Universidade Federal do Rio Grande do Sul
Instituto de Matemática e Estatística
Departamento de Estatística

Anais

VII SEMANÍSTICA

VII Semana Acadêmica do Departamento de Estatística

da UFRGS

http://www.ufrgs.br/semanistica

Porto Alegre - 12 a 14 de setembro de 2016

Organização:

Departamento de ESTATÍSTICA
IME-UFRGS

Promoção:

Instituto de MATEMÁTICA E ESTATÍSTICA
UFRGS

UFRGS
UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

Conteúdo

1 Introdução **4**

2 Agradecimentos **4**

3 Comissão Organizadora Docente **5**

4 Comissão Científica **5**

5 Comissão Organizadora Discente **5**

6 Apresentação **5**

7 Programação **6**

8 Conferências **7**

9 Seções de Comunicações **10**

1 Introdução

A VII Semana Acadêmica da Estatística (VII SEMANÍSTICA) foi realizada de 12 a 14 de setembro de 2016, no Instituto de Matemática e Estatística - IME, Campus do Vale da UFRGS, Porto Alegre, RS. O evento engloba os mais variados temas dentro da área acadêmica e profissional.

O objetivo principal da SEMANÍSTICA é promover o desenvolvimento, aprimoramento e a divulgação da Estatística, entre diferentes perspectivas, acadêmica e/ou prática no campo de aplicação. A proposta da SEMANÍSTICA é promover a integração entre estudantes, professores e profissionais de diversas áreas que utilizam a Estatística como suporte de decisão em suas respectivas áreas de conhecimento. Propõe-se que o evento seja um cenário de aproximação e troca de experiências entre professores e alunos em diferentes áreas de conhecimento.

Como objetivos específicos da SEMANÍSTICA, podem-se citar: divulgar as contribuições recentes dos pesquisadores participantes promovendo-se o intercâmbio entre cientistas, alunos e profissionais aplicados; promover um maior contato entre pesquisadores do Departamento de Estatística da UFRGS e pesquisadores de outros departamentos, propiciando futuros trabalhos de pesquisa conjuntos; intensificar o contato e o intercâmbio científico entre profissionais da Região Sul e a iniciativa privada dentro das realidades do Estado do Rio Grande do Sul e do MERCOSUL; divulgar os diferentes métodos e aplicações de Estatística para discentes da graduação em Estatística, bem como discentes de pós-graduação e graduação das mais diversas áreas correlatas, tais como: Economia, Administração, Engenharia e Biomédicas.

Para maiores informações sobre a VII SEMANÍSTICA (Semana Acadêmica da Estatística 2016) ver www.ufrgs.br/semanistica.

2 Agradecimentos

A VII SEMANÍSTCA - Semana Acadêmica do Departamento de Estatística da UFRGS não teria sido possível sem o apoio das seguintes agências financiadoras e instituições:

- DEST-UFRGS - Departamento de Estatística da UFRGS
- IME-UFRGS - Instituto de Matemática e Estatística da UFRGS
- UFRGS - Universidade Federal do Rio Grande do Sul

A Comissão Organizadora da VII SEMANÍSTICA agradece a colaboração de todos que se dedicaram anonimamente e sem interesses pessoais, em promover a integração entre alunos, professores e profissionais em estatística.

Comissão Organizadora

3 Comissão Organizadora Docente

- Cleber Bisognin (Coordenador - Departamento de Estatística-UFRGS)
- Danilo Marcondes Filho (Departamento de Estatística-UFRGS)
- Guilherme Pumi (Departamento de Estatística-UFRGS)

4 Comissão Científica

- Guilherme Pumi (Coordenador - Departamento de Estatística-UFRGS)
- Cleber Bisognin (Departamento de Estatística-UFRGS)
- Danilo Marcondes Filho (Departamento de Estatística-UFRGS)

5 Comissão Organizadora Discente

- Gabriel da Cunha (Curso de Estatística - UFRGS)
- Ivanhoé Angelo Fachél Filho (Curso de Estatística - UFRGS)
- Luana Giongo Pedrotti (Curso de Estatística - UFRGS)
- Matias Segelis Vieira (Curso de Estatística - UFRGS)

6 Apresentação

O programa da VII SEMANÍSTICA - Semana Acadêmica do Departamento de Estatística da Universidade Federal do Rio Grande do Sul englobou as seguintes atividades:

- 8 Conferências envolvendo pesquisas realizadas em diversas áreas da Estatística proferidas por pesquisadores convidados de Universidades do Rio Grande do Sul e do Brasil e ainda profissionais em Estatística;
- Comunicações orais apresentadas pelos participantes do evento;

7 Programação

Horário	12/9/2016	13/9/2016	14/9/2016
	Segunda-Feira	Terça-Feira	Quarta-Feira
08:30-09:15	C1	C5	Apresentações Alunos
09:15-10:00	C2		
10:00-10:30	Coffe Break	Coffe Break	Coffe Break
10:30-11:15	C3	C6	C7
11:15-12:00	C4		C8

Conferências:

(C1) Conferência 1 - Prof. Dr^a. Luciana Neves Nunes - DEST - UFRGS
Título: A Ética em Pesquisa

(C2) Conferência 2 - Amanda de Mello Martins e Giovana da Silva Lenzi
Título: Diversidade na Universidade

(C3) Conferência 3 - Pedro Gerhardt Gavronski
Título: Risco sistêmico das instituições financeiras dos EUA

(C4) Conferência 4 - Angela Silva e Filippo Petroli
Título: Desafio Francis Galton

(C5) Conferência 5 - Pietro Tiaraju Giavarina dos Santos
Título: Introdução à interfaces gráficas utilizando RStudio e Shiny

(C6) Conferência 6 - João Henrique Ferreira Flores
Título: Introdução a Modelos de Redes Neurais

(C7) Conferência 7 - Álvaro Vigo
Título: Programa de Pós-Graduação em Epidemiologia/UFRGS: Linhas de Pesquisa e Possibilidades de Formação para Estatísticos

(C8) Conferência 8 - Fernando Neves Hugo e Denise Antunes
Título: Desafios analíticos na Análise da Provisão de Cuidados Odontológicos na Atenção Primária á Saúde na Perspectiva da Equidade

8 Conferências

Conferência 1

A Ética em Pesquisa

Prof^a. Dr^a. Luciana Neves Nunes
Departamento de Estatística - UFRGS

Resumo

Por definição, ética é a investigação geral sobre aquilo que é bom. Numa perspectiva histórico-dialética, é querer um certo bem geral, uma vez que existam as condições materiais e técnicas indispensáveis para a concretização desse bem. Para orientar os pesquisadores das diversas áreas de pesquisa, algumas resoluções servem como norteadoras: Resolução 466/12 do CNS, Resolução, Lei 11.794/08 que diz respeito a pesquisa com animais, a Minuta de resolução sobre 'especificidades éticas das pesquisas nas ciências humanas e sociais e de outras que se utilizam de metodologias próprias dessas áreas" de 2016 e outros documentos. Para as Pesquisas de Mercado e Pesquisas Sociais, existe o Código de Ética a ser seguido. Toda essa documentação e regulamentação da ética na pesquisa traz informações que atingem diretamente a prática dos profissionais estatísticos que se envolvem em pesquisa.

Conferência 2

Ações Afirmativas na UFRGS

Amanda de Mello Martins e Giovana da Silva Lenzi
Núcleo de Graduação do Instituto de Matemática e Estatística - UFRGS

Resumo

A oficina será apresentada pelas Técnicas em Assuntos Educacionais Amanda Martins e Giovana Lenzi e irá propor reflexão a respeitos de políticas públicas inclusivas em nosso país, especialmente aquelas voltadas para a inclusão de estudantes negros e indígenas nas universidades amparados pelas reserva de vagas. Nesse sentido abordarão os conceitos de raça, racismo institucional, identidade, diversidade cultural, entre outros, bem como as ações de permanência promovidas pela UFRGS para alunos cotistas.

Conferência 3

Risco sistêmico das instituições financeiras dos EUA

Pedro Gerhardt Gavronski
Programa de Pós-Graduação em Economia (PPGE) - UFRGS

Resumo

Os objetivos deste trabalho são propor uma nova forma de mensurar o risco sistêmico do sistema financeiro de uma economia e identificar as instituições financeiras que mais influenciam o risco de uma economia. Essa medida será construída a partir da teoria dos valores extremos, e utilizaremos um modelo do tipo GAS para descrever a dinâmica desta medida ao longo do tempo.

Conferência 4

Desafio Francis Galton

Angela Silva e Filippo Petroli
Curso de Bacharelado em Estatística - UFRGS

Resumo

O Desafio Francis Galton, através da Escola Convexo, tem como objetivo buscar melhorias nas estruturas física e operacional do Instituto de Matemática e Estatística/UFRGS. Através de encontros com estudantes de Estatística e Matemática, abordamos tópicos no intuito de implementar mudanças que auxiliarão no aprendizado e desenvolvimento dos estudantes do nosso instituto, como a criação de um espaço com computadores para os alunos complementarem o conteúdo visto em sala de aula. Contamos com o apoio de toda comunidade acadêmica para darmos seguimento ao projeto.

Conferência 5

Introdução à interfaces gráficas utilizando RStudio e Shiny

Pietro Tiaraju Giavarina dos Santos
Curso de Bacharelado em Estatística - UFRGS

Resumo

Devido o crescimento recente de recursos computacionais para a Estatística, a necessidade da proximidade e interação com a programação se faz cada vez mais relevante. O objetivo do minicurso é mostrar uma pouco sobre uma nova ferramenta chamada Shiny, que surgiu em 2014 e tem grande potencial para facilitar não só o estatístico individualmente como também da comunidade estatística como um todo.

Conferência 6

Introdução a Modelos de Redes Neurais

João Henrique Ferreira Flores
Departamento de Estatística - UFRGS

Resumo

Trata-se de uma apresentação de alguns modelos de Redes Neurais, inclusive do ponto de vista histórico. Além disso, pretendo apresentar um panorama geral e atual dos modelos e seus usos, principalmente de modelos multi-layer perceptron (MLP). Pauta-se a apresentação de forma a servir como uma motivação, assim como divulgar, a um possível curso de Modelos de Redes Neurais a ser iniciado em 2017/1.

Conferência 7

Programa de Pós-Graduação em Epidemiologia/UFRGS: Linhas de Pesquisa e Possibilidades de Formação para Estatísticos

Álvaro Vigo, Stela Maris de Jezus Castro e Patrícia Klarmann Ziegelmann
Departamento de Estatística - UFRGS

Resumo

Descrição da atuação dos professores do Departamento de Estatística credenciados no Programa de Pós-Graduação em Epidemiologia, Faculdade de Medicina, contemplando aspectos como disciplinas de Estatística, linhas de pesquisa e orientações. Serão brevemente descritos alguns trabalhos publicados da linha da pesquisa "Métodos Estatísticos Aplicados em Epidemiologia", bem como de trabalhos colaborativos com pesquisadores de outras linhas de pesquisa. Estas informações podem ser interessantes para se entender o perfil de formação e de atuação de um Estatístico com pós-graduação em Epidemiologia.

Conferência 8

Desafios analíticos na Análise da Provisão de Cuidados Odontológicos na Atenção Primária á Saúde na Perspectiva da Equidade

Fernando Neves Hugo e Denise Antunes
CPOS - Centro de Pesquisas em Odontologia Social - UFRGS

Resumo

A apresentação tem por objetivo descrever nosso problema de pesquisa, pois representa os desafios a serem enfrentados pelo estatístico na prática profissional, e discutir com os participantes o modelo de análise que pretendemos adotar. O nosso problema acontece em áreas de cobertura das chamadas Equipes de Saúde da Família. Em cada área vivem em torno de 4 mil pessoas. A legislação vigente no Brasil define que o cuidado em saúde deve ser fornecido pelas instituições de saúde pública a todos os cidadãos, conforme a necessidade de cada um. O nosso problema envolve a distribuição do cuidado odontológico segundo a necessidade de cada morador de áreas coberturas pela Saúde da Família, de tal forma a que a espera por tratamento dentário não leve à progressão das doenças bucais.

9 Seções de Comunicações

Comunicações Orais

Agrupamento de séries temporais financeiras aplicado a otimização de portfólios.

Cristiano Sulzbach [1 2]

Marcio Valk[3]

Hudson Torrent[4]

Resumo: A construção de carteiras de ativos financeiros que tenham risco mínimo para um determinado nível de retorno esperado é objetivo na teoria de portfólios. Neste trabalho propomos uma estratégia para construção de portfólios baseada em agrupamento de séries temporais. A definição dos grupos é feita através do método hierárquico, juntamente com teste para homogeneidade de grupos baseado em U-estatísticas. A alocação de capital nos grupos é definida por um critério original que distribui os pesos de acordo com o desempenho de cada grupo, segundo alguma medida de desempenho. Os resultados mostram que o método de agrupamento, juntamente com o critério para alocação de capital, é superior à estratégia 1/N e ao método de mínima variância para a amostra e período considerados.

Palavras-chave: *Séries Temporais, Otimização de Portfólio, Agrupamento*.

1 Introdução

Em um cenário ideal um investidor deseja um alto grau de retorno com um baixo nível de risco. A teoria moderna de portfólios estipula que investidores construam carteiras optimizando o retorno esperado para um determinado nível de risco de mercado. De acordo com essa teoria, proposta por [3], busca-se construir uma *fronteira eficiente*, em que os portfólios tenham retorno esperado máximo para um dado nível de risco. Portfólios que proporcionam retorno máximo para um dado nível de risco são chamados de *portfólios eficientes*.

Recentemente, alguns trabalhos propõem técnicas que combinam métodos de agrupamento com métodos clássicos de alocação de ativos ([6, 1, 4]). A justificativa para a utilização de clusters é que na técnica de Markowitz os retornos esperados e a matriz de variâncias e covariâncias não podem ser estimadas com precisão a partir dos dados. Para evitar o viés de estimação, uma solução é adotar uma estratégia simples que não dependa da estimação de parâmetros, a qual consiste em alocar igualmente

[1]UFRGS - Universidade Federal do Rio Grande do Sul. Email: *cristiano.sulzbach@hotmail.com*
[2]O autor agradece ao CNPQ pela bolsa de iniciação científica
[3]UFRGS - Universidade Federal do Rio Grande do Sul. Email: *marciovalk@gmail.com*
[4]UFRGS - Universidade Federal do Rio Grande do Sul. Email: *hudsontorrent@gmail.com*

o capital nos ativos. Essa estratégia chamada *"naive strategy"* ou 1/N, tem sido discutida na literatura e, embora não atenda aos pressupostos para alocação eficiente de portfólios, tem-se apresentado um *benchmark* difícil de ser superado.

Um método de classificação e agrupamento baseado em U-estatísticas é apresentado no trabalho de [5]. Esse método é baseado em um teste de homogeneidade para dois grupos de séries temporais e se mostra competitivo para encontrar a melhor divisão possível em dois grupos. No entanto, para o caso em que existem mais de dois grupos, esse método não foi estudado. Como todos os métodos de classificação e agrupamento, o método baseado em U-estatística depende da escolha da métrica. No contexto de volatilidade é mais adequado considerar medidas chamadas *Garch model based*.

Neste trabalho exploramos a construção de metodologias para otimização de portfólios de investimento baseada na utilização dessas técnicas de agrupamentos em séries temporais.

2 Medidas de Performance

Estamos interessados em medir a qualidade dos nossos portfólios, portanto é interessante conhecer índices que forneçam uma ideia da performance do portfólio. O índice mais conhecido é o *índice de Sharpe*. Este índice mede o excesso de retorno médio ganho em relação a uma taxa livre de risco (que no trabalho será a *Treasury bills*) por unidade de volatilidade. Numericamente temos que:

$$\widehat{SR_k} = \frac{\hat{\mu_k} - R_f}{\hat{\sigma_k}}$$

Em seguida, tem-se o *índice de Treynor*, que, analogamente ao índice de Sharpe, mensura o retorno ganho em excesso. A única diferença entre eles é que o índice de Treynor usa Beta para mensurar as incertezas. Também, pode-se dizer que quanto mais alto o valor deste índice, melhor a performance do portfólio. Em seguida, temos a fórmula:

$$\widehat{TR_k} = \frac{\hat{\mu_k} - R_f}{\beta}$$

Por último, tem-se *Certainty-Equivalent Return(CEQ)*. Este é um retorno garantido (livre de risco) que um determinado agente estaria disopsto a aceitar; ao invés de correr risco para uma chance de maior de retorno. Em seguida, tem-se a expressão para calcular a taxa CEQ, na qual γ pode assumir diversos valores, neste trabalho será usado sempre com o valor 1.

$$CEQ_k = \hat{\mu_k} - \frac{\gamma \hat{\sigma_k}}{2}$$

2.1 Desempenho de alguns setores do mercado americano

Para compreender toda a tecnicalidade inerente ao cálculo das medidas de desempenho, repetimos um estudo realizado por [2]. Os dados para esse exemplo são provenientes do *S&P500 sector data* e seguem o Padrão de Classificação Industrial Global. Fazem parte dele os 10 seguintes setores: 1, Consumo Facultativo; 2, Consumo Básico; 3, Energia; 4, Financeiro; 5, Saúde; 6, Industrial; 7, Tecnologia da Informação; 8, Materiais; 9 Telecomunicações e 10, Utilidades. Os dados foram obtidos semanalmente do período de 22 de setembro de 1989 a 28 de dezembro de 2012. Além destes, foi obtido o retorno das T-bills (usadas como taxa livre de risco). Os resultados podem ser vistos na seguinte tabela:

	SR	CEQ(%)	TR	Ann.Ret.(%)	Ann.risk(%)
Índice Obtido	0.0253	-1.5831	0.0095	0.4815	20.3208
Índice real	0.025	-1.59	0.010	0.51	20.50

Esse estudo foi fundamental para que pudéssemos verificar o cálculo dos índices e então ter maior confiabilidade na comparação de métodos de formação de portfólios.

3 Aplicação à dados do Ibovespa

Para essa aplicação consideramos dados diários do preço de fechamento dos ativos que compõem o índice Ibovespa, mas que tem histórico de comercialização no período de janeiro de 2010 a setembro de 2016. Dentre esses ativos, consideramos os mais líquidos, ou seja, desconsideramos os ativos em que não houve comercialização em mais de 5% dos dias. Assim, o banco de dados é composto por 63 ativos (ver Figura 1) com 1652 observações. Os dados foram obtidos da base de dados do google utilizando o pacote *"quantmod"* do R. Para a montagem do portfólio foram calculados os retornos diários e utilizadas 252 observações, o que é padrão para dados diários, representando o período de um ano. Os clusters são obtidos utilizando o método hierárquico (*"hclust"*) e a medida de distância utilizada é uma *Garch-Model-based*, apresentada [4], que mede a distância entre coeficientes garchs. Particularmente para um processo GARCH(1,1) a distância entre duas séries temporais é definida como

$$d_{arch} = \left[\frac{\alpha_{1,1}^2}{1-\beta_{1,1}^2} + \frac{\alpha_{2,1}^2}{1-\beta_{2,1}^2} - \frac{2\alpha_{1,1}\alpha_{2,1}}{1-\beta_{1,1}\beta_{2,1}} \right]^{1/2}$$

em que α_i's e β_i's, i=1,2, são os coeficientes do modelo GARCH(1,1) da série i. Um dos problemas ao se utilizar o método hieráquico é que o número de grupos existente não é conhecido. Assim, é necessário que o usuário defina previamente o número de grupos ou use alguma técnica empírica para encontrar um número ótimo de grupos. É nesse ponto que utilizamos o teste de homogeneidade baseado em U-estatísticas proposto por [5]. Basicamente, seguimos as ramificações do método hierárquico e definimos

os grupos. Enquanto os grupos forem significativamente distintos, continuamos subdividindo. Se o teste atesta que os grupos são homogêneos, paramos de procurar mais grupos naquele ramo. Fazendo isso, encontramos 8 grupos significativos, os quais podem ser vistos em cada uma das caixas da Figura 1. Sendo homogêneo, é natural que dentro de cada grupo os ativos recebam a mesma carga de investimento, ou seja, os pesos dentro de cada grupo são alocados seguindo a estratégia 1/N. Para definir quanto vamos alocar em cada grupo, vamos adotar a estratégia de [6] e distribui de acordo com o índice de Sharpe (ou Treynor) de cada grupo. Os grupos que apresentam maior índice de Sharpe recebem uma fatia maior do capital e os grupos com índices menores recebem uma penalização. Propomos o seguinte critério para obtenção dos pesos de cada grupo: Seja SR_i, $i = 1, \cdots, ng$ o Sharpe Ratio de cada grupo, em que ng é o número de grupos (aqui $ng = 8$). Defina μ_{SR} como sendo a média dos valores absolutos de SR_i. Considere

$$SRc_i = \frac{SR_i}{cp * \mu_{SR}},$$

uma correção do Sharpe Ratio de cada grupo, em que $cp > 0$ é o *coeficiente de penalização* que permitirá valorizar grupos com maior Sharpe e penalizar os grupos com Índice baixo ou negativo. Agora defina os pesos como sendo

$$w_i = \frac{e^{SRc_i}}{\sum_{i=1}^{ng}(e^{SRc_i})}.$$

Nessa aplicação os grupos de 1 a 8 estão representados pelas cores, *preto, magenta, vermelho, verde, cyan, amarelo, azul* e *cinza*. Os índices de Sharpe para cada grupo, calculado para 252 observações foi SR={-0.018690361, 0.009399394, 0.034921843, -0.040598077, 0.039889066, 0.044910269, 0.076644713, -0.015720382}. Usando a estratégia acima com $cp = 0.7$, os pesos para cada um dos 8 grupos ficaram definidos como w={0.011473638, 0.035996137, 0.101724901, 0.004703584, 0.124518831, 0.152755559, 0.555879349, 0.012948001}. Com essa estratégia, obtemos um Sharpe ratio para o portfólio via agrupamento de 0.002416785, contra -0.01989971 do método 1/N e -0.001533676 do método de mínima variância.

4 Conclusão

Neste trabalho propusemos uma estratégia para construção de portfólios baseada em agrupamento de séries temporais através do método hierárquico, juntamente com teste para homogeneidade de grupos baseado em U-estatísticas. A alocação de capital nos grupos foi definida de uma maneira diferente da normamente adotada na literatura. Optamos por um critério que distribui os pesos de acordo com o desempenho de cada grupo, segundo alguma medida de desempenho. Os resultados mostram que o

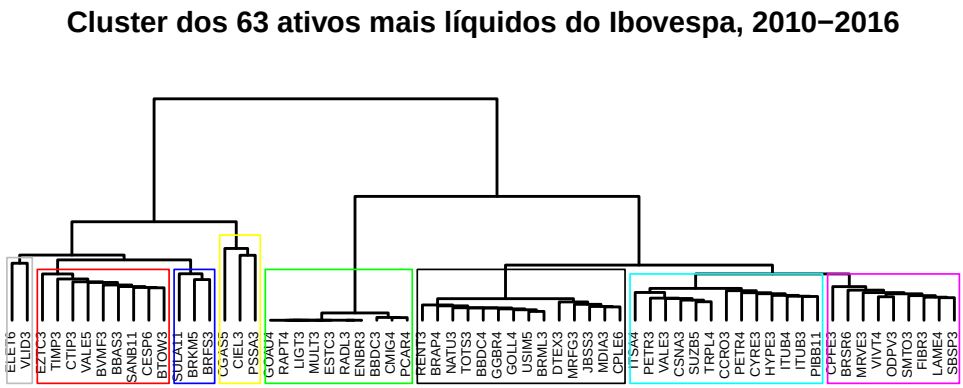

Figura 1: Agrupamento dos 63 ativos mais líquidos do Ibovespa no período de 01/2010 a 09/2016. A métrica utilizada é baseada na estimativa de parâmetros do modelo GARCH para a volatilidade dos retornos.

método de agrupamento, juntamente com o critério para alocação de capital, é superior à estratégia 1/N e ao método de mínima variância. Considerar rebalanceamento da carteira e custos de transação são pontos a serem abordados em pesquisa futura.

Referências

[1] D'Urso, P., Cappelli, C., Lallo, D. D., and Massari, R. (2013). Clustering of financial time series. *Physica A: Statistical Mechanics and its Applications*, 392(9):2114 – 2129.

[2] Koundouri, P., Kourogenis, N., Pittis, N., and Samartzis, P. (2016). Factor models of stock returns: Garch errors versus time-varying betas. *Journal of Forecasting*, 35(5):445–461. for.2387.

[3] Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91.

[4] Otranto, E. (2008). Clustering heteroskedastic time series by model-based procedures. *Computational Statistics and Data Analysis*, 52(10):4685 – 4698.

[5] Valk, M. and Pinheiro, A. (2012). Time-series clustering via quasi U-statistics. *J. Time Ser. Anal.*, 33(4):608–619.

[6] Zhang, J. and Maringer, D. (2010). *A Clustering Application in Portfolio Management*, pages 309–321. Springer Netherlands, Dordrecht.

Uso do MCMC para Estimação dos Parâmetros dos Processos ARFIMA(p, d, q)

Letícia Menegotto [1]
Cleber Bisognin[2]

Resumo: Neste trabalho estamos interessados em estudar os processos ARFIMA(p, d, q), mais especificamente comparamos alguns métodos de estimação dos parâmetros destes processo. Utilizamos o estimador proposto por Fox e Taqqu (1986) e Whittle (1951), o estimador proposto por Beran (1995), o estimador proposto por Shimotsu e Phillips (2005). Além disso, também propomos, para os processos ARFIMA(p, d, q), um novo estimador utilizado o método MCMC. Para comparação dos estimadores utilizamos o vício, erro quadrático médio e variância. Na maioria dos casos estudados e simulados o estimador MCMC apresentou desempenho tão bom quanto ou melhor comparado com os estimadores Whittle, Beran e LW.

Palavras-chave: *Processos ARFIMA, Longa Dependência, Estimação, Simulações de Monte Carlo*.

1 Introdução

O principal objetivo de se estudar uma série temporal é encontrar um modelo que nos possibilite fazer previsões sobre os futuros valores dessa série. Um dos modelos, amplamente utilizado com este objetivo, é o proposto por G. E. P. Box e G. M. Jenkins na década de 70, que sugerem descrever uma série temporal na forma de polinômios: são os chamados modelos autoregressivo integrado média móvel, denotado por ARIMA(p, d, q). A metodologia proposta consiste em identificar (estimar) os parâmetros do modelo e, posteriormente, analisar a adequação desse ajuste através da análise dos resíduos.

Neste trabalho estamos interessados em estudas séries temporais que apresentam a característica de longa dependência. A longa dependência é uma característica onde valores observados, de uma série temporal, em momentos distantes, são correlacionados. Esta característica foi identificada inicialmente por Hurst (1956) em seus estudos sobre a influência dos nível do Rio Nilo na fertilidade de suas margens. Também podemos encontrar esta característica em séries temporais de diversas áreas como Economia, Genoma, entre outras. Vários processos foram propostos afim de modelar séries temporais com esta característica. Mandelbrot e Ness (1968) apresentou o fractional Brownian motion e o fractional Gaussian noise process. Hosking (1981) apresentou uma generalização do modelo ARIMA(p, d, q), o chamado modelo autoregressivo fracionário integrado de m édia móvel, denotado por ARFIMA(p, d, q). No caso dos modelos ARIMA(p, d, q), o parâmetro $d \in \mathbb{N}$, para os modelos ARFIMA(p, d, q), o parâmetro $d \in \mathbb{R}$. Neste trabalho, estamos interessados em estudar os modelos ARFIMA(p, d, q) com o parâmetro $d \in (0, 0.5)$. Quando $d \in (0, 0.5)$, o modelo ARFIMA(p, d, q) é utilizado para analisar séries temporais com longa dependência, quando $d \in (0.5, 0)$ estamos na região de curta dependência e quando $d \in (0.5, 1)$, estamos na região de dependência intermediária.

2 Processos ARFIMA(p, d, q)

Nesta seção definimos os processos *Auto-regressivos fracionariamente integrados de médias móveis* denotados por ARFIMA (p, d, q) e apresentamos algumas de suas propriedades.

[1]UFRGS - Universidade Federal do Rio Grande do Sul. Bolsista de Iniciação Científica - BIC/UFRGS.
Email: *leticia.menegotto@gmail.com*
[2]UFRGS - Universidade Federal do Rio Grande do Sul. Email: *cbisognin@ufrgs.br*

Definição 2.1. Seja $\{X_t\}_{t\in\mathbb{Z}}$ um processo estocástico que satisfaz a equação

$$\phi(\mathcal{B})(1-\mathcal{B})^d(X_t-\mu)=\theta(\mathcal{B})\varepsilon_t, \qquad t\in\mathbb{Z} \tag{2.1}$$

onde μ é a média do processo, $\{\varepsilon_t\}_{t\in\mathbb{Z}}$ é um processo ruído branco, $\mathcal{B}$ é o operador defasagem ou de retardo, isto é, $\mathcal{B}^kX_t=X_{t-k}$, d é o *parâmetro ou grau de diferenciação*, $\phi(\cdot)$ e $\theta(\cdot)$ são polinômios de ordem p e q, respectivamente dados por

$$\phi(z)=\sum_{i=0}^{p}\phi_i z^i \quad \text{e} \quad \theta(z)=\sum_{j=0}^{q}\theta_j z^j, \tag{2.2}$$

com ϕ_i, $1\leq i\leq p$, e θ_j, $1\leq j\leq q$ pertencentes ao reais e $\phi_0=1=\theta_0$. Então o processo $\{X_t\}_{t\in\mathbb{Z}}$ é chamado de *processo auto-regressivo fracionariamente integrado de média móvel de ordem* (p,d,q), denotado por ARFIMA (p,d,q).

A seguir vamos apresentar propriedades fundamentais dos modelos ARFIMA (p,d,q). A prova do teorema a seguir pode ser encontrada em Brockwell e Davis (1991). Sem perda de generalidade utilizamos $\mu\equiv 0$.

Teorema 2.1. *Seja $\{X_t\}_{t\in\mathbb{Z}}$ um processo ARFIMA(p,d,q), dado pela Definição 2.1. Suponha que $d\in(-0.5,0.5)$ e que as equações $\phi(z)=0$ e $\theta(z)=0$ não possuem raízes em comum e as suas raízes estão fora do círculo unitário. Então valem as seguintes afirmações.*

(i) O processo $\{X_t\}_{t\in\mathbb{Z}}$ é estacionário e inversível e possui representações autoregressiva e média móvel infinitas dadas, respectivamente, por

$$\sum_{k\geq 0}\pi_k X_{t-k}=\varepsilon_t \quad e \quad X_t=\sum_{k\geq 0}\psi_k\varepsilon_{t-k}$$

onde π_k e ψ_k são os coeficientes de $\mathcal{B}^k$ nas expensões

$$\pi(B)=\frac{\phi(\mathcal{B})}{\theta(\mathcal{B})}(1-\mathcal{B})^d \quad e \quad \psi(B)=\frac{\theta(\mathcal{B})}{\phi(\mathcal{B})}(1-\mathcal{B})^{-d},$$

respectivamente.

(ii) A função densidade espectral do $\{X_t\}_{t\in\mathbb{Z}}$ é dada por

$$f_X(\omega)=\frac{\sigma_\varepsilon^2}{2\pi}\frac{|\theta(e^{-i\omega})|^2}{|\phi(e^{-i\omega})|^2}\left[2\operatorname{sen}\left(\frac{\omega}{2}\right)\right]^{-2d}, \quad \textit{para todo} \quad 0<\omega\leqslant\pi. \tag{2.3}$$

Além disso, $\lim_{w\to 0} f_X(\omega)=\dfrac{\sigma_\varepsilon^2}{2\pi}\dfrac{|\theta(1)|^2}{|\phi(1)|^2}|w|^{-2d}$.

(iii) Quando $0<d<0.5$, as funções de autocovariância e autocorrelação assintóticas do processo $\{X_t\}_{t\in\mathbb{Z}}$ são dadas, respectivamente, por (quando $|k|\to\infty$)

$$\gamma_X(k)\sim c_\gamma(d,\phi,\theta)|k|^{2d-1},$$

onde

$$c_\gamma(d,\phi,\theta)=\frac{\sigma_\varepsilon^2}{\pi}\frac{|\theta(1)|^2}{|\phi(1)|^2}\Gamma(1-2d)\operatorname{sen}(\pi d)$$

e

$$\rho_X(k) \sim c_\rho(d,\phi,\theta)|k|^{2d-1},$$

onde

$$c_\rho(d,\phi,\theta) = \frac{c_\gamma(d,\phi,\theta)}{\int_\pi^\pi f_X(\omega)dw}.$$

Maiores detalhes sobre os processos ARFIMA(p,d,q) podem ser encontrados em Hosking (1981) e Brockwell e Davis (1991).

3 Estimação de Parâmetros

Na literatura de séries temporais com propriedade de longa dependência existem diversos métodos para a estimação dos parâmetros. Em geral, os estimadores podem ser semi-paramétricos e paramétricos. Neste trabalho utilizaremos o estimador proposto por Fox e Taqqu (1986) e Whittle (1951) na qual utiliza o método da função de máxima verossimilhança aproximada e é denotado por Whittle, o estimador proposto por Beran (1995) que utiliza a representação autoregressiva infinita do processo e o estimador proposto por Shimotsu e Phillips (2005) que utiliza o comportamento da função densidade espectral próximo da origem, denotado por LW (*local Whittle*). Também propomos um novo estimador utilizado o método MCMC para estimar os parâmetros do processo ARFIMA(p,d,q) via estimador de Whittle. Por serem estimadores bastante conhecidos na literatura de séries temporais com longa dependência apresentamos apenas o estimador MCMC.

3.1 Estimator Whittle MCMC

Neste método para calcular o estimador de verossimilhança utlizando a aproximação de Whittle utilizamos o método de Monte Carlo baseado em Cadeias de Markov. Seja C uma constante tal que $C = \int_{-\pi}^{\pi} I_n(\lambda)d\lambda$. Assim, assim temos que a equação

$$\sigma_n^2(\boldsymbol{\eta}) = \int_{-\pi}^{\pi} \frac{I_n(\lambda)}{f_X(\lambda,\boldsymbol{\eta})}d\lambda, \tag{3.4}$$

por ser reescrita como

$$\sigma_n^2(\boldsymbol{\eta}) = C\int_{-\pi}^{\pi} \frac{f(\lambda)}{f_X(\lambda,\boldsymbol{\eta})}d\lambda = C\,\mathbb{E}\left(\frac{1}{f_X(\lambda,\boldsymbol{\eta})}\right), \tag{3.5}$$

onde $f(\lambda) = \frac{1}{C}I_n(\lambda)$ é a função densidade em $[-\pi,\pi]$, $I_n(\cdot)$ é a função periodograma e $f_X(\lambda,\cdot)$ é a função densidade espectral dos processos ARFIMA(p,d,q) dada pela equação (2.3). O valor esperado em (3.5) pode ser aproximado pela média empírica

$$\overline{\sigma}_n^2(\boldsymbol{\eta}) = \frac{1}{N}\sum_{j=1}^{N} \frac{1}{f_X(\lambda_j,\boldsymbol{\eta})}, \tag{3.6}$$

onde N é suficientemente grande para satisfazer a lei dos grandes números. Aqui consideramos $N = 10000$ e o algoritmo de Metropolis-Hastings para gerar a amostra $(\lambda_1,\cdots,\lambda_N)$ como segue

i) Gerar λ_k, para $k \in \{1,\cdots,N\}$;

ii) Gerar Y_k de uma distribuição uniforme no intervalo $[-\pi,\pi]$ e denote o valor obtido por y_k, para $k \in \{1,\cdots,N\}$;

iii) Tome

$$\lambda_{k+1} = \begin{cases} y_k, & \text{com probabilidade } p(\lambda_k,y_k), \\ \lambda_k, & \text{com probabilidade } 1-p(\lambda_k,y_k), \end{cases}$$

onde $p(\lambda_k, y_k) = \min\left\{\frac{f(y_k)}{f(\lambda_k)}, 1\right\} = \min\left\{\frac{I_n(y_k)}{I_n(\lambda_k)}, 1\right\}$. Então, o estimador para $\boldsymbol{\eta}$, denotado por MCMC, é o valor $\widehat{\boldsymbol{\eta}}$ o qual minimiza (3.6) com respeito a $\boldsymbol{\eta}$. Maiores detalhes sobre o algoritmo de Metropolis-Hastings ver Gilks et al. (1998). Ndongo et al. (2010) consideram este método para estimar os parâmetros dos processos ARFISMA$(0, d, 0) \times (0, D, 0)_s - S\alpha S$.

4 Simulações de Monte Carlo

Nesta subseção apresentamos os resultados de simulações de Monte Carlo, mais especificamente da estimação dos parâmetros do processo ARFIMA(p, d, q). Todas as simulações realizadas neste trabalho foram realizadas através do software estatístico R. Cada estatística provem de $re = 500$ replicações com $n \in \{250, 500, 1000, 2000\}$, $d \in \{0.05, 0.3, 0.45\}$, $p \in \{0, 1\}$, $q \in \{0, 1\}$, $\phi_1 \in \{-0.5, 0.5\}$ e $\theta_1 \in \{-0.5, 0.5\}$. No caso do estimador de Shimotsu e Phillips (2005), necessitamos especificar o número de frequências de Fourier serão utilizadas na estimação dos parâmetros. O número de frequências utilizadas é dada por $m = n^\beta$. Nestas simulações utilizamos $\beta \in \{0.60, 0.65, 0.70, 0.75, 0.80, 0.85\}$.

Tabela 4.1: Resultados de estimação para um processo ARFIMA(p, d, q) utilizando o Estimador LW, quando $p = 0 = q$, $d \in \{0.05, 0.3, 0.45\}$, $\beta \in \{0.60, 0.65, 0.70, 0.75, 0.80, 0.85\}$ e $n \in \{250, 500, 1000\}$.

		$d = 0.05$				$d = 0.3$				$d = 0.45$			
β	n	Média	Vício	EQM	VAR	Média	Vício	EQM	VAR	Média	Vício	EQM	VAR
0.60	250	0.1094	0.0594	0.0098	0.0063	0.2878	-0.0122	0.0137	0.0136	0.4213	-0.0287	0.0098	0.0090
	500	0.0888	0.0388	0.0052	0.0037	0.2969	-0.0031	0.0081	0.0081	0.4376	-0.0124	0.0052	0.0050
	1000	0.0746	0.0246	0.0031	0.0025	0.2994	-0.0006	0.0052	0.0052	0.4452	-0.0048	0.0029	0.0029
0.65	250	0.0948	0.0448	0.0063	0.0043	0.2896	-0.0104	0.0104	0.0103	0.4260	-0.0240	0.0070	0.0064
	500	0.0819	0.0319	0.0040	0.0030	0.2970	-0.0030	0.0062	0.0062	0.4418	-0.0082	0.0039	0.0038
	1000	0.0663	0.0163	0.0021	0.0018	0.3001	0.0001	0.0032	0.0032	0.4468	-0.0032	0.0024	0.0024
0.70	250	0.0830	0.0330	0.0043	0.0033	0.2874	-0.0126	0.0078	0.0077	0.4289	-0.0211	0.0056	0.0051
	500	0.0727	0.0227	0.0028	0.0023	0.2967	-0.0033	0.0041	0.0041	0.4422	-0.0078	0.0030	0.0029
	1000	0.0621	0.0121	0.0015	0.0014	0.3003	0.0003	0.0021	0.0021	0.4477	-0.0023	0.0019	0.0019
0.75	250	0.0760	0.0260	0.0032	0.0025	0.2869	-0.0131	0.0056	0.0055	0.4305	-0.0195	0.0044	0.0040
	500	0.0644	0.0144	0.0018	0.0016	0.2948	-0.0052	0.0029	0.0029	0.4423	-0.0077	0.0023	0.0022
	1000	0.0561	0.0061	0.0011	0.0011	0.2980	-0.0020	0.0016	0.0016	0.4469	-0.0031	0.0015	0.0015
0.80	250	0.0683	0.0183	0.0023	0.0019	0.2819	-0.0181	0.0044	0.0041	0.4264	-0.0236	0.0039	0.0033
	500	0.0594	0.0094	0.0014	0.0013	0.2902	-0.0098	0.0021	0.0020	0.4387	-0.0113	0.0019	0.0018
	1000	0.0524	0.0024	0.0008	0.0008	0.2935	-0.0065	0.0012	0.0011	0.4429	-0.0071	0.0011	0.0011
0.85	250	0.0595	0.0095	0.0015	0.0014	0.2716	-0.0284	0.0037	0.0029	0.4121	-0.0379	0.0043	0.0028
	500	0.0540	0.0040	0.0010	0.0010	0.2807	-0.0193	0.0018	0.0015	0.4257	-0.0243	0.0021	0.0015
	1000	0.0490	-0.0010	0.0006	0.0006	0.2865	-0.0135	0.0010	0.0008	0.4315	-0.0185	0.0012	0.0008

Tabela 4.2: Resultados de estimação para um processo ARFIMA(p, d, q) utilizando os estimador Whittle, Beran e MCMC, quando $p = 0 = q$, $d \in \{0.05, 0.30, 0.45\}$ e $n \in \{250, 500, 1000\}$.

	Whittle				Beran				MCMC			
					$d = 0.05$							
n	Média	Vício	EQM	VAR	Média	Vício	EQM	VAR	Média	Vício	EQM	VAR
250	0.0632	0.0132	0.0017	0.0015	0.0475	-0.0025	0.0006	0.0006	0.0611	0.0111	0.0015	0.0014
500	0.0556	0.0056	0.001	0.001	0.0481	-0.0019	0.0003	0.0003	0.0562	0.0062	0.001	0.001
1000	0.0496	-0.0004	0.0005	0.0005	0.0461	-0.0039	0.0005	0.0005	0.0500	0.000	0.0006	0.0006
					$d = 0.3$							
n	Média	Vício	EQM	VAR	Média	Vício	EQM	VAR	Média	Vício	EQM	VAR
250	0.2954	-0.0046	0.0031	0.0031	0.2798	-0.0202	0.0034	0.003	0.2981	-0.0019	0.0028	0.0028
500	0.2993	-0.0007	0.0014	0.0014	0.2906	-0.0094	0.0015	0.0014	0.2992	-0.0008	0.0015	0.0015
1000	0.3009	0.0009	0.0007	0.0007	0.296	-0.004	0.0007	0.0007	0.3011	0.0011	0.0008	0.0008
					$d = 0.45$							
n	Média	Vício	EQM	VAR	Média	Vício	EQM	VAR	Média	Vício	EQM	VAR
250	0.4442	-0.0058	0.0024	0.0024	0.4304	-0.0196	0.003	0.0026	0.4508	0.0008	0.0023	0.0023
500	0.4519	0.0019	0.0012	0.0012	0.4429	-0.0071	0.0013	0.0012	0.4539	0.0039	0.0013	0.0013
1000	0.4533	0.0033	0.0007	0.0007	0.4473	-0.0027	0.0007	0.0007	0.4535	0.0035	0.0008	0.0008

Tabela 4.3: Resultados de estimação para um processo ARFIMA(p, d, q) utilizando os estimador Whittle, Beran e MCMC, quando $p = 1$ e $q = 0$, $d = 0.30$, $\phi_1 \in \{-0.5, 0.5\}$ e $n \in \{250, 500, 1000\}$.

$n = 250$												
	$\phi_1 = 0.5$						$\phi_1 = -0.5$					
	Whittle		Beran		MCMC		Whittle		Beran		MCMC	
	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\phi}_1$	$\widehat{d}$
Média	0.5081	0.2809	0.4295	0.2620	0.5083	0.2850	-0.4944	0.2937	-0.4855	0.2724	-0.4947	0.2980
Vício	0.0081	-0.0191	-0.0705	-0.0380	0.0083	-0.0150	0.0056	-0.0063	0.0145	-0.0276	0.0053	-0.0020
EQM	0.0169	0.0165	0.1366	0.0245	0.0164	0.0146	0.0046	0.0046	0.0049	0.0054	0.0046	0.0043
VAR	0.0168	0.0162	0.1319	0.0231	0.0164	0.0144	0.0046	0.0046	0.0047	0.0047	0.0046	0.0043
$n = 500$												
	Whittle		Beran		MCMC		Whittle		Beran		MCMC	
	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\phi}_1$	$\widehat{d}$
Média	0.4985	0.2929	0.4122	0.2800	0.4982	0.2950	-0.4948	0.2947	-0.4898	0.2830	-0.4949	0.2969
Vício	-0.0015	-0.0071	-0.0878	-0.0200	-0.0018	-0.0050	0.0052	-0.0053	0.0102	-0.0170	0.0051	-0.0031
EQM	0.0122	0.0119	0.1410	0.0190	0.0134	0.0113	0.0024	0.0022	0.0025	0.0025	0.0026	0.0022
VAR	0.0122	0.0118	0.1336	0.0186	0.0134	0.0113	0.0024	0.0022	0.0024	0.0022	0.0026	0.0022
$n = 1000$												
	Whittle		Beran		MCMC		Whittle		Beran		MCMC	
	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\phi}_1$	$\widehat{d}$
Média	0.5066	0.2896	0.4576	0.2856	0.5032	0.2945	-0.4993	0.2994	-0.4971	0.2904	-0.4992	0.2998
Vício	0.0066	-0.0104	-0.0424	-0.0144	0.0032	-0.0055	0.0007	-0.0006	0.0029	-0.0096	0.0008	-0.0002
EQM	0.0081	0.0075	0.0659	0.0126	0.0089	0.0075	0.0011	0.0009	0.0010	0.0010	0.0013	0.0011
VAR	0.0081	0.0074	0.0647	0.0125	0.0089	0.0075	0.0011	0.0009	0.0010	0.0009	0.0013	0.0011

Tabela 4.4: Resultados de estimação para um processo ARFIMA(p, d, q) utilizando os estimador Whittle, Beran e MCMC, quando $p = 0$ e $q = 1$, $d = 0.3$, $\theta_1 \in \{-0.5, 0.5\}$ e $n \in \{250, 500, 1000\}$.

$n = 250$												
	$\theta_1 = 0.5$						$\theta_1 = -0.5$					
	Whittle		Beran		MCMC		Whittle		Beran		MCMC	
	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{d}$	$\widehat{\theta}_1$
Média	0.2606	0.4577	0.2266	0.4335	0.2672	0.4610	0.2998	-0.4959	0.2961	-0.4342	0.3013	-0.4942
Vício	-0.0394	-0.0423	-0.0734	-0.0665	-0.0328	-0.0390	-0.0002	0.0041	-0.0039	0.0658	0.0013	0.0058
EQM	0.0163	0.0191	0.0223	0.0252	0.0142	0.0173	0.0042	0.0051	0.0076	0.0766	0.0044	0.0060
VAR	0.0148	0.0174	0.0170	0.0208	0.0132	0.0158	0.0042	0.0051	0.0076	0.0724	0.0044	0.0059
$n = 500$												
	Whittle		Beran		MCMC		Whittle		Beran		MCMC	
	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{d}$	$\widehat{\theta}_1$
Média	0.2821	0.4777	0.2551	0.4569	0.2852	0.4798	0.2991	-0.4980	0.3059	-0.4360	0.3020	-0.4983
Vício	-0.0179	-0.0223	-0.0449	-0.0431	-0.0148	-0.0202	-0.0009	0.0020	0.0059	0.0640	0.0020	0.0017
EQM	0.0110	0.0126	0.0142	0.0172	0.0089	0.0109	0.0023	0.0025	0.0055	0.0657	0.0025	0.0033
VAR	0.0107	0.0121	0.0122	0.0154	0.0087	0.0105	0.0023	0.0025	0.0055	0.0617	0.0025	0.0033
$n = 1000$												
	Whittle		Beran		MCMC		Whittle		Beran		MCMC	
	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{d}$	$\widehat{\theta}_1$
Média	0.2932	0.4921	0.2717	0.4740	0.2921	0.4898	0.2993	-0.4993	0.3120	-0.4264	0.2992	-0.5015
Vício	-0.0068	-0.0079	-0.0283	-0.0260	-0.0079	-0.0102	-0.0007	0.0007	0.0120	0.0736	-0.0008	-0.0015
EQM	0.0059	0.0069	0.0068	0.0082	0.0056	0.0068	0.0010	0.0012	0.0046	0.0729	0.0013	0.0020
VAR	0.0059	0.0068	0.0061	0.0075	0.0056	0.0067	0.0010	0.0012	0.0044	0.0676	0.0013	0.0020

5 Conclusões

No casos da estimação do parâmetro d dos processos ARFIMA(p, d, q), com $p = 0 = q$, o estimador LW (*Local Whittle*), ver Tabela 4.1, obteve os melhores resultados $d = 0.05$ quando $\beta \in \{0.8, 0.85\}$, independente do tamanho de n. Para $d = 0.3$ todas as simulações apresentam bons resultados. Para $d = 0.45$, as estimativas apresentam um vício maior, independente do tamanho amostral e de β. Comparando as estimativas de LW com os outros estimadores, quando $\beta \in \{0.8, 0.85\}$, as estimativas são equivalentes e apresentem semelhante EQM e VAR.

Para os processos ARFIMA(p, d, q), com $p = 0 = q$, ver Tabela 4.2, comparando os estimador MCMC

Tabela 4.5: Resultados de estimação para um processo ARFIMA(p, d, q) utilizando os estimador Whittle, Beran e MCMC, quando $p = 1 = q$, $d = 0.3$, $\phi_1 = -0.5$ e $\theta_1 = 0.5$ e $n \in \{250, 500, 1000\}$.

$n = 250$									
	Whittle			Beran			MCMC		
	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\theta}_1$
Média	-0.5023	0.2505	0.4359	-0.5024	0.2265	0.4252	-0.5023	0.2709	0.4543
Vício	-0.0023	-0.0495	-0.0641	-0.0024	-0.0735	-0.0748	-0.0023	-0.0291	-0.0457
EQM	0.0052	0.0167	0.0256	0.0053	0.0231	0.0314	0.0051	0.0144	0.0227
VAR	0.0052	0.0143	0.0215	0.0053	0.0177	0.0259	0.0051	0.0136	0.0206
$n = 500$									
	Whittle			Beran			MCMC		
	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\theta}_1$
Média	-0.4974	0.2730	0.4694	-0.4997	0.2452	0.4472	-0.4977	0.2838	0.4780
Vício	0.0026	-0.0270	-0.0306	0.0003	-0.0548	-0.0528	0.0023	-0.0162	-0.0220
EQM	0.0028	0.0107	0.0148	0.0028	0.0142	0.0180	0.0031	0.0108	0.0148
VAR	0.0028	0.0100	0.0139	0.0028	0.0112	0.0152	0.0031	0.0106	0.0143
$n = 1000$									
	Whittle			Beran			MCMC		
	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\theta}_1$
Média	-0.4999	0.2881	0.4865	-0.5013	0.2668	0.4679	-0.5016	0.2931	0.4897
Vício	0.0001	-0.0119	-0.0135	-0.0013	-0.0332	-0.0321	-0.0016	-0.0069	-0.0103
EQM	0.0012	0.0057	0.0075	0.0017	0.0067	0.0090	0.0015	0.0062	0.0087
VAR	0.0012	0.0056	0.0073	0.0017	0.0056	0.0080	0.0015	0.0062	0.0087

Tabela 4.6: Resultados de estimação para um processo ARFIMA(p, d, q) utilizando os estimador Whittle, Beran e MCMC, quando $p = 1 = q$, $d = 0.3$, $\phi_1 = 0.5$ e $\theta_1 = -0.5$ e $n \in \{250, 500, 1000\}$.

$n = 250$									
	Whittle			Beran			MCMC		
	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\theta}_1$
Média	0.5111	0.2711	-0.4890	0.3556	0.2885	-0.4681	0.5124	0.2755	-0.4876
Vício	0.0111	-0.0289	0.0110	-0.1444	-0.0115	0.0319	0.0124	-0.0245	0.0124
EQM	0.0196	0.0187	0.0062	0.2045	0.0272	0.0774	0.0186	0.0140	0.0112
VAR	0.0195	0.0179	0.0061	0.1841	0.0272	0.0765	0.0184	0.0134	0.0110
$n = 500$									
	Whittle			Beran			MCMC		
	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\theta}_1$
Média	0.5127	0.2759	-0.4927	0.3890	0.2884	-0.4866	0.5002	0.2867	-0.5040
Vício	0.0127	-0.0241	0.0073	-0.1110	-0.0116	0.0134	0.0002	-0.0133	-0.0040
EQM	0.0144	0.0130	0.0029	0.1590	0.0213	0.0450	0.0176	0.0111	0.0095
VAR	0.0143	0.0125	0.0029	0.1470	0.0212	0.0449	0.0176	0.0110	0.0095
$n = 1000$									
	Whittle			Beran			MCMC		
	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\theta}_1$	$\widehat{\phi}_1$	$\widehat{d}$	$\widehat{\theta}_1$
Média	0.5086	0.2856	-0.4978	0.4327	0.2848	-0.4703	0.4950	0.2966	-0.5076
Vício	0.0086	-0.0144	0.0022	-0.0673	-0.0152	0.0297	-0.0050	-0.0034	-0.0076
EQM	0.0077	0.0071	0.0013	0.1189	0.0144	0.0533	0.0114	0.0072	0.0071
VAR	0.0077	0.0069	0.0013	0.1145	0.0142	0.0526	0.0114	0.0072	0.0070

com os estimadores Beran, Whittle, em relação ao EQM, o estimador MCMC possui EQM muito semelhante aos estimadores Whittle e Beran. Em relação ao vício, na maioria dos casos possui o vício um pouco menor em relação ao Whittle. Em comparação com o estimador Beran, para d pequeno possui vício um pouco maior, mas quando d cresce, o estimador MCMC possui menor vício.

Para os processos ARFIMA(p, d, q), com $p = 1 = q$, nos casos em que $\phi_1 = -0.5$ e $\theta_1 = 0.5$, ver Tabela 4.5, o estimador MCMC apresenta melhores resultados, no sentido de menor vício, EQM e VAR, que o estimador BERAN, para todos os tamanhos amostrais, resultados melhores que o estimador Whittle para amostras de tamanho $n \in \{250, 500\}$ e resultados semelhantes aos do Whittle quando $n = 1000$.

Analisando a estimação dos parâmetros dos processos ARFIMA(p, d, q), com $p = 1$, $q = 0$ e $d = 0.3$, nos casos em que $\phi_1 = 0.5$, ver Tabela 4.3, o estimador MCMC apresenta melhores resultados, no sentido de menor vício, EQM e VAR, que o estimador BERAN, para todos os tamanhos amostrais, e resultados

semelhantes aos do Whittle para todos os tamanhos amostrais. Para $\phi_1 = -0.5$, o estimador MCMC apresenta melhores resultados que BERAN para todos os tamanhos amostrais no sentido de menor vício, e resultados parecidos aos obtidos com o estimador de Whittle em todos os quesitos e em todos os tamanhos amostrais. Para os processos ARFIMA(p, d, q), com $p = 0$, $q = 1$ e $d = 0.3$, nos casos em que $\theta_1 = 0.5$, ver Tabela 4.4, o estimador MCMC apresenta melhores resultados, no sentido de menor vício, EQM e VAR, que o estimador BERAN, para todos os tamanhos amostrais. Quando comparado com o estimador de Whittle, MCMC é melhor estimador no sentido de menor vício, EQM e variância quando $n \in \{250, 500\}$ e resultados semelhantes aos do Whittle quando $n = 1000$. Quando $\theta_1 = -0.5$, o estimador MCMC é melhor em todos os sentidos e tamanhos amostrais que o BERAN, e possui resultados parecidos com os de Whittle para todos os tamanhos amostrais. Para os processos ARFIMA(p, d, q), com $p = 1 = q$, nos casos em que $\phi_1 \in \{-0.5, 0.5\}$ e $\theta_1 \in \{-0.5, 0.5\}$, ver Tabela 4.5 e 4.6, o estimador MCMC, em geral, apresenta melhores resultados, no sentido de menor vício, EQM e VAR, que o estimador BERAN, para todos os tamanhos amostrais e resultados semelhantes que o estimador Whittle para amostras de todos os tamanhos.

Referências

[1] Beran, J. (1995). "Maximum Likelihood Estimation of the Differencing Parameter for Invertible Short and Long Memory Autoregressive Integrated Moving Average Models". *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. **57**(4), pp. 659-672.

[2] Brockwell, P.J. e R.A. Davis (1991). *Time Series: Theory and Methods*. New York: Springer-Verlag.

[3] Fox, R. e M.S. Taqqu (1986). "Large-sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series". *The Annals of Statistics*, Vol. **14**, pp. 517-532.

[4] Hosking, J.R.M. (1981). "Fractional Differencing". *Biometrika*, Vol. **68**(1), pp. 165-176.

[5] Shimotsu, K. e P.C.B. Phillips (2005). "Exact Local Whittle Estimation o Fractional Integration". *The Annals of Statistics*, Vol. **33**(4), pp. 1890-1933.

O aprendizado de conceitos de Estatística através de um estudo sobre os óbitos dos escravos do Rio Grande do Sul no séc. XIX: uma experiência interdisciplinar

Leila Inês Pagliarini de Mello [1]

Luciana Neves Nunes[2]

Resumo: Apresenta-se um trabalho interdisciplinar desenvolvido em uma escola privada do município de Canoas, com o tema as epidemias e a escravidão na província de São Pedro no século XIX. O interesse no tema deste trabalho de pesquisa justifica-se na tentativa de conhecer um pouco das condições de vida e morte dos escravos em nosso estado e de poder contribuir para a compreensão de que as doenças que atingiam esse grupo social não se restringiam a ele. Considerando esse interesse e a possibilidade de realização dessa pesquisa num curto espaço de tempo, os alunos do sétimo ano foram convidados a fazer uma experiência de pesquisa de iniciação científica, orientada pelos seus professores de Matemática, História, Geografia e Ciências. Com esse intuito, em Matemática foram orientados a criar um banco de dados sobre os registros de óbitos dos escravos que constam no livro "Africanos na Santa Casa de Porto Alegre: Óbitos dos escravos sepultados no cemitério da Santa Casa (1850-1885)" uma obra de FLORES (2007) que traz dados relativos aos escravos que foram sepultados no Cemitério Santa Casa, entre os anos de 1850 e 1885, somando aproximadamente sete mil registros de óbitos. Tais registros contêm informações sobre a data do falecimento, prenome de quem morreu, idade, naturalidade ou nação, estado civil, a causa da morte e o sexo. A causa da morte tem importância fundamental, principalmente, pelo estudo das doenças, que são indicativos das condições de vida desses sujeitos no local e período analisado. Através da análise de dados e uso de conceitos de estatística descritiva de estudos históricos, foram avaliadas as causas de óbito, ou seja, as doenças, obtidas através da análise dos registros de óbitos. Com isso, pretendeu-se conhecer as condições de vida dos escravos: Quem eram? Como viviam e do que morriam Para atingir os objetivos propostos, foram seguidas as seguintes etapas: elaboração de um formulário, que foi preenchido com os dados de uma amostra aleatória sobre os óbitos de escravos e, a partir disso, criou-se o banco de dados, tabelas e gráficos; leituras teóricas sobre escravidão, estatística, demografia e doenças; visita ao Arquivo Histórico da Santa Casa de Porto Alegre; análise dos gráficos e discussões sobre os resultados a respeito dos escravos, seu estilo de vida, importância, etc.; redação de textos sobre causa mortis dos escravos e as doenças que acometiam os escravos, visando entender as condições de vida destes. Analisou-se 1559 casos de óbitos. Destes, 63% dos escravos eram crioulos, fruto da reprodução endógena, 35% eram africanos e 2% tem a origem desconhecida. Observa-se

[1]Email: *by.leila@gmail.com*

[2]UFRGS - Universidade Federal do Rio Grande do Sul. Email: *lunenu@gmail.com.*

que entre as principais causas de morte destacaram-se as doenças infectocontagiosas e parasitárias que acometeram 33% dos escravos. Observou-se que entre as doenças, algumas eram desconhecidas ou mal definidas, pois 19% dos óbitos tinham como causa mortis sintomas mal definidos, como, por exemplo, a tosse. As disciplinas envolvidas trabalharam com diferentes questões: etnias, escravidão, preconceitos, mortes diferenciais, padrões demográficos e epidemias de época, relacionando-as com as causas das mortes dos cativos, visando instigar e preparar o aluno para o exercício da cidadania crítica, reflexiva e participativa, em decisões individuais e/ou coletivas. Em síntese, considerando as precárias condições de vida das populações levadas ao cativeiro submetidas à constante violência, reflete-se sobre como a problemática da saúde e da doença foi pensada para os negros e por sua vez, em que nível elas eram tratadas.

Palavras-chave: *Epidemias, Doenças, Escravidão*.

Mortalidade Neonatal no Rio Grande do Sul: fatores associados e principais causas no período de 2006-2013.

Danielle Rodrigues Guidini [1]

Marilyn Agranonik[2]

Resumo: A mortalidade infantil tem destaque como indicador de saúde especialmente por estar relacionada e condições socioeconômicas e sanitárias da população e por refletir as condições de saúde da parcela da população mais vulnerável: os menores de um ano. No RS a taxa de mortalidade infantil caiu nas últimas duas décadas, especialmente devido à redução dos óbitos no período pós-neonatal (óbitos de 28 a 364 dias). Os óbitos neonatais (entre 0 a 27 dias de vida) passaram a representar 67% do total de óbitos infantis ocorridos no RS, tornando-se o principal foco de ações de prevenção. Portanto, este estudo tem por objetivo: (1) descrever a tendência da taxa de mortalidade neonatal (TMIN) por principais causas e (2) analisar os principais fatores de risco relacionados à mortalidade neonatal no Rio Grande do Sul no período de 2006-2013. Metodologia: Foram utilizados dados do Sistema de Informações de Nascidos Vivos (SINASC) e do Sistema de Informações de Mortalidade (SIM) obtidos através do site do DATASUS. Foram analisadas variáveis sociodemográficas maternas - idade materna ($<$ 20, 20-34, $\geqslant$ 35 anos) e escolaridade materna ($<$ 4, 4-7, $\geqslant$ 8 anos) -, de assistência pré e perinatal - tipo de parto (vaginal ou cesáreo) - e do recém-nascido - duração da gestação ($<$ 32, 32-36, $\geqslant$ 37 semanas), peso ao nascer (500-1.499g, 1.500-2.499g, 2.500-3.999g, $\geqslant$ 4.000g) e tipo de gravidez (única ou múltipla). As principais causas de óbito foram selecionadas a partir do CID10 e com base na literatura. Para cada variável, a tendência no período 2000-13 foi medida através teste Qui-Quadrado de tendência. O nível de significância foi de 5%. Resultados: Entre 2006 e 2013 ocorreram 1.094.771 nascimentos e 8.657 óbitos neonatais no RS. A TMIN caiu de 8,8 por 1.000 em 2006 para 7,0 por 1.000 em 2013 ($p < 0,001$). A prematuridade continua apresentando TMIN elevada, embora tenha decrescido significativamente no período estudado, caindo 23% para os recém-nascidos (RNs) com menos de 32 semanas. Com relação ao peso ao nascer, ocorreram quedas acentuadas na TMIN para o grupo nascido com menos de 1499g e para o grupo com peso entre 2500 e 3999g. A TMIN relacionada à idade materna apresentou decréscimo de 40% para aqueles RNs cujas mães tinham menos de 20 anos. A TMIN de RNs cujas mães possuem baixa escolaridade aumentou, passando de 15,5 por 1.000 em 2006 para 19,6 por 1.000 em 2013. Ao mesmo tempo, de RNs com mães de alta escolaridade decresceu, chegando a 5,4 óbitos por 1.000 em 2013. Ocorreu decréscimo da TMIN para os RNs de gestações única e múltipla, partos vaginal e cesáreo. Na análise de causas de óbitos neonatais, prevaleceram as causas formadoras do bloco de afecções

[1]Graduanda em Estatística - UFRGS / Bolsista FAPERGS.

[2]Pesquisadora da FEE / Orientadora.

do período perinatal, representando 75,1% dos óbitos, seguida pelas malformações congênitas (24,3%). Em relação ao primeiro grupo, cinco causas se destacam: (1) fatores maternos relacionados à gravidez que aumentaram de 16% para 21%; (2) afecções respiratórias que decresceram 27,5% entre 2006 e 2013, caindo para 12%; (3) prematuridade que se manteve durante todo período como a quinta colocada entre as causas de óbito, embora tenha caído de 7,6% para 6,3% dos óbitos; (4) as infecções e (5) asfixia/hipóxia que se mantiveram estáveis ao longo do período, em torno de 14% e 5,5%, respectivamente. Conclusões: Os resultados demonstram que no período neonatal, os óbitos estão majoritariamente relacionados a fatores maternos e ao parto. Portanto, é necessária especial atenção à qualidade do atendimento oferecido à mãe a ao recém-nascido e maiores investimentos em tecnologias voltadas ao cuidado perinatal.

Palavras-chave: *Mortalidade Neonatal, Fatores de Risco, Causas de Óbito*.

Explorando os recursos estatísticos presentes no software GeoGebra por meio de construções

Bruno Marques Collares [1]

Jéssica Carolini da Silva Laurindo [2]

Lucas Caitano[3]

Resumo: : Este trabalho tem como objetivo evidenciar as potencialidades do software GeoGebra como um possível recurso computacional para o ensino e para a aprendizagem de conceitos relacionados à Estatística na Educação Básica. O GeoGebra é um software livre, de matemática dinâmica, multiplataforma que vem despertando o interesse de muitos professores por reunir diversos recursos e apresentar diferentes representações de um mesmo objeto que interagem entre si e que podem ser modificadas pelo usuário. Apesar de não ser um software estatístico, desde 2009 (a partir da sua versão 3.2) o GeoGebra passou a disponibilizar diversos recursos relacionados à Estatística. No entanto, o pequeno número de trabalhos acadêmicos voltados à exploração destas ferramentas evidencia o pouco conhecimento da comunidade de usuários frente às possibilidades oferecidas pelo software para o ensino de conceitos estatísticos. Desse modo, são apresentadas três construções que possibilitam explorar alguns recursos oferecidos pelo software relativos à criação de gráficos, tabelas e cálculos estatísticos. Além disso, argumenta-se que a multiplicidade de representações expostas em uma mesma tela, possibilitadas pelo software, pode auxiliar o estudante na compreensão de diversos conceitos referentes à análise e interpretação de dados por meio da experimentação e da manipulação concreta, proporcionando, assim, o desenvolvimento do letramento estatístico.

Palavras-chave: *GeoGebra, Estatística, Educação Básica, Múltiplas Representações*.

[1]UFRGS - PPGEMAT. Email: *collares.bruno@hotmail.com*
[2]UFRGS - PPGEMAT. Email: *jessica.rec@gmail.com*
[3]UFRGS -PPGEMAT. Email: *lucaitano@gmail.com*

Análises Utilizadas nos Ensaios Clínicos Controlados com Medidas antes e depois da Intervenção

Daniela Benzano Bumaguin [1]

Patricia Klarmann Ziegelmann[2]

Introdução: Os trabalhos científicos na área do comportamento utilizam frequentemente um delineamento do tipo ensaio clinico controlado com medidas antes e depois da intervenção. Uma análise adequada dos dados coletados e uma correta comunicação dos resultados são aspectos fundamentais para a compreensão clínica e para a unificação da informação entre os pesquisadores.

Objetivos: O nosso trabalho tem como objetivo fazer uma revisão dos métodos de análise e medidas de efeito utilizadas em trabalhos publicados na área da psiquiatria e do comportamento.

Métodos: Foram selecionados artigos de revistas nacionais e internacionais com diferentes fatores de impacto da área da psiquiatria e levantadas as análises utilizadas assim como os resultados comunicados.

Resultados: Entre as análises mais frequentes figuravam as comparações do depois, a comparação dos deltas e a comparação do depois ajustado pelo valor basal (via ANCOVA). Como medida de efeito é frequentemente utilizado o d de Cohen, nem sempre compreendido e justificado. Muitos artigos apresentam esta medida sem uma real necessidade de comparar escalas entre si, e bastaria a apresentação de uma diferença bruta entre as médias com o respectivo intervalo de confiança.

Conclusão: Sugerimos a análise via ANCOVA com ajuste para os dados no basal e como medida de efeito a diferença bruta entre as medias ajustadas quando o objetivo não é a comparação entre escalas. Em caso de comparação entre escalas o d de Cohen nos parece uma medida conveniente porque coloca as escalas na mesma unidade porém discutível em relação ao seu significado do ponto de vista clínico.

[1]UFRGS - Universidade Federal do Rio Grande do Sul.

[2]UFRGS - Universidade Federal do Rio Grande do Sul. Email: *patricia.ziegelmann@ufrgs.br*