Solicitação de Institucionalização de Projeto de Pesquisa – Sem Financiamento

Título: Cadeia de Markov: Exemplos de Aplicação

1. Dados do Projeto

• Nome da coordenadora: Profa. Dra. Roziane Sobreira dos Santos.

• Cargo na UNIR: Docente.

• Está atuando como docente em algum curso de Pós-Graduação da UNIR? Não.

• Núcleo/Campus de lotação: *Campus* de Ji-Paraná

• Departamento/Unidade de lotação: Departamento de Matemática e Estatística

• Titulação do Coordenador do Projeto de Pesquisa: Doutora

• Área da Titulação: Ciências Agrárias

• Curso da titulação: Doutorada em Meteorologia Agrícola

• Instituição de titulação: Universidade Federal de Viçosa - UFV

• Linhas de Pesquisa em que atua: Estatística Aplicada e Biometria; Estatística e Saúde; Estatística Teórica

• E-mail: roziane@unir.br

• Telefone: (69) 998116-2540

• Link para o Currículo Lattes: http://lattes.cnpq.br/4983021820079917

• Grupo de Pesquisa ao qual o Projeto está vinculado: Grupo de Estudo e Pesquisa em Análise de Registros da Região Amazônica: http://dgp.cnpq.br/dgp/espelhogrupo/251343

• O projeto de pesquisa está vinculado a algum Programa de Pesquisa? Não.

• Título do Projeto de Pesquisa: **Cadeia de Markov: Exemplos de Aplicação**

• Relacione os pesquisadores vinculados ao projeto de pesquisa e a instituição de vínculo:

Dilson Henrique Ramos Evangelista - Universidade Federal do Sul e Sudeste do Pará – UNIFESSPA;

Robson Alves de Oliveira – Universidade Federal de Rondônia – UNIR.

• Relacione os estudantes vinculados ao projeto de pesquisa, a instituição de vínculo, o tipo de curso (graduação, pós-graduação) e o nome do curso.

Maria Aparecida Fritz de Alencar. Universidade Federal de Rondônia - UNIR, Graduação – Bacharelado em Estatística.

• Relacione os técnicos vinculados ao projeto de pesquisa, a instituição de vínculo e o cargo na instituição: Não se aplica.

• Link do Currículo Lattes de todos os membros do Projeto de Pesquisa (atualizado nos últimos seis meses):

Roziane Sobreira dos Santos - http://lattes.cnpq.br/4983021820079917

Dilson H. R. Evanglista - http://lattes.cnpq.br/1851435739271286

Robson Alves de Oliveira - http://lattes.cnpq.br/8666796019262442

Maria Aparecida Fritz de Alencar - http://lattes.cnpq.br/8204265175915812

• O projeto está sendo desenvolvido em parceria com outras instituições? Se a resposta for “sim”, relacione as instituições: Sim. Universidade Federal do Sul e Sudeste do Pará – UNIFESSPA

• Área do Projeto de Pesquisa: Ciências Exatas e da Terra

• Data de início do Projeto: fevereiro – 2023

• Data de término do Projeto: outubro – 2024

• Palavras-Chave: Probabilidade, Ações, R, Precipitação, Processos estocásticos

Resumo

O presente estudo terá como objetivo principal descrever o uso e aplicação de processos estocásticos para elaborar recursos didáticos com abordagem teórica e prática que forneça o detalhamento algébrico, bem como a implementação de um Software de Código Aberto (Open Source) para utilização prática de Cadeia de Markov. Ainda, objetiva desenvolver pesquisas aplicadas com dados advindos das diversas áreas do conhecimento, tais como: ambiental, econômica e de saúde, etc., ilustrando e exemplificando as abordagens estatísticas estudadas.

2. Problema, Justificativa e hipóteses do problema

Na sociedade atual, a análise de dados é um procedimento fundamental, que está cada vez mais presente nos diversos setores, abrangendo todas as áreas do conhecimento. Os avanços tecnólogos tem possibilitado uma volumosa quantidade de dados, gerados a todo instante. Sendo cada vez mais urgente a necessidade de utilizar essa grande quantidade de dados para gerar informações úteis nos processos de compreensão, gerenciamento e tomada de decisão (RAUTENBERG e CARMO, 2019). Analisar estes dados estatisticamente de forma eficiente tem sem tornando cada dia mais importante.

A disponibilidade de sistemas computacionais capazes de processar um grande volume de informações e softwares para análise de dados, permitiram que análises estatísticas sejam abordadas por diversos usuários, muitas vezes sem conhecimento das formulações matemáticas, pressupostos e aplicação das diversas técnicas estatísticas utilizadas. Rautenberg e Carmo (2019) destacam que os profissionais trabalham com dados precisam entender o funcionamento dos métodos e saber interpretar seus resultados, estatisticamente.

Destarte, para realizar uma análise estatística não é suficiente apenas saber utilizar um software, pois o conhecimento e entendimento da fundamentação teórica das metodologias estatísticas e detalhamento da formulação algébrica é essencial para aplicações práticas. Sendo fundamental para compreender os resultados da análise, interpretar e discutir de maneira aprofundada e coesa os dados, para obter informações úteis que irão auxiliar na gestão e na tomada de decisão.

A análise estatística permite organizar os dados e calcular probabilidades das tendências futuras com base nas informações passadas, ajudando a melhorar o desempenho em atividades nas mais variadas áreas, como econômica, engenharia, Meio Ambiente, Medicina, Agricultura, Marketing, entre outras, transformando e modelando dados, e descobrindo informações úteis.

Uma técnica estatística que tem ganhado destaque no processo decisório, por possibilitar modelar a incerteza é a cadeia de Markov, que contribui para o entendimento do comportamento imprevisível do processo ao longo do tempo (ALMIÑANA e SILVA, 2022).

Embora a cadeia Markov sejam uma técnica conhecida e implementada nos softwares, muitos não dominam seus cálculos matemáticos passo a passo, e, por vezes, não possuem um bom entendimento de como a análise é realizada com o software, principalmente os softwares livre e de código aberto, como por exemplo o R.

Neste contexto, este projeto surge da necessidade de elaborar recursos didáticos que forneçam embasamento teórico, com rigoroso detalhamento da formulação algébrica apresentando a fundamentação teórica e aplicação prática de cadeia de Markov, para que profissionais das mais variadas áreas possam ter um material de alta qualidade, de fácil entendimento, mas não deixando de considerar uma abordagem criteriosa para aplicação e análise dos resultados. Bem como, fornecer um tutorial elucidativo e minucioso que apresente o passo a passo da aplicação da técnica no software R.

Principais Resultados e Impactos Esperados

1) Difusão de conhecimento, ampliando a qualidade do ensino articulando-o a pesquisa;
2) Promover o crescimento e a consolidação dos Grupos de Pesquisa: Grupo de Estudo e Pesquisa em Análise de Registros da Região Amazônica-GEPARRA e Grupo de Pesquisa em Águas Superficiais e Subterrâneas - GPEASS;
4) Publicar artigos científicos em revistas especializadas indexadas;
5) Formação de recursos humanos aptos a analisar e processar dados, aplicando a técnica estatística de cadeia de Markov na Região Norte do Brasil.

3. Objetivos

3.1 Objetivo Geral:

Desenvolver pesquisas na área de processos estocásticos, demonstrando a aplicabilidade de cadeias de Markov em estudos com dados de diferentes áreas do conhecimento. Bem como, elaborar recursos didáticos que forneçam embasamento teórico, com rigoroso detalhamento da formulação algébrica apresentando a fundamentação teórica e aplicação prática de cadeia de Markov.

3.2 Objetivos Específicos:

1) Elaborar material didático com textos elucidativos combinados ao uso de ferramentas computacionais;
2) Produzir conhecimento científico com a publicação de artigos em áreas aplicadas empregando Cadeia de Markov;
3) Elaborar um tutorial, de fácil entendimento, do uso de cadeia de Markov no software R.

4. Material e Métodos*

Os dados para aplicação prática serão obtidos em plataformas de dados abertos (público) oriundos de diversas fontes em áreas do conhecimento que podem ser aplicados Cadeia de Markov como ferramenta para obtenção de informações.

Para aplicação prática da cadeia de Markov nas pesquisas, inicialmente, serão estudadas a fundamentação teórica de processos estocásticos e cadeia de Markov, apresentando conceitos matemáticos e estatísticos das metodologias que serão utilizadas. Esse conhecimento é necessário para tornar as pesquisas aplicadas que utilizam cadeia de Markov mais criteriosas.

Na primeira etapa será realizado estudos e pesquisas bibliográfica, em livros teóricos, artigos científicos e materiais acadêmicos a fim de elaborar um material teórico sobre processo estocástico e cadeia de Markov que serão aplicadas, explicitando a formulação matemática, descrevendo-a de forma que apresente:

- O tipo de variável em que a metodologia estatística pode ser aplicada, como deve ser o conjunto de dados inicial;

- Quais os objetivos, quais os pressupostos necessários e quais a limitações para a utilização da metodologia estatística;
- Detalhamento do desenvolvimento algébrico;
- Elaborar um tutorial com aplicação prática da técnica no software livre R. Descrevendo os pacotes e funções no R disponíveis para realização da análise, detalhando os argumentos de entrada e saída da função, juntamente com a interpretação deles.

As pesquisas de aplicação utilizando Cadeia de Markov, baseado nas pesquisas bibliográficas, necessidades da área de estudo na região norte e os dados disponíveis, o próximo passo será a definição dos estados da cadeia para a estimação da matriz de probabilidade de transição de Markov (Tabela 1). Uma matriz quadrada, formada pelas probabilidades de um estado permanecer em sua classe ou migrar para diferentes estados no próximo passo (n+1).

Tabela 1 – Matriz de transição

n/n+1	Estado 1	Estado 2	Estado 3	··
Estado 1	p	p	p	··
Estado 2	p	p	p	··
Estado 3	p	p	p	··
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$
Estado i	p_i	p_i	p_i	··
...	$\vdots$	$\vdots$	$\vdots$	$\ddots$

Fonte: Elaborado pelos autores

Em seguida, será estimada a distribuição inicial de cada estado. Essas probabilidades formam o vetor de probabilidade inicial especificado na equação 4.

Em sequência, a cadeia de Markov será executada por um tempo longo, para obter a distribuição estacionária (Tabela 2), quando esta existir.

Tabela 2 – Probabilidade do estado estacionário

Intervalo	Probabilidade de estado
Estado 1	π_0
Estado 2	π_1
Estado 3	π_2
$\vdots$	$\vdots$

Estado m	π_m

Fonte: Elaborado pelos autores

Com a aplicação da Cadeia de Markov, será possível obter estimação de probabilidade do processo em cada estado n passo no futuro..

Atualmente duas pesquisas aplicadas estão em fase inicial: uma pesquisa para determinar a probabilidades de ocorrência de dias secos e dias chuvosos consecutivos em Ji-Paraná usando cadeia de Markov, com dados extraídos da plataforma Agritempo (Sistema de Monitoramento Agrometeorológico); e uma pesquisa empregando cadeia de Markov para prever e analisar o comportamento do mercado de ações a fim de descrever suas variações, a partir dos dados conhecidos, utilizada uma série temporal da acão do Itaú Unibanco (ITUB4), obtida do site do yahoo Finance.

Este projeto também contempla o uso do software R (R CORE TEAM, 2022) para a organização, análise dos dados e exibição dos resultados.

5. Referencial Teórico Resumido*

5.1 Processo Estocástico

Um processo estocástico $\{X_t(s), t \in T\, e\, s \in S\}$, também chamado de processo aleatório, é definido como uma coleção de variáveis aleatórias que evoluem, de forma imprevisível, ao longo do tempo (KULKARNI, 2017).

Na definição de um processo estocástico $\{X_t(s): t \in T, s \in S\}$;

$X_t(s)$ é o estado do processo no tempo t;

T é o conjunto de parâmetros (ou índices);

S é o espaço amostral comum das variáveis aleatórias, também chamado de espaço de estados do processo.

O processo estocástico $X_t(s)$ assume um estado ou valor de $x_{(t)}$ no tempo t; ou seja, é uma função aleatória de tempo ou sequências aleatórias. Para especificar seu comportamento, precisamos especificar as funções de distribuição de probabilidade para X_t em cada valor possível de t. Isto é, a cada ponto t no conjunto T, observa uma variável aleatória X_t (ROSS, 2014).

- Os processos estocásticos são caracterizados por: pelo seu conjunto de parâmetros T, que pode ser parâmetro discreto se T é finito ou enumerável ou parâmetro contínuo se T é infinito ou não-enumerável; pelo espaço de estados S, que também pode assumir valores discretos ou contínuos e pelas relações de dependência entre as variáveis aleatórias X_t (KULKARNI, 2017).

O objetivo do estudo de processos estocásticos é a modelagem e análise de variações de um evento através da teoria da probabilidade, para prever o comportamento do processo no futuro, dado que determinado comportamento no passado foi observado (JONES e SMITH, 2017).

5.2 Cadeias de Markov

Uma Cadeia de Markov descreve um processo cujo estado muda ao longo do tempo. Este é um processo que se move aleatoriamente dentro de um conjunto S de estados de acordo com algumas probabilidades de transição denotadas por $p_{ij}[i, j \in S]$.

Essas distribuições de probabilidade incorporam um tipo simples de estrutura de dependência, onde a distribuição condicional dos estados futuros do sistema, dada alguma informação sobre os estados passados, depende apenas da informação mais recente (GILLI et al., 2019). Dessa forma, na previsão futura, apena o estado atual importa, e não o caminho pelo qual o processo chegou ao seu estado atual.

Um processo estocástico $\{X_n, n = 0, 1, 2, ...\}$ que assume um número finito ou contável de valores possíveis. X_n = i, diz-se que o processo está no estado i no instante n. Supondo que sempre que o processo estiver no estado i, existe uma probabilidade fixa p_{ij} de que ele vá no estado j. Uma cadeia de Markov possui a propriedade markoviana (ROSS, 2014):

$$P\{X_{n+1}=j|X_n=i, X_{n-1}=i_{n-1}, \ldots, X_1=i_1, X_0=i_0\}=P\{X_{n+1}=j|X_n=i\}=p_{ij} \quad (1)$$

para todo$i_0, \ldots, i_{n-1}, i, j \in S e n \in N_0$,

em que, o próximo estado do processo $\{X_{n+1}\}$ depende apenas do estado presente $\{X_n=i\}$.

O processo estocástico $\{X_n$, n ≥ 0$\}$ com espaço de estados contáveis S usado para descrever um sistema com propriedade Markoviana é chamado de cadeia de Markov de tempo discreto.

Uma cadeia de Markov de tempo discreto $\{X_n, n \geq 0\}$ com espaço de estados contável S é dito homogênea no tempo se:

$$P\{X_{n+1}=j|X_n=i\}=p_{ij} \quad (2)$$

Da equação 2, a cadeia de Markov de tempo discreto homogêneo no tempo está no estado i no tempo n, se move para o estado j no tempo $n+1$ com probabilidade p_{ij}, para todos os valores de n (ROSS, 2014).

5.2.1 Matriz de Probabilidade de Transição e Distribuição Inicial

A probabilidade fazer uma transição do estado i para o estado j, em uma unidade de tempo $\{n+1\}$, é denominada probabilidade de transição (WERNER, 2020).

As transições são especificadas por probabilidades representadas pela matriz P.

$$P=\begin{bmatrix} p_{00} & p_{01} & p_{02} & \cdots \\ p_{10} & p_{11} & p_{02} & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ p_{i0} & p_{i1} & p_{i2} & \cdots \\ & \vdots & & \end{bmatrix} \quad (3)$$

em que,

(i) p_{ij} de probabilidade de transição do estado$i$ para o estado j;

(ii) $p_{ij} \geq 0$ para todo $i, j \in S$;

(iii) $\sum_{j \in S} p_{ij} = 1$para todo $i \in S$.

Considere $a_i = P(X_0 = i)$, a distribuição inicial do processo no estado i.

$$a_i = P(X_0 = i), i \in S \quad (4)$$

em que a é a distribuição inicial da cadeia de Markov.

Uma cadeia de Markov de tempo discreto homogênea no tempo é completamente descrita por sua distribuição inicial a e pela matriz de probabilidade de transição P (ROSS, 2014).

O comportamento de uma matriz de probabilidade de transição P pode ser representada graficamente por um diagrama de transição, que é um grafo direcionado com tantos nós quantos estados no espaço de estados, e um arco direcionado do nó i para o nó j se $p_{ij} \geq 0$. A figura 1 exibe uma representação do diagrama de transição para uma cadeia de Markov de tempo discreto com dois estados.

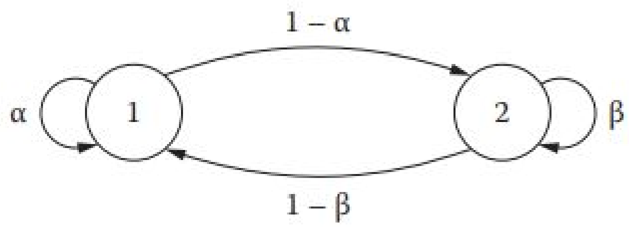

Figura 1- Diagrama de transição de uma cadeia de Markov de dois estados.
Fonte: Kulkarni (2017).

Assim, se a cadeia de Markov de tempo discreto homogênea no tempo está no estado 1, salta para o estado 2 com probabilidade $1-\alpha$; se estiver no estado 2, salta para o estado 1 com probabilidade $1-\beta$, independente de todo o resto.

5.2.2 Distribuição Estacionária

A distribuição estacionária de uma Cadeia de Markov com matriz de transição P é algum vetor, π que satisfaz:

$$\pi P=\pi \quad (5)$$

Ou seja, a distribuição estacionária de uma cadeia de Markov é uma distribuição de probabilidade que permanece inalterada na cadeia de Markov com o passar do tempo (ROSS, 2014).

5.4 Software:

O R é um ambiente de software livre para computação estatística e gráficos (R CORE TEAM, 2022), de código aberto, sendo uma linguagem de programação completa, simples e eficaz. Pode ser instalado em diferentes sistemas operacionais.

Um dos benefícios do uso de R é a possibilidade de inspecionar o código da função para entender/aprender como as análises são realizadas. Ademais, O R permite melhorar ou ampliar seus códigos/funções, pela criação livre de novos pacotes (SALDANHA et al., 2019), além ser versátil, possuindo uma ampla e variada gama de pacotes nas mais variadas áreas do conhecimento, que podem ser combinadas entre si.

Com o R é possível trabalhar os dados desde a sua obtenção até a utilização de métodos estatísticos mais avançados. Ademais, o software é são amplamente utilizados na comunidade científica alguns órgãos governamentais e de iniciativa privada.

6. Beneficiários e/ou usuários da atividade*

Pesquisadores, estudantes e profissionais das diversas áreas do conhecimento, interessados em aprender e/ou ampliar o entendimento da aplicação de cadeia de Markov.

7. Bibliografia e Referências*

ALMIÑANA, CESAR; SILVA, LEANDRO AUGUSTO da. Cadeias de Markov em Estratégia de Negociação de Ações em Alta Frequência. Revista Mackenzie de Engenharia e Computação, São Paulo, v. 22, n. 1, p. 10-35, 2022. https://doi.org/10.5935/RMEC.v22n1p10-35

GILLI, Manfred; MARINGER, Dietmar; SCHUMANN, Enrico. Numerical methods and optimization in finance. Second Edition. Academic Press, 2019. https://doi.org/10.1016/C2017-0-01621-X

JONES, PETER W.; SMITH, PETER. **Stochastic processes an introduction**. Chapman and Hall/CRC, 2017.

KULKARNI, VIDYADHAR G. Modeling and analysis of stochastic systems. Third edition. Boca Raton: CRC Press, 2017.

R CORE TEAM (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Disponível em: < https://www.R-project.org/>. Acesso em: 02 aug. 2022.

RAUTENBERG, SANDRO; CARMO, PAULO RICARDO VIVIURKA do. Big Data e Ciência de Dados: complementariedade conceitual no processo de tomada de decisão. // Brazilian Journal of Information Studies: Research Trends. 13:1 (2019) p.56-p.67. ISSN 1981-1640.https://doi.org/10.36311/1981-1640.2019.v13n1.06.p56

ROSS, SHELDON M. **Introduction to probability models**. Academic press, 2014.

SALDANHA, RAPHAEL DE FREITAS; BASTOS, RONALDO ROCHA; BARCELLOS, CHRISTOVAM. Microdatasus: pacote para download e pré-processamento de microdados do Departamento de Informática do SUS (DATASUS). **Cadernos de Saúde Pública**, v. 35, p. e00032419, 2019. https://doi.org/10.1590/0102-311X00032419

WERNER, CAMILA et al. Variáveis aleatórias e Cadeias de Markov: o problema de completar um álbum de figurinhas e uma proposta de aplicação no Ensino Médio. 2020.

8. Cronograma*

Descrição da Atividade	2023	2024
Elaboração do projeto	X	
Pesquisa bibliográfica	X	X
Estudo Inicial do Tema	X	X
Aprendizado teórico do método estatístico	X	X
Aprendizado Software R	X	X
Obtenção dados e elaboração banco dados	X	X
Tratamento estatístico inicial nos dados	X	X
Aplicação metodologia	X	X
Análise dos resultados	X	X
Análise final e conclusão	X	X
Elaboração de Artigos Científicos e Materiais Didáticos	X	X
Apresentação final	X	X

9. Outras informações relevantes

Os recursos necessários para realização deste projeto serão disponilizados pelo Departamento Acadêmico de Matemática e Estatística do Grupo de Estudo e Pesquisa em Análise de Registros da Região Amazônica da UNIR *Campus* Ji-Paraná, sem nenhum outro custo adicional a instituição.