disputado até o ano analisado).

E por fim, as variáveis referentes aos gols: gols pró por partida (soma do número de gols pró dividido pelo número de partidas disputadas até o ano analisado) e saldo de gols por partida (soma dos gols pró e gols contra dividido pelo número de partidas disputados até o ano analisado).

Para concluir quais os clubes que possuem melhores desempenho vamos utilizar técnicas de agrupamento com as oito variáveis já mencionadas.

A análise de agrupamento (cluster) consiste em agrupar objetos segundo suas características, com isso criam-se grupos homogêneos, ou seja, as observações que pertencem ao mesmo grupo são similares e signi-

cativamente diferentes das demais observações dos outros grupos. Neste trabalho utilizamos o métodos hierarquicos de Ward. Esse método consiste em agrupar elementos de modo que a medida de homogeneidade interna seja mínima. Para mais informações consultar [2], [3] ou [4]

3 Conclusão

O principal objetivo deste trabalho era analisar o desempenho ao longo dos anos de cada clube que disputou pelo menos uma vez o campeonato brasileiro na série A. A cada ano acumulado podemos ver a melhora ou piora do clube em relação ao seu desempenho.

Alguns clubes tiveram uma evolução muito boa enquanto outros só tiveram uma ou duas chances de disputar o campeonato, isso fez com que seu desempenho não pudesse obter melhora ao longo do tempo.

Em 2014 vimos que o Cruzeiro e o São Paulo obtiveram resultados muito bons ao longo dos anos. Se observarmos os dados, vemos que em média o São Paulo e o Cruzeiro terminaram o

campeonato na quarta e sexta colocação, respectivamente. Além disso, cada um deles obteve três títulos de campeão brasileiro entre os anos de 2003 e 2014.

O Internacional, Corinthians e Santos, também na maioria das vezes obtiveram boas colocações ao longo dos anos.

Uma evolução bem interessante foi a do Grêmio. Em 2003, o Grêmio obteve um desempenho bem ruim fazendo com que ele ficasse no grupo de péssimo desempenho. Em 2004 ele foi rebaixado e continuou no grupo dos piores. Em 2006 o Grêmio foi resgatado da série B e obteve um excelente desempenho, isso fez com que ele fosse para o grupo de desempenho ruim, ou seja, dois grupos a frente do que ele estava anteriormente. Em 2007 ele passou para o grupo de desempenho regular e permaneceu neste grupo até 2009. Em 2010 o Grêmio passou para o grupo de bom desempenho e permaneceu neste grupo até 2014.

O Palmeiras e o Goiás na maioria das vezes obtiveram resultados bons ao longo desses anos, porém em algumas ocasiões foram rebaixados. Como esses clubes obtiveram bons resultados ao longo dos anos o rebaixamento não comprometeu tanto seu desempenho acumulado.

Atlético Paranaense e Flamengo em 2003 pertenciam ao grupo com desempenho ruim, Atlético Mineiro em 2003 pertencia ao grupo de desempenho regular e Fluminense pertencia ao grupo de péssimo desempenho. Em 2004 o Atlético Paranaense subiu para o grupo com desempenho bom, enquanto o Flamengo, Atlético Mineiro e Fluminense estavam no grupo com desempenho ruim. E ao longo dos anos eles vieram se desempenhando melhor fazendo com que em 2014 eles chegassem aos grupos com melhores desempenhos.

Referências

- [1] **História do campeonato brasileiro (PARTE 1)**. Disponível em:<http://www.campeoesdofutebol.com.br/brasileiro_historia.html>. Acesso em: 18 de março de 2015.
- [2] Johnson, A.R.; Wichern, D.W. **Applied Multivariate Statistical**. ed. 6
- [3] Linder, R. **Técnicas de Agrupamento**. Revista de Sistemas da Informação da FSMA. v.4, p. 18-36, 2009
- [4] Rencher, A.C. **Methods of Multivariate Analysis**. ed. 2

ANÁLISE DE SOBREVIVÊNCIA: Uma aplicação para a evasão universitária na UFG, 2009

Nathalia Rodrigues Damasceno e-mail: *nathaliadamascenoufg2@gmail.com* IME-UFG

Mario Ernesto Piscoya Diaz e-mail: *mario.piscoya@gmail.com* IME-UFG

Helton Saulo Bezerra dos Santos e-mail: *heltonsaulo@gmail.com* IME-UFG

1 Introdução

A educação superior é um componente do capital humano essencial para o desenvolvimento dos países. Evidências observadas nos atuais países desenvolvidos mostraram que as mudanças realizadas nos sistemas educativos desses países, com o objetivo de torná-los mais eficientes na melhora de seus processos de produção, têm contribuído positivamente na melhora das condições de vida dessas populações [8].

A evasão universitária é um problema importante nas universidades públicas brasileiras, pois está relacionada com uma série de impactos negativos na formação de recursos humanos em uma sociedade assim como individuais (depressão, frustração, etc).

A literatura internacional aponta uma série de fatores socioeconômicos e demográficos que encontram-se associados com a evasão como, por exemplo, o nível de escolaridade dos país ou responsáveis, a renda domiciliar, o sexo do individuo entre outros [9], [6], [5], [4], [3] e [1].

Identificar os fatores associados com a evasão universitária é uma dos objetivos principais dos formuladores de política educacional posto que o entendimento da relação desses fatores com a evasão pode ser de grande utilidade para a implementação de ações afirmativas que permitam prevenir a evasão nas instituições de ensino superior brasileiras. Sendo assim, este trabalho procura responder a seguinte pergunta: Quais são as condições socioeconômicas dos alunos que estão associadas com a evasão?

2 Dados

Os dados foram cedidos pela Pró-Reitoria de Graduação da UFG (PROGRAD), correspondentes a 5.108 estudantes de graduação, que ingressaram e se matricularam na Universidade Federal de Goiás no primeiro semestre do ano de 2009. Os alunos foram acompanhados desde o dia de ingresso a universidade até o meio do ano de 2014. O evento de interesse é se o aluno evadiu do curso durante o período de estudo.

As informações socioeconômicas e demográficas foram obtidas dos questionários socioeconômicos preenchidos pelos discentes durante sua inscrição no processo seletivo da universidade.

Para efeitos deste trabalho, foram considerados alunos evadidos àqueles que: reprovaram por média/falta em todas as disciplinas por 2 semestres consecutivos; reprovaram por média/falta em todas as disciplinas no semestre de ingresso; se transferiram para outra IES; reprovaram por falta 3 vezes na mesma disciplina e reprovaram por falta em todas as disciplinas por 2 semestres consecutivos; reprovaram por falta em todas as disciplinas por 2 semestres consecutivos; não efetuaram a matrícula; optaram por outro curso ou desistiram do curso.

3 Métodos

Inicialmente verificamos a associação das condições socioeconômicas dos alunos com a evasão. Para isto, fez-se um teste qui-quadrado de associação para tabelas de contingência ([2]).

Posteriormente, a partir dos dados da coorte de alunos ingressantes na Universidade Federal de Goiás, no ano de 2009 realizamos uma análise de sobrevivência, considerando como variável principal o tempo desde a primeira matrícula até a evasão.

E por fim, foi feita a comparação das experiências de evasão entre os discentes usando testes de hipótese da família Tarone-Ware. A hipótese nula deste teste é a inexistência de diferenças nas experiências de evasão dos discentes segundo alguma característica socioeconômica. Os resultados serão considerados significativos para um erro de tipo I igual a 0,05.

Maiores detalhes sobre o procedimento de estimação das curvas de sobrevivência e aplicação dos testes podem ser encontrados em [7].

4 Resultados

Os resultados obtidos a partir dos testes de associação Quadrado sugerem a existência de uma associação estatística entre a evasão e todas as características socioeconômicas consideradas neste trabalho (resultados não mostrados).

Após estimar as curvas de sobrevivência pelo método de Kaplan-Meier e compará-las através dos testes de hipótese da família Tarone-Ware, observam-se a existência de diferença significativa entre a evasão dos discentes e todas as características socioeconômicas consideradas, por exemplo, observamos que alunos unidos (ou alguma vez unidos) se evadem em menor tempo em comparação com discentes solteiros ($p\text{-valor} < 0,01$). Um resultado similar é observado quando é feita a comparação da evasão entre aqueles que trabalham e que não trabalham. Quem trabalha se evade mais rápido ($p\text{-valor} < 0,01$).

Os resultados sugerem diferenças na evasão por sexo, observando-se um menor tempo de evasão para os homens ($p\text{-valor} < 0,01$). Discentes que cursaram o ensino público em escolas públicas apresentam uma sobrevida de evasão menor comparados com discentes de escolas particulares ($p\text{-valor} < 0,01$). De forma similar, quem não realizou curso pré-vestibular apresenta

um tempo de evasão menor em comparação com quem sim fez tal curso (p-valor $< 0,01$).

Observamos a existência de diferenças no tempo de evasão segundo a escolaridade dos pais. Por exemplo, discentes cujos pais não tem nenhuma escolaridade, ou tem nível fundamental, apresentam um tempo de evasão menor em comparação com aqueles cujos pais têm nível médio ou superior (p-valor $< 0,01$). Um resultado similar é observado quando se analisa a evasão por escolaridade da mãe (p-valor $< 0,01$).

Não observamos diferenças na evasão segundo a raça. No entanto, observa-se que discentes auto-declarados como de raça indígena se evadem mais rápido em comparação com discentes de outras raças (p-valor $< 0,01$).

Em relação com a renda, observa-se que estudantes que recebem até 2 salários mínimos se evadem mais rápido em comparação com aqueles que não possuem nenhuma renda (p-valor $< 0,01$).

5 Conclusão

Segundo o estado civil, os estudantes unidos (ou alguma vez unidos) são aqueles que se evadem em menor tempo. Esta evasão pode ser explicada pelas responsabilidades adicionais que ter uma família, tais como a realização de uma atividade remunerada e o cuidado dos filhos.

A condição socioeconômica do domicílio parece estar associada com a evasão. Os nossos resultados identificaram que estudantes de escola pública, que trabalham, que não realizaram curso pré - vestibular e cujos pais têm um baixo nível de escolaridade se evadem em um período de tempo mais curto. A renda familiar, que está associada com o grau de escolaridade dos pais, é um dos determinantes para a escolha da escola de uma criança. Famílias com menor renda optaram pelo sistema de ensino público. Esse sistema apresenta uma série de deficiências em comparação com o ensino particular. A má formação escolar pode resultar em uma frustração acadêmica do indivíduo durante os primeiros anos na universidade e fazer com que ele se evada.

A baixa renda também força o indivíduo, uma vez que saiu do ensino médio, a procurar uma atividade remunerada, com o intuito de contribuir nas despesas do domicílio. Uma hipótese levantada para explicar a evasão daqueles discentes que trabalham é que esta se encontraria associada com a quantidade de tempo livre disponível, fora de sala de aula, destinado à realização das atividades da universidade. Pesquisas adicionais nesse tema devem ser realizadas para identificar diferenças na evasão de estudantes que trabalham, segundo o número de horas disponíveis para desenvolver aquelas atividades.

Não foi possível estabelecer diferenças na evasão segundo a raça do candidato. No entanto, esses resultados devem ser interpretados de forma cuidadosa pelo fato de que a informação da raça é auto-declarada, podendo essa declaração mudar em um curto período de tempo.

Espera-se que os resultados obtidos neste trabalho sejam de utilidade para a redução da evasão universitária nesta faculdade.

Referências

- [1] Abbott, R. D.; Hill, K. G.; Catalano, R. F.; Hawkins, J. D. **Predictors of Early High School Dropout: A test of Five Theories.** Journal of Educational Psychology. v.92, p. 568-582, 2000
- [2] Agresti, A. **An Introduction to Categorical Data Analysis.** John Wiley & Sons. v.2. 2007
- [3] Arulampalam, W.; Naylor, R. A.; Smith, J. P. **A hazard model of probability of medical school drop-out in the UK.** Journal of the Royal Statistical Society Series A. v.167, p. 157-178, 2004
- [4] Bradley, S.; Lenton, P. **Dropping out of post-compulsory education in the uk: an analysis of determinants and outcomes.** Journal of Population Economics. v.20, p. 299-328, 2007
- [5] DeJardins, S. L.; Ahlburg, D. A.; McCall, B. P. **An event history model of student departure.** Economics of Education Review. v.18, p. 375-390, 1999
- [6] Lassibille, G. and Gómez, L. N. **Why do higher education students drop out? evidence from Spain.** Why do higher education students drop out? evidence from Spain. v.16, p. 89-105, 2008
- [7] Kleinbaum, D. **Survival Analysis. A self-learning text** Springer, 1995
- [8] Montmarquette, C.; Mahseredjian, S.; Houle, R. **The determinantes of university dropouts: a bivariate probability model with sample selection.** Economics of Education Review. v.20, p. 475-484, 2001
- [9] Murtaught, P. A. and Burns, L. D. and Schuster, J. **Predicting the retention of university students.** Predicting the retention of university students. v.40, p. 355-371, 1999

Casualidade do programa UFGInclui sobre o desempenho acadêmico usando propensity scoring matching

Navarro Mendes Santos Rosa - *navarro_msr@hotmail.com* - Universidade Federal de Goiás

Mario E. Piscoya Díaz - *mario.piscoya@gmail.com* - Universidade Federal de Goiás

1 Introdução

A partir do ano de 2000, as instituições de ensino superior brasileiras, através de uma série de políticas destinadas a reduzir as desigualdades sociais observadas na população. Estas políticas consistem na reserva de uma quantidade de vagas destinadas para o ingresso de indivíduos de grupos socioeconômica e culturalmente vulneráveis. Um exemplo deste tipo de ações é o caso das chamadas "cotas raciais", onde candidatos autodeclarados como sendo de raça negra e aqueles procedentes das comunidades indígenas possui um determinado número de vagas reservadas, a fim de facilitar sua entrada na instituição de ensino superior. Um outro exemplo é o reconhecimento por parte do do governo federal das desigualdades existentes entre indivíduos que cursam o ensino médio em escolas da rede pública e da rede particular, reservando um determinado número de vagas fixas para indivíduos oriundos de escolas públicas [1].

A Universidade Federal de Goiás (UFG) tem como propósito institucional a promoção de uma formação acadêmica de qualidade para a sociedade, formando profissionais que possam contribuir positivamente no crescimento econômico e na melhora das condições de vida no estado e no país. No entanto, a UFG historicamente reflete as desigualdades sociais existentes na população. Por exemplo, com base em dados institucionais observou-se que no ano de 1998, 51% dos ingressos na instituição correspondiam a indivíduos procedentes de escolas públicas. Após de 8 anos, no ano de 2006, os novos estudantes oriundos de escolas públicas representavam em torno de 42% do total de estudantes [1].

Perante esta realidade, a UFG entende que deve promover ações afirmativas para incluir aqueles indivíduos excluídos, seja por questões socioeconômicas ou culturais, de um ensino superior de qualidade, contribuindo positivamente para o exercício de sua plena cidadania. É assim que surge no ano de 2008, o programa de ações afirmativas UFGInclui. Esse programa facilita o ingresso de estudantes procedentes de escolas públicas, autodeclarados como sendo de raça negra, procedentes das comunidades dos remanescentes dos quilombos do estado e aqueles que apresentam uma deficiência auditiva (LIBRAS). Além disso, o programa UFGInclui procura facilitar a permanência desses estudantes na instituição, implementando uma série de ações tais como: o auxílio alimentação, auxílio moradia, bolsas de estágio, bolsas de extensão, entre outros. Sabemos que a media global é um indicador para o desempenho do estudante no ensino superior. Então, faz-se interessante avaliar qual desempenho daqueles estudantes que entraram por ação afirmativa, e aqueles que ingressaram pelo sistema universal.

2 Objetivos

É interessante avaliar o desempenho acadêmico entre os estudantes ingressos pelo programa de ação afirmativa UFGInclui e pelo sistema universal. A hipótese inicial que queremos testar, é que, os alunos ingressos pelo programa UFGInclui possui um desempenho inferior aos ingressos pelo sistema universal.

3 Metodologia

A avaliação do impacto que programa UFGInclui origina no desempenho acadêmico dos estudantes, aproximado pelas médias globais, se faz necessário aderir aos métodos que permitam inferir a existência de uma relação de causalidade entre a política e o desempenho Acadêmico [4].

Para isso, é fundamental a identificação das variáveis que mascaram a causalidade. São variáveis que possuem impacto no desempenho estudantil. Uma vez identificados as variáveis, necessita-se um método para o isolamento destas no impacto que o programa causa no desempenho acadêmico [4].

A técnica de nome Propensity Scoring Matching (PSM) surge como modo de tornar isto possível. Dentre o grupo controle (aqueles que não participam do programa) e o grupo tratamento (estudantes ingressos pelo UFGInclui) o PSM seleciona indivíduos de ambos os grupos de tal forma que eles sejam equivalentes em características previamente selecionadas.

O PSM constrói uma comparação estatística que é baseada na probabilidade de se pertencer ao grupo de tratamento dada certas características que não são afetadas pelo programa. Os indivíduos pertencentes ao grupo tratamento são anexados a “base de probabilidade”, ou propensity score, contra os indivíduos do grupo controle.

Assim é selecionado uma amostra de cada grupo, com mesmas características sociais. Agora através do método de análise de variância não paramétrico Kruskal-Wallis é possível observar a existência do impacto do UFGInclui sobre o desempenho acadêmico.

4 Resultados

Após a aplicação da metodologia, tendo o cuidado para isolar as variáveis que mascaram o efeito do programa de ação afirmativa, foi constatado, através dos dados, este por sua vez não afeta o desempenho estudantil para os estudantes da UFG no ano de 2010. De acordo com a metodologia empregada, o resultado significa que para cada estudante que entrou pelo programa UFGInclui, existe um outro com mesmas condições sociais que tem condições cognitivas de ingressar a UFG pelo sistema universal, tendo o mesmo desempenho que o primeiro.

Referências

- [1] B.F. Marques, Resolução CONSUNI N° 29/2008, **UFG INCLUI Programa de Inclusão da UFG**, Goiânia (2008).
- [2] M.A. Vieira e R.M. Tenório, **Impacto Da Escolaridade Dos Pais E Nível Socioeconômico Familiar Nos Resultados De Testes Cognitivos**.
- [3] R. Barros, R. Mendonça, D. Santos e G. Quintaes, **Determinantes do desempenho educacional no Brasil**, IPEA Outubro de 2001. Disponível em: <http://repositorio.ipea.gov.br/bitstream/11058/2160/1/TD_834.pdf>. Acesso em: 2015.
- [4] S.L. Morgan e C. Winship, **“Methods and principles for social research”, Counterfactuals and casual inference**, primeira publicação: 2007. Cambridge University.
- [5] S.R. Khandker, G.B. Koolwal e H.A. Samad, **“Quantitative methods and practices”, Handbook on Impact Evaluation**, 2010 The International Bank for Reconstruction and Development / The World Bank.

Distribuição Birnbaum-Saunders baseada no núcleo logístico e alguns problemas de inferência

Navarro Mendes Santos Rosa - *navarro_msr@hotmail.com* - Universidade Federal de Goiás

Helton Saulo - *heltonsaulo@gmail.com* - Universidade Federal de Goiás

1 Introdução

Em geral, na literatura, o termo “tempo de vida” muitas vezes é relacionado à variáveis aleatórias positivas que representam o tempo passado até que determinado evento ocorra. Porém, este termo abrange um conceito muito mais geral, como por exemplo, a quantidade de quilômetros rodados até que um carro apresente falha, a resistência de uma peça sob determinada pressão. Estas variáveis não são medidas cronológicas, então o tempo de vida é mensurado de acordo com outros termos. Distribuições de probabilidade normalmente usadas para modelagem do “tempo de vida” frequentemente são chamadas de distribuições de vida.

Sem dúvidas a distribuição mais utilizada em análises estatísticas é a normal (curva Gaussiana), ela possui muitas propriedades úteis e conceitos conhecidos. Porém, quando o assunto é análise de “tempo de vida” esta distribuição não se adequa. Geralmente, os dados são assimétricos à direita, positivos ($T > 0$), onde T é uma variável tempo de vida. Nesta área a distribuição chave é o modelo exponencial.

Frente a novos desafios, o ser humano passa a desenvolver técnicas específicas para lidar com tais obstáculos, e isto não foi diferente quanto a modelagem de tempo de vida. Em Davis (1952), veja [2], mostrou inúmeros dados de falha, junto com testes de qualidade de ajuste para o modelo exponencial, se tornando então uma base para o engajamento deste modelo como uma distribuição de tempo de vida, usado como justificativa da escolha do modelo exponencial ainda hoje. Posteriormente outros trabalhos vieram para promover a escolha deste modelo para dados de tempo de vida.

Através desses trabalhos o modelo exponencial se tornou chave nos estudos de tempo de vida. O modelo probabilístico exponencial se baseia numa população finita N , e quando este tende ao infinito propriedades interessantes surgem, como por exemplo a perda de memória, que facilita simplificações de inferência associado ao modelo, porém esta propriedade também significa uma limitação por ser uma suposição forte a se aplicar a dados reais, como por exemplo itens onde ocorre processo de envelhecimento. Além disto, o modelo apresenta taxa de falha constante, e cálculo facilitado, grande atrativo para época, levando em consideração os recursos computacionais limitados.

O modelo gamma surge de forma natural, modelando uma soma de tempos de vida com distribuições exponenciais. Gerada a partir de convolução de distribuições exponenciais, tem

como vantagem ser um caso mais geral destas, porém não é algebricamente tratável, elevando assim os custos computacionais.

A partir deste ponto, cada vez mais se procurou dados e formas de modelagem de tempo de vida que fogem do clássico modelo exponencial. Por exemplo, no trabalho de Lieblein e Zelen (1956), veja [4], e Kao (1959), ver [3], propuseram o uso da distribuição Weibull, pois esta se adequava bem a dados de falha. Por sua vez esta distribuição foi pensada por Weibull (1951), veja [6], para descrever a força necessária para ruptura de materiais. Esta distribuição é muito utilizada, pois sua função de taxa de falha é flexível, além de ser matematicamente tratável.

Motivados pela fadiga de materiais relacionados a aviões, provocada pela constante vibração desses, Birnbaum e Saunder (1968), veja [1], propuseram um novo modelo probabilístico para análise do tempo de vida de materiais expostos à fadiga, stress ou tensão. No ano seguinte, os autores formalizaram a distribuição de vida-fadiga e apresentaram métodos de estimação para distribuição de probabilidade bi-paramétrica.

No estudo de variáveis de tempo de vida, em geral, há algumas classes distintas e bem definidas, como por exemplo, dano cumulativo, degradação cumulativa, em fases, valores extremos, entre outros. Normalmente é pela definição destas classes em que a escolha do modelo probabilístico apropriado é feita. Em particular a classe de dano cumulativo se faz interessante pela quantidade de áreas onde ela pode ser empregada, tais como engenharia, dados ambientais, ciência médica (como doenças cardíacas e câncer), e outros. Logo o estudo e aplicação da distribuições Birnbaum-Saunders (BS) se torna relevante.

Neste trabalho será feito a apresentação da distribuição BS baseada no núcleo logístico e os estimadores de máxima verossimilhança para α e β . Após isto, é feito um estudo de Monte Carlo para avaliação do desempenho destes estimadores.

2 BS-Logística

Um caso particular da distribuição GBS (*Generalized Birnbaum Saunders*) é a BS-Logística. Nela a distribuição associada z segue uma distribuição logística de parâmetros $\mu = 0$ e $s = 1$, sendo $Z \sim L(0, 1)$. A variável Z tem uma forma semelhante à distribuição normal, porém possui caudas mais pesadas. Então, Z pode ser escrita da seguinte forma,

$$f(z; \mu = 0, s = 1) = \frac{e^z}{\left(1 + e^z\right)^2}. \quad (1)$$

Conhecida a forma de Z pode ser encontrado a função de densidade e probabilidade da BS-Logística através da equação (2). Seja $T \sim BS-L(\alpha, \beta)$, então a função de densidade e

probabilidade de T pode ser escrita como,

$$f_T(t) = \frac{\exp\left[\frac{1}{\alpha}\left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right)\right]}{\left\{1 + \exp\left[\frac{1}{\alpha}\left(\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right)\right]\right\}^2} \left[\sqrt{\frac{\beta}{t}} + \sqrt{\frac{\beta^3}{t^3}}\right]; t > 0. \quad (2)$$

A BS-Logística é um caso particular da GBS. Logo, todos os resultados válidos para a GBS também valem para a BS-Logística, incluindo as propriedades, estimadores, propriedade de estimadores e métodos para a verificação de ajuste. A seguir será estudado a forma da BS-Logística baseado no pacote gbs no software R.

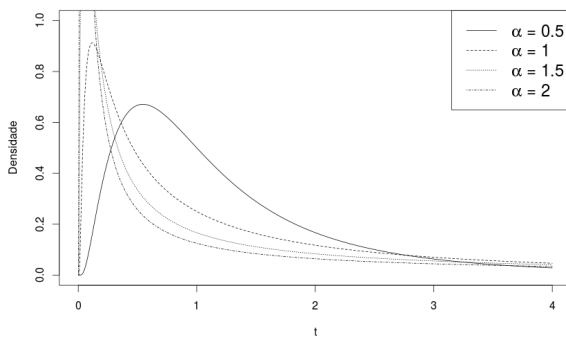


Figura 1: Variação do parâmetro α para a BS-Logística

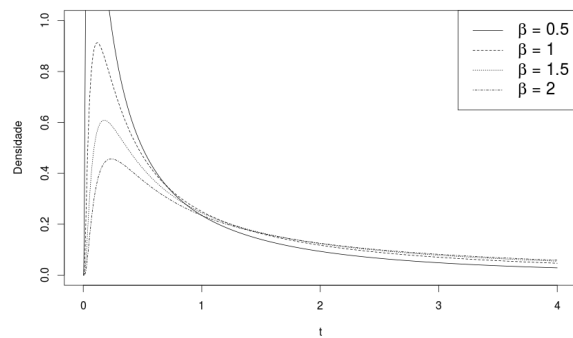


Figura 2: Variação do parâmetro β para a BS-Logística

Pode-se observar na FIG.1 que o parâmetro β possui um impacto muito grande na assimetria da distribuição. De fato ele é o parâmetro referente à escala, quanto menor este é, mais concentrados os dados estarão em torno da moda. Já na FIG.2 temos as variações do parâmetro de forma, onde este afeta a assimetria da distribuição.

Seja a variável aleatória U , onde $U = Z^2$ e $u_i = \mathbb{E}(U^i)$. Seja uma variável aleatória $H \sim GBS(\alpha, \beta; g)$, onde o núcleo g é uma variável aleatória Z . Então os primeiros quatro momentos de H , veja [5], são dados como,

$$\begin{aligned} \mathbb{E}(H) &= \frac{\beta}{2}(2 + u_1\alpha^2); \\ \mathbb{E}(H^2) &= \frac{\beta^2}{2}(2 + 4u_1\alpha^2 + u_2\alpha^4) \\ \mathbb{E}(H^3) &= \frac{\beta^3}{2}(2 + 9u_1\alpha^2 + 6u_2\alpha^4 + u_3\alpha^6); \\ \mathbb{E}(H^4) &= \frac{\beta^4}{2}(2 + 16u_1\alpha^2 + 20u_2\alpha^4 + 8u_3\alpha^6 + u_4\alpha^8). \end{aligned} \quad (3)$$

Para o caso particular onde o núcleo g segue uma distribuição logística $Z \sim L(0, 1)$. Sabemos que $\mathbb{E}(Z) = 0$, logo $Var(Z) = \mathbb{E}(Z^2) = \mathbb{E}(U) = u_1 = \frac{\pi^2}{3}$. A distribuição logística padrão ($L(0, 1)$) é simétrica, logo $\mathbb{E}(Z^3) = 0$ e $\mathbb{E}(Z^4) = \mathbb{E}(U^2) = u_2 = \frac{7\pi^4}{15}$. Assim, através das equações em (3),

pode-se obter as média e variância de $T \sim BS-L(\alpha, \beta)$, como

$$\mathbb{E}(T) = \frac{\beta}{2} \left(2 + \frac{\pi^2 \alpha^2}{3} \right) \quad (4)$$

$$\text{Var}(T) = \frac{\beta^2 \alpha^2}{4} \left[\frac{4\pi^2}{3} + \left(\frac{14\pi^4}{15} - \frac{\pi^4}{9} \right) \alpha^2 \right] \quad (5)$$

Em geral, a BS-Logística é uma distribuição assimétrica à direita. Como esta assimetria se torna maleável pelos parâmetros, a distribuição é uma boa candidata para lidar com dados de tempo de vida assimétricos e referentes ao dano cumulativo. Pois ela é um caso particular da GBS.

3 Estudo de Monte Carlo

Utilizando o núcleo logístico para a distribuição GBS, têm-se uma distribuição com novas aplicações. É imprescindível então, a análise de desempenho dos estimadores de máxima verossimilhança para os parâmetros α e β que neste trabalho será realizada usando o método de Monte Carlo. Este método consiste em, replicar o estimador para diferentes amostras, para então ser computado as estatísticas de interesse para o experimento.

4 Estimadores de Máxima verossimilhança

Para os estimadores de máxima verossimilhança, veja [5], foram usadas amostras de tamanho 20, 30, 50 e 100. Para cada tamanho de amostra simulou-se dez mil repetições de Monte Carlo, ou seja, $R = 10.000$ estimações para avaliação. Com as R repetições foi calculado o viés, erro quadrático médio, desvio padrão e média dos estimadores.

	$N = 20$	$N = 30$	$N = 50$	$N = 100$
MEDIA- α	0.48	0.49	0.49	0.50
MEDIA- β	1.02	1.01	1.01	1.00
SD- α	0.09	0.08	0.06	0.04
SD- β	0.19	0.15	0.12	0.08
MSE- α	0.01	0.01	0.00	0.00
MSE- β	0.04	0.02	0.01	0.01
VIES- α	-0.02	-0.01	-0.01	-0.00
VIES- β	0.02	0.01	0.01	0.00

Tabela 1: Estudo de Monte Carlo para estimadores de máxima verossimilhança para $\alpha = 0.5$ e $\beta = 1$.

Pode-se observar em TAB.1 que, a medida que N cresce os estimadores se aproximam do valor real da amostra. Além do fato que, medidas de dispersão e erro diminuem de acordo com o crescimento do tamanho amostra aleatória da distribuição $BS - L(\alpha = 0.5, \beta = 1)$.

5 Conclusão

Através dos resultados apresentados, pode-se observar a existência de estimadores funcionais para os parâmetros da distribuição BS-Logística, possibilitando a modelagem de dados reais com esta última. A existência de novas distribuições para modelagem de dados favorece a comunidade científica com novas ferramentas que se adequam melhor a determinadas tarefas.

Referências

- [1] BIRNBAUM, Z., AND SAUNDERS, S. A probabilistic interpretation of miner's rule. **SIAM Journal of Applied Mathematics** **16** (1968), 637–652.
- [2] DAVIS, D. An analysis of some failure data. **Journal of the American Statistical Association** **47** (1952), 113–150.
- [3] KAO, J. A graphical estimation of mixed Weibull parameters in life-testing of electron tubes. **Technometrics** **1** (1959), 389–407.
- [4] LIEBLEIN, J., AND ZELEN, M. Statistical investigation of the fatigue life of deep ball bearing. **Journal of research of the National Bureau of Standards** **57** (1956), 273–316.
- [5] SANHUEZA, A., LEIVA, V., AND BALAKRISHNAN, N. The generalized Birnbaum-Saunders distribution and its theory, methodology and application. **Communications in Statistics: Theory and Methods** **37** (2008), 645–670.
- [6] WEIBULL, W. A statistical distribution function of wide applicability. **Journal of Applied Mechanics** **1** (1951), 293–297.

AJUSTE SAZONAL DA PRODUÇÃO INDUSTRIAL EM GOIÁS: UMA ANÁLISE DA INDÚSTRIA DE TRANSFORMAÇÃO E SUAS DESAGREGAÇÕES

Alex Felipe Rodrigues Lima - *alexfrl@segplan.go.gov.br* - Instituto Mauro Borges - SEGPLAN

Sergio Borges Fonseca Júnior - *seergio-bj@segplan.go.gov.br* - Instituto Mauro Borges -
SEGPLAN

Dinamar Maria Ferreira Marques - *dinamar-mfm@segplan.go.gov.br* - Instituto Mauro
Borges - SEGPLAN

Mario Ernesto Piscocya Díaz - *mario.piscocya@gmail.com* - IME - UFG

O objetivo deste artigo é realizar o ajuste sazonal e obter o melhor modelo de ajuste para a indústria de transformação e em suas desagregações - indústria alimentícia, de minerais não-metálicos e de metalurgia - no estado de Goiás. O ajuste sazonal é importante, pois permite a comparação entre meses subsequentes, o que é imprescindível, para o melhor acompanhamento da evolução dos indicadores da indústria, sobretudo, em anos que ocorrem profundas mudanças na atividade econômica. Os resultados foram obtidos através do software R em conjunto com o X-13-ARIMA-SEATS. Verificou-se que as variáveis relacionadas ao calendário (carnaval, anos com maior número de dias trabalhados) e outliers exercem influencia significativa para o melhor ajuste do modelo.

Palavras-chave: *Ajuste sazonal; Indústria de Transformação; Efeito calendário; X-13-ARIMA-SEATS; Estado de Goiás.*

APLICAÇÃO DA TEORIA DAS FILAS NO RESTAURANTE CIDADÃO

Ana Carolina do Couto Andrade e-mail: *carol.goiana@hotmail.com* Instituto de Matemática
e Estatística - IME

Fabiano Fortunato Teixeira dos Santos e-mail: *fortunato@ufg.br* Instituto de Matemática e
Estatística - IME

Aguardar em filas para ser atendido torna-se cada vez mais desgastante e incômodo para a população. Diante disso, este estudo tem por finalidade equacionar os indicadores do Sistema de Filas existente no Restaurante Cidadão de Goiânia, para que a posteriori seja possível dimensionar e planejar as instalações do local. Será feita uma coleta de dados que possibilitará, através de testes de aderência, a validação do modelo proposto, $M/M/c$.

Palavras-chave: *Filas, Planejamento, Restaurante.*

Teoria das Filas Aplicada ao Transplante de Córneas em Goiás

Lucas Siqueira Cheim - *lucas_siqueira11@hotmail.com* - IME/UFG
Fabiano Fortunato Teixeira dos Santos - *fortunato@ufg.br* - IME/UFG

O objetivo deste estudo é modelar de forma eficiente a fila de espera para o transplante de córneas em Goiás, bem como analisar as variáveis que interferem no tempo de espera dos pacientes, utilizando prioritariamente a Teoria das Filas. Para tal, foi realizado o estudo dos conceitos básicos para o entendimento dos sistemas de filas. Contando com a colaboração da Central de Notificação, Captação e Distribuição de Órgãos do Estado de Goiás (CNCDO/GO), que se propôs a oferecer o banco de dados para estudo, espera-se encontrar valores para importantes variáveis a fim de definir o modelo mais apropriado ao caso. Em seguida, deseja-se realizar simulações para observar como as variáveis responderiam a melhorias ou sobrecargas no sistema. Espera-se encontrar o modelo de fila conhecido como M/M/s, com o intervalo de chegada e o tempo de atendimento distribuído exponencialmente, independente e identicamente distribuído (i.i.d.). Por fim, considerando a carência de produções acadêmicas para filas de transplante, este estudo vem representar uma pequena parte do cenário de transplantes de órgãos em Goiás, a fim de fomentar a pesquisa na área e trazer mais visibilidade à causa.

Palavras-chave: *Teoria das Filas, Transplante de Córnea, Modelo, CNCDO/GO.*

Inferência e Correção de Vício em Filas Markovianas de Servidor Único

Luiz Carlos R. Pereira - *luizreal05@ufmg.br* - Universidade Federal de Minas Gerais

Márcio A. C. Almeida - *estatimarcio@gmail.com* - Universidade Federal do Pará

Frederico R. B. Cruz - *fcruz@est.ufmg.br* - Universidade Federal de Minas Gerais

Fernando L. P. Oliveira - *fernandoluiz@iceb.ufop.br* - Universidade Federal de Ouro Preto

As filas de espera estão presentes em diversas situações do nosso dia a dia, tais como na lanchonete, no supermercado, no banco e no posto de gasolina. Várias são as medidas que descrevem o desempenho das filas. Neste trabalho daremos foco à estimativa da intensidade do tráfego ρ (também denominado fator de utilização da estação de serviço), para o caso de filas $M/M/1$, que, na notação de Kendall, representa filas com tempos entre chegadas e de serviço exponenciais (markovianos), servidor único e capacidade de espera infinita. Experimentos computacionais apresentados mostram que o estimador de máxima verossimilhança para ρ apresenta vício para amostras pequenas e moderadas ($n < 200$). Adicionalmente, uma versão do estimador, corrigida pelo método *bootstrap* não paramétrico, foi testada. Assim, com um esforço computacional extra, o pesquisador pode alcançar, com amostras de tamanho $n = 5$, estimativas para ρ com erro médio equivalente a estimativas com amostras de tamanho $n = 20$. Esta redução pode ser bastante importante do ponto de vista prático, pela diminuição que pode acarretar nos custos e no tempo para obtenção das amostras.

Palavras-chave: *Bootstrap, desempenho, filas markovianas, inferência, vício.*

Desempenho dos times que disputaram o campeonato brasileiro de futebol (Série A) de 2003 a 2014

Nathalia Rodrigues Damasceno e-mail: *nathaliadamascenoufg2@gmail.com* IME-UFG

Valdivino Vargas Júnior e-mail: *vvjunior@gmail.com* IME-UFG

O campeonato brasileiro de futebol é uma disputa nacional que consegue reunir milhares de brasileiros para seus jogos. O desempenho dos clubes é um assunto muito polêmico, pois cada um tem um carinho por algum dos clubes, o que acaba atrapalhando essa discussão. O trabalho fará análises sobre o desempenho dos clubes que disputaram pelo menos uma vez o campeonato brasileiro de 2003 a 2014, somente usando os dados oficiais de cada ano. Para obter esses resultados serão utilizadas oito variáveis referentes aos dados oficiais acumuladas ano após ano. Para definir os clubes com melhores ou piores desempenho foi utilizada a técnica de análise de agrupamento (cluster). A análise de agrupamento consiste em agrupar objetos segundo suas características, com isso criam-se grupos homogêneos e significativamente diferentes dos demais grupos.

Palavras-chave: *Campeonato Brasileiro, Futebol, Desempenho, Análise de Agrupamento.*

ANÁLISE DE SOBREVIVÊNCIA: Uma aplicação para a evasão universitária na UFG, 2009

Nathalia Rodrigues Damasceno e-mail: *nathaliadamascenoufg2@gmail.com* IME-UFG

Mario Ernesto Piscoya Diaz e-mail: *mario.piscoya@gmail.com* IME-UFG

Helton Saulo Bezerra dos Santos e-mail: *heltonsauro@gmail.com* IME-UFG

A evasão universitária é um problema importante dentro das universidades públicas brasileiras que está relacionado com uma série de impactos negativos na formação de recursos humanos e, também individuais. A literatura internacional aponta uma série de fatores socioeconômicos e demográficos que encontram-se associados com a evasão. O objetivo deste trabalho é identificar a existência de uma relação entre tais fatores e a evasão para o caso da Universidade Federal de Goiás, usando técnicas de análise de sobrevivência tais como o estimador Kaplan-Meier e testes de hipótese da família Tarone-Ware. Os resultados mostram que todas as características socioeconômicas estudadas estão associados com a evasão nesta instituição.

Palavras-chave: *Evasão, Fatores Socioeconômicos, Educação Superior.*

Casualidade do programa UFGInclui sobre o desempenho acadêmico usando propensity scoring matching

Navarro Mendes Santos Rosa - *navarro_msr@hotmail.com* - Universidade Federal de Goiás

Mario E. Piscoya Díaz - *mario.piscoya@gmail.com* - Universidade Federal de Goiás

O trabalho consiste em averiguar o impacto que o programa de ação afirmativa UFGInclui tem sobre o desempenho acadêmico, através da média global. Para isto se faz necessário isolar fatores externos que influenciam no desempenho acadêmico, para isto foi utilizado a técnica propensity scoring mathing. E para validar o impacto do programa foi utilizado o teste de Kruskal-Wallis.

Palavras-chave: *propensity scoring matching, UFGInclui, casualidade, desempenho acadêmico.*

Distribuição Birnbaum-Saunders baseada no núcleo logístico e alguns problemas de inferência

Navarro Mendes Santos Rosa - *navarro_msr@hotmail.com* - Universidade Federal de Goiás

Helton Saulo - *heltonsauro@gmail.com* - Universidade Federal de Goiás

Os modelos de tempo de vida são importantes para a realização de análises de sobrevivências. Cada dia surge novos problemas e tipos de dados que as distribuições de probabilidade clássicas não conseguem abranger. Dito isto, é imperioso o desenvolvimento de novos modelos e técnicas para possibilitar a análise destes novos dados. A distribuição BS surge como proposta a problemas de dano cumulativo. O trabalho tem como objetivo estudar a distribuição BS-Logística como proposta de modelagem de dados.

Palavras-chave: *inferência, Birnbaum-Saunders, Logístico.*

ESTATÍSTICA E BIG DATA NA INDÚSTRIA E AGRONEGÓCIO

Doris S M Fontes - e-mail: *dsfontes@gmail.com* - CONRE-3

O mercado de trabalho do estatístico continua muito aquecido. Além das áreas que costumam contratar bastante, o crescimento na utilização do Big Data está favorecendo o surgimento de mais oportunidades de trabalho, inclusive em segmentos como Indústria e Agronegócio. Como o estatístico pode aproveitar esse momento para garantir o seu espaço?

Palavras-chave: *Estatístico, agronegócio, big-data.*

Introdução ao mapeamento de QTL's.

Renato Rodrigues Silva - e-mail: *renato.rrsilva@ufg.br* - IME-UFG

Por muitas vezes, em genética vegetal, animal e ou humana, o pesquisador tem interesse em conhecer a base genética da variação dos caracteres quantitativos. Com o advento dos marcadores moleculares, é possível realizar estudos sobre associações entre determinadas regiões no genoma e a variação de um determinado caráter observado. Essas regiões são denominadas como QTL's ou "quantitative trait loci". Assim, o mapeamento de QTL's consiste em estimar o número, a posição e os efeitos dos QTL's. Nessa palestra, serão apresentados os principais métodos estatísticos utilizados para o mapeamento dos QTL's.

Palavras-chave: *Caracteres quantitativos, quantitative trait loci, estatística genética.*

Likelihood-based Inference for Mixed-Effects Models with Censored Response Using the Multivariate-t Distribution

Victor Hugo Lachos Davila - e-mail: *hlachos@ime.unicamp.br* - Instituto de Matemática e Estatística e Computação Científica, IMECC-UNICAMP.

Mixed-effects models are commonly used to fit longitudinal or repeated measures data. A complication arises when the response is censored, for example, due to limits of quantification of the assay used. Although the normal distributions are commonly assumed for the random effects and the residual errors, such assumptions make inferences vulnerable in the presence of outliers. The sensitivity to outliers and the need of heavy tailed distributions for random effects and residual errors motivate us to develop a likelihood-based inference for linear and nonlinear mixed effects models with censored response (NLMEC/LMEC) based on the multivariate Student-t distribution. An ECM algorithm is developed for computing the maximum likelihood estimates for NLMEC/LMEC with the standard errors of the fixed effects and the likelihood function as a by-product. The algorithm uses the closed-form expressions at the E-step, which relies on the formulas of the mean and variance of a truncated multivariate-t distribution. The proposed algorithm is implemented in the R package *tlmec*. The proposed methodology is applied to analyze longitudinal HIV viral load data in the two recent AIDS studies. In addition, a simulation study is conducted to examine the performance of the proposed method and carry out the comparison with the approach by Vaida and Liu (2009).

Palavras-chave: *Mixed-Effects Models, Censored Response, Multivariate-t Distribution.*