

MAT02036 - Amostragem 2

Aula 00 - Apresentações

Markus Stein

Departamento de Estatística, IME/UFRGS

2022/2

Apresentações

0 professor esse semestre

- Eu!



via GIPHY


O professor esse semestre

Minha experiência com Estatística

- Graduação em Estatística em **2006** aqui no DEst, título da monografia *Estudos de Correlação Ecológica - Uma aplicação a dados de saúde em Porto Alegre.*;
- Mestrado em Estatística na **UFMG** com conclusão em **2008**: dissertação intitulada *Testes da Razão de Verossimilhanças em Modelos Lineares Mistos.*
- Ingressei na **UFPEL** em **2009** como Professor Assistente no **Departamento de Matemática e Estatística.**
- Em novembro de **2010** voltei para o **DEst**, onde estou como docente desde então.

0 professor nesse semestre

Minha experiência com Estatística

- Em **2013** fui contemplado com bolsa de estudos e cursei o Doutorado em Estatística da **Universidade de Auckland**, Nova Zelândia; porém não concluí a tese, intitulada provisoriamente *Combining aggregate and individual level data in contingency tables*.
- Continuarei minha formação de Doutorado em **2023** no **Programa de Pós-Graduação em Epidemiologia** da UFRGS.
- Meus interesses de pesquisa são: **inferência paramétrica; teoria de verossimilhança e aproximações, melhoramentos de testes de hipóteses, equações de estimação ponderadas e amostragem complexa**.
- Em estatística aplicada: **modelos mistos, dados (correlacionados) de área e/ou longitudinais**. 

E sobre vocês?

Sugestões

- Qual o seu nome?
- Etapa do curso?
- Quantas/quais disciplinas nesse semestre?
- O que é estatística/amostragem/inferência?
- O que espera da disciplina?
- Possui alguma experiência profissional na área?
- Qual sua cidade?

A disciplina

Objetivos

- Estender o conhecimento de técnicas de **obtenção de dados** primários e métodos de **inferência** sobre características populacionais baseados em **amostragens complexas**.
- Estudar outros delineamentos: **Amostragem Estratificada; Amostragem por Conglomerados; e extensões**, vantagens e desvantagens em relação à AAS com e sem Reposição.

A disciplina

Organização

- **Disciplina:** Amostragem 2
- **Turma:** U
- **Modalidade:** Ensino presencial
- **Professor:** Markus Chagas Stein
 - e-mail: markus.stein@ufrgs.br
 - Sala: B120 do IME


A disciplina

Aulas e material didático

- **Aulas** (teóricas e práticas)
 - Exposição e **discussão** dos conteúdos
 - **Leituras semanais de artigos e capítulos de livros**
 - Exemplos
- **Notas de aula**
 - Slides
 - Arquivos de rotinas em R
- **Exercícios**
 - Listas de exercícios

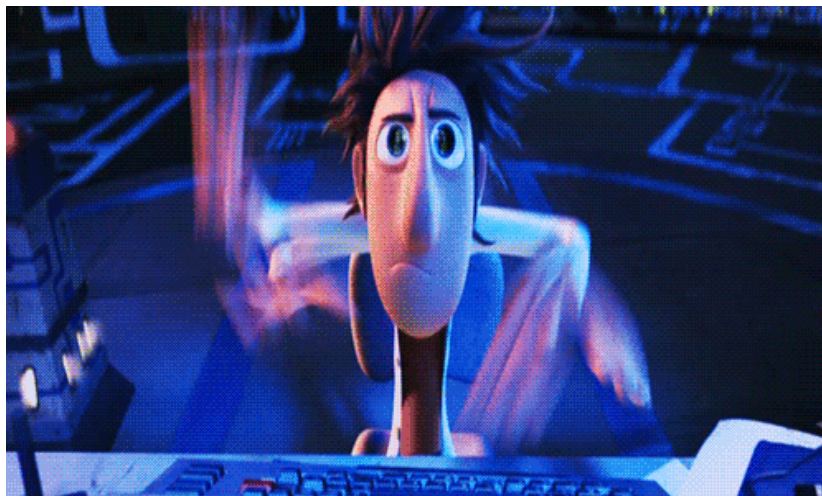
A disciplina

Aulas e material didático


- **Canais de comunicação:**
 - Durante as aulas
 - Moodle: aulas, materiais e **fórum geral**
 - e-mail do professor
- **Aulas:** terças e quintas, das 8hs 30min às 10hs 10min, na Sala F115 do Instituto de Química - Campus do Vale
 - *8hs 30min:* chegada
 - *8hs 40min:* início 
 - *10hs:* fim/dúvidas
 - *10hs 10min:* saída
- **Covid** e recomendações.

A disciplina

Tecnologias e Linguagem



Fonte: ghipy.com




- Exemplos e exercícios com o apoio do computador:
 -  e RStudio

A disciplina

Tecnologias e Linguagem

- GitHub, .Rproj e .Rmd.
 - revisão(?)

Porque usar .

- Aprendemos  para análises estatística, mas  é uma linguagem de programação (geral).
- Ao contrário de linguagens específicas, como SQL para manipulação de bases de dados.
-  foi criado na **Universidade de Auckland** em 1993,
 - continua sendo uma das linguagens mais utilizadas porque sua comunidade cresce e desenvolve milhares de pacotes e produtos.
- CRAN

A disciplina

Conteúdo programático

- **Área 1 - Amostragem Estratificada**
 - Características. Decomposição da variância total. Parâmetros. Composição dos estratos. Critérios de repartição da amostra. Estimadores. Erro padrão dos estimadores. Eficiência das repartições. Dimensionamento de amostras.
- **Área 2 - Amostragem por conglomerados**
 - Características, as variâncias dentro e entre conglomerados, comparação com estratos. Amostragem por conglomerados com etapa única, em duas etapas com sub-amostras de tamanhos iguais e tamanhos variáveis. Estimadores e tamanhos de amostras.
- **Área 3**
 - Amostragem sistemática, seleção de amostras
 - Amostragem com probabilidades variáveis
 - Planos amostrais complexos

A disciplina

Avaliação

- Serão realizadas quatro atividades de avaliação (pelo menos uma em de cada área):
 - três provas (P_1 , P_2 e P_3) presenciais e individuais;
 - um trabalho em grupo (T_1).
- Cada atividade de avaliação vale 10 pontos
- Será realizado uma prova presencial e individual como atividade de recuperação (PR)
 - Para os alunos que não atingirem o conceito mínimo
 - **Esta prova abrange todo o conteúdo da disciplina**

A disciplina

Avaliação !

$$MF = \frac{P_1 + P_2 + P_3 + T_1}{4}$$

- **A:** $9 \leq MF \leq 10$
- **B:** $7,5 \leq MF < 9$
- **C:** $6 \leq MF < 7,5$
- **D:** $MF < 6$
- **FF:** se o aluno tiver frequência inferior a 75% da carga horária prevista no plano da disciplina

A disciplina

Avaliação

- Se $MF < 6$ e frequência mínima de 75% o aluno poderá realizar a prova de recuperação e neste caso

$$MF' = MF \times 0,4 + PR \times 0,6$$

- **C:** $MF' \geq 6$
- **D:** $MF' < 6$

A disciplina

Datas das avaliações

- *P1*: 22/12/2022
- *P2*: 02/03/2023
- *P3*: 11/04/2023
- *T1*: a definir
- *PR*: 18/04/2023

Tempo (esperado) de dedicação

- aulas presenciais - **4h semanais**
- listas de exercícios, laboratórios e revisão - **6h semanais**

A disciplina

Referências bibliográficas

- Principais
 - Bolfarine, H. e Bussab, W. O. **Elementos de amostragem**, Blucher, 2005. (*sabi+*)
 - Cochran, W. G. **Técnicas de Amostragem**, Fundo de Cultura, 1965.
- Complementares
 - Leotti, V. B. **Amostragem 2**, Notas de aula, 2019.
 - Silva, P. L. N. e Dias, A. J. R. **Amostragem: teoria e prática usando o R**, 2021. (*github*)
 - Djalma Pessoa e Pedro Nascimento Silva, **Análise de Dados Amostrais Complexos** (*github*)
 - Lumley, T. **Complex Surveys: a guide to analysis using R**, 2010. (<https://r-survey.r-forge.r-project.org/svybook/>)

Dúvidas, sugestões, críticas, ...? ?

O que é amostragem?

De modo geral

- teoria de **amostragem** = **planejamento** + **obtenção** + **análise** de dados amostrais obtidos de populações finitas.

Em amostragem 2 estudamos diferentes formas de obtenção de dados e (des)vantagem em relação à AAS.

- **Model versus design approach**
 - verossimilhança \times equação de estimação ponderada?
 - v.a.: $Y_i \sim f(y, \theta)$ ou $R_i \in \{0, 1\}(?)$.

O que é Amostragem 2?

- Análise de dados de amostras dependentes (possivelmente) com probabilidades desiguais de seleção.
- Na prática pesquisas utilizam a combinação de
 - AAS (ou sistemática)
 - estratos
 - *clusters*
- Aspectos computacionais
 - tradicionalmente livros de amostragem focam nas diferentes configurações
 - computacionalmente é mais eficiente implementar o caso geral (de equações de estimação ponderadas)

Na prática

- Exemplos de Pesquisas no Brasil
 - PNS do IBGE -
 - PED do DEE? FEEE? - <https://arquivofee.rs.gov.br/publicacoes/ped-rmpa/>
 - POF -
 - PNAD -

Institutos/Organizações

- No Brasil
 - ENCE/IBGE -
 - DEE/RS - FEEE???
- Exterior
 - StatsNZ - Nova Zelândia
 - Statscan - Canadá
- IASS/ISI - <http://isi-iass.org/home/>

Na prática

- *Epicovid*
 - EPIDEMIOLOGIA DA COVID-19 NO RIO GRANDE DO SUL: Estudo de base populacional - https://wp.ufpel.edu.br/covid19/files/2020/09/Coletiva-RS-Fase-8-20200908_v2.pdf
 - *Estimando a prevalência de covid19 no RS* - https://github.com/markus-stein/MCML-notes/blob/master/covid-prevalence_RS/prevalencia_covid19_RS2.pdf
- Artigo caso-controle e Covid no RS
 - *Social Distancing, Mask Use and the Transmission of SARS-CoV-2: A Population-Based Case-Control Study* - https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3731445
 - https://github.com/markus-stein/MCML-notes/blob/master/case-control_mask-use_and-covid/case-control_mask-use-and-covid.txt
- Tenho também alguma experiência em análise de dados caso-controle em epidemiologia veterinária.

Viéses em amostragem

Definição de viés:

- Dicionário Priberam - <https://dicionario.priberam.org/vi%C3%A9s>
- Cambridge dictionary - <https://dictionary.cambridge.org/dictionary/english/bias>



Fonte: <https://www.adweek.com/agencies/6-ways-you-might-be-unintentionally-introducing-bias-into-your-marketing-surveys/>

Viéses em pesquisas

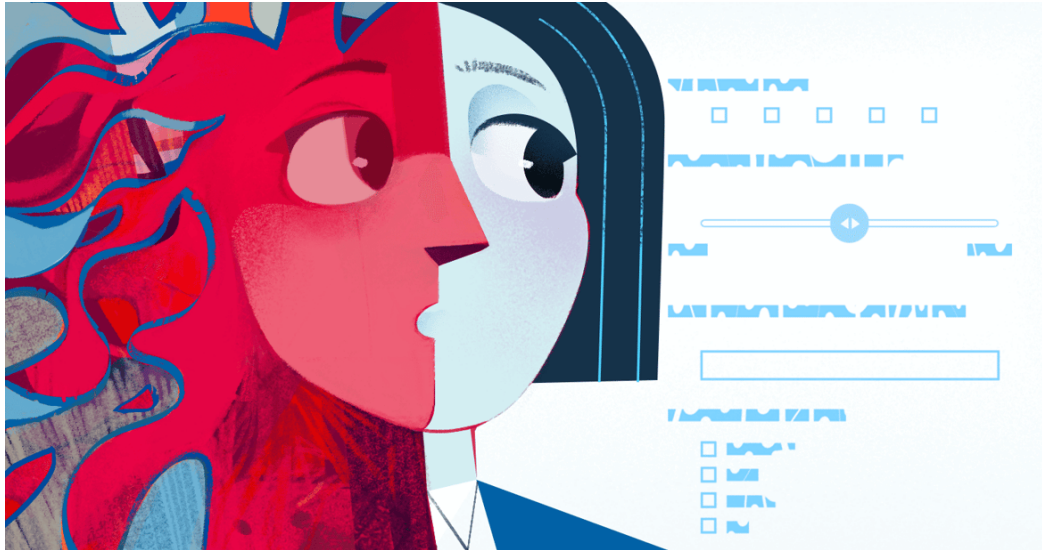
Nonresponse - Interviewer bias



Fonte: <https://blog.cruxresearch.com/2013/08/27/the-top-5-errors-and-biases-in-survey-research/>

Viéses em pesquisas

Response bias



Fonte: <https://www.zef.fi/blog/response-bias-its-cramping-your-survey-style>

Viéses em pesquisas

Questionnaire bias



Fonte: <https://news.nnlm.gov/nec/2017/03/06/the-dark-side-of-questionnaires-how-to-identify-questionnaire-bias/> -->

Por fim

Coisas que acredito...

- (duas) disciplinas que diferenciam nosso currículo (da Estatística em relação a outras graduações em **ciência de dados**) são o estudo de planejamento e análise de
 - experimentos
 - (amostragens complexas)
- porque?
 - estimular o pensamento crítico sobre a obtenção dos dados
 - desenvolver análises e conclusões rigorosamente seguindo a teoria
 - flexibilidade para trabalhar em novos delineamentos/amostragens

Próxima aula

Revisão de conceitos básicos da teoria de:

- probabilidade,
 - definições,
 - variáveis aleatórias;
- inferência,
 - estimação (pontual e intervalar),
 - teste de hipóteses;
- **amostragem (1) aleatória simples (AAS),**
 - parâmetros e estimadores,
 - AAS com e sem reposição
 - estimação dos parâmetros e das variâncias dos estimadores, viés,
 - resultados assintóticos, tcl,
 - ICs e THs.

Sejam bem-vinda(o)s!



Fonte: imagem do livro *Combined Survey Sampling Inference: Weighing of Basu's Elephants: Weighing Basu's Elephants*.