

**MI427 – Análise de dados hierárquicos (acompanha ME 913, disciplina de
graduação, homônima)
Segundo Semestre de 2020
Programa**

Professor: Caio L. N. Azevedo

Sala do professor: As aulas serão remotas (veja abaixo)

e-mail: cnaber@ime.unicamp.br

Página do curso: link

Sistemática

1. Aulas

- (a) O professor gravará as apresentações da matéria (via Google Meet), de forma antecipada, disponibilizando-as regularmente (Google Drive). Os respectivos links de acesso as gravações serão disponibilizados no moodle.
- (b) Durante os dias e horários das aulas (Terças e Quintas das 16h00 às 18h00), via Google Meet, serão tiradas dúvidas a respeito da matéria (as quais foram apresentadas nas gravações descritas no item 1.(a)). Os links de acesso (convite) serão disponibilizados no moodle (em princípio, será o mesmo link para todas as aulas). Estas aulas também serão gravadas e a disponibilização delas será feita conforme descrito no item 1(a)

2. Tira-dúvidas

- (a) Professor: Durante os dias e horários das aulas (Terças e Quintas das 16h00 às 18h00) (não haverá monitoria: PAD/PED nem atendimento em outro dia/horário)

3. Comunicação

- (a) Aluno-Professor: 1. Via e-mail do professor (veja acima, não enviar e-mail via moodle). 2. Durante as aulas (item 1.2). Deverá ser enviado e-mail ao Professor, somente em caso de correções necessárias nos slides/programas em R/site do curso, problemas com links de acesso/arquivos, solicitação de Prova Substitutiva e Exame (veja mais adiante), bem como para tirar dúvidas sobre o conteúdo da disciplina e regras (veja item 2, ou seja, procure tirar suas dúvidas durante as aulas). Utilize seu e-mail acadêmico. O título do e-mail deve conter o assunto seguido de (MI427/ME913) e o corpo do e-mail deve conter (ao final) o nome completo e RA do aluno.

- (b) Professor-aluno: moodle (para disponibilização dos links das aulas/gravações e para enviar e-mail para todos os alunos, quando necessário) e página do curso (slides, programas em R, referências, informações sobre o trabalho etc).
- (c) Recomenda-se visitar a página do curso/moodle, diariamente.

Ementa

Introdução aos dados hierárquicos. Análise exploratória de dados hierárquicos. Conceitos básicos, terminologia e notações. Modelos lineares hierárquicos gaussianos de um, dois e vários níveis. Modelos lineares hierárquicos não gaussianos de um, dois e vários níveis. Estimacão, análise de resíduos, comparação de modelos e testes de hipótese sob a ótica frequentista. Inferência Bayesiana. Modelos não lineares. O suporte computacional será desenvolvido/apresentado na plataforma R. Naturalmente, o conteúdo referente aos pré-requisitos para esta disciplina, será necessário para o seu desenvolvimento. Buscar-se-á: 1. O desenvolvimento do aluno na compreensão/desenvolvimento Estatístico de metodologias e das necessárias ferramentas matemáticas, 2. O desenvolvimento de habilidades computacionais, 3. E o aprendizado de modelagem e resolução de problemas com dados hierárquicos.

Conteúdo

1. Introdução aos dados hierárquicos: exemplos
2. Análise exploratória de dados hierárquicos
3. Revisão de modelos de regressão normais lineares homocedásticos
4. Modelos lineares gaussianos hierárquicos com dois níveis
 - (a) Introdução
 - (b) Estimacão
 - (c) Testes de hipótese
 - (d) Verificacão da qualidade de ajuste do modelo
5. Modelos lineares gaussianos hierárquicos com três níveis
 - (a) Introdução
 - (b) Estimacão
 - (c) Testes de hipótese
 - (d) Verificacão da qualidade de ajuste do modelo

6. Modelos lineares não gaussianos hierárquicos para dados positivos assimétricos com dois, três e vários níveis
 - (a) Introdução
 - (b) Estimação
 - (c) Testes de hipótese
 - (d) Verificação da qualidade de ajuste do modelo
7. Modelos lineares não gaussianos hierárquicos para dados binários com dois, três e vários níveis
 - (a) Introdução
 - (b) Estimação
 - (c) Testes de hipótese
 - (d) Verificação da qualidade de ajuste do modelo
8. Modelos lineares não gaussianos hierárquicos para dados de contagem com dois, três e vários níveis
 - (a) Introdução
 - (b) Estimação
 - (c) Testes de hipótese
 - (d) Verificação da qualidade de ajuste do modelo
9. Outros tópicos
 - (a) Inferência Bayesiana
 - (b) Modelos não lineares

Bibliografia (em ordem alfabética), [eventualmente, outras referências serão indicadas, ao longo do curso]

- Azevedo, C. L. N. (2020). Notas de aula disponíveis no site do curso (e outras referências a serem disponibilizadas/divulgadas no próprio site).
- Clark, J. S., Gelfand, A. E. (2006) Hierarchical Modelling for the Environmental Sciences, Oxford University Press.
- Congdon, P. D. (2019) Bayesian Hierarchical Models: With Applications Using R, second edition, Chapman & Hall/CRC.

- de Leeuw, J., Meijer, E. (2007). Handbook of Multilevel Analysis, Springer-Verlag, New York, NY (disponível na biblioteca do IMECC).
- Finch, W. H., Bolin, J. E., Kelley, K. (2019) Multilevel Modeling Using R, second edition, Chapman & Hall/CRC, Statistics in the Social and Behavioral Sciences.
- Galecki, A.; Burzykowski, T. (2013). Linear Mixed-Effects Models Using R: A Step-by-Step Approach, Springer-Verlag, New York, NY (disponível na biblioteca do IMECC).
- Gelman, A., Hill, J. (2006) Data Analysis Using Regression and Multilevel/Hierarchical Models, Cambridge University Press (disponível na biblioteca do IMECC).
- Goldstein, H. (2011) Multilevel Statistical Models, fourth edition, Wiley Series (disponível na biblioteca do IMECC).
- Hox, J. J., Moerbeek, M., van de Schoot, R. (2017) Multilevel Analysis: Techniques and Applications, Third Edition, Quantitative Methodology Series, Routledge (disponível na biblioteca do IMECC).
- Hox, J. J., Roberts. J. K. (2011) Handbook of Advanced Multilevel Analysis, Routledge.
- Lee, Y., Ronnegard, L., Noah, M. (2017) Data Analysis Using Hierarchical Generalized Linear Models with R, Chapman and Hall/CRC.
- Little, T. D., Shcnabel, K. U., Baumert, J. (2000) Modeling Longitudinal and Multilevel Data: Practical Issues, Applied Approaches, and Specific Examples, Psychology Press.
- Migon, H. S., Souza, A. D. P., Schmidt, A. M. (2008). Modelos hierárquicos e aplicações. Associação Brasileira de Estatística, 18º SINAPE.
- Natis, L. (2008). Superdispersão em dados binomiais hierárquicos. Tese de Doutorado. IME-USP (<https://teses.usp.br/teses/disponiveis/45/45133/tde-19062008-132744/es.php>).
- Natis, L. (2001). Modelos Lineares Hierárquicos, Estudos em Avaliação, 23, pp. 3-29. (<http://publicacoes.fcc.org.br/ojs/index.php/eae/article/view/2206/2162>).
- Natis, L. (2000). Modelos lineares hierárquicos. Dissertação de Mestrado. IME-USP.
- O'Connell, A. A., McCoah, D. B. (2008) Multilevel Modeling of Educational Data, Information Age Publishing.
- Pinheiro, J., Bates, D. (2009). Mixed-Effects Models in S and S-PLUS, Springer-Verlag, New York, NY (disponível na biblioteca do IMECC).

- Raudenbush, S. W., Bryk, A. S. (2001) Hierarchical Linear Models: Applications and Data Analysis Methods, second edition, (Advanced Quantitative Techniques in the Social Sciences), SAGE Publications (disponível na biblioteca do IMECC).
- Scott, M. A., Simonoff, J. S., Marx, B. D. (2014) The SAGE Handbook of Multilevel Modeling, SAGE Publications.
- Snijders, T., Bosker, R. (2011) Multilevel Analysis: An Introduction to Basic and Advanced Multilevel Modeling, Sage publications (disponível na biblioteca do IMECC).

Para acessar alguns dos livros digitais (alguns deles estão disponíveis nesse formato) de fora da Unicamp, você poderá precisar do VPN. Veja instruções de instalação aqui. Veja também: Tutorial da Plataforma Minha Biblioteca - versão para alunos: [link](#)

Critérios de avaliação

1. Metodologia de avaliação: duas provas, um trabalho, um seminário e um exame (este último, se necessário).
 - (a) Seminários: apresentação de um artigo que verse sobre modelos e/ou metodologias que não foram e/ou não serão apresentadas em sala de aula, mas que estejam relacionados(as) ao conteúdo da disciplina.
 - i. Os seminários devem ser desenvolvidos e apresentados individualmente ou em dupla.
 - ii. Cada aluno ou dupla deve enviar (por e-mail) um artigo como sugestão para sua apresentação. É vedada a apresentação de artigos em que um dos autores seja aluno desta disciplina. O artigo pode estar apenas “aceito” (sem ter sido publicado ainda), contanto que ele esteja disponível no site da respectiva revista científica. O Professor julgará a adequabilidade do artigo sugerido e, caso ele não seja aceito, o(a) aluno(dupla) deverá enviar outro. Naturalmente, o mesmo artigo não pode ser apresentado por mais de um aluno/dupla. Será disponibilizado no site do curso uma lista com os artigos a serem apresentados, com o(s) respectivo(s) apresentador(es), a medida que os artigos forem sendo aceitos para a apresentação. As sugestões de artigos devem ser enviadas, ao supracitado e-mail, em tempo hábil para que os artigos sejam escolhidos, até o dia 17/12/2020. Ou seja, de modo a haver tempo para que o professor avalie e aceite ou não a sugestão.
 - iii. Regras adicionais, relativas aos seminários serão divulgadas posteriormente.
 - (b) Listas de exercícios serão disponibilizadas periodicamente. Recomenda-se resolvê-las na íntegra. Delas serão selecionados exercícios para serem entregues, valendo nota,

em forma de trabalho. O trabalho poderá ser resolvido e entregue em grupos de um a dois alunos. Ele poderá ser manuscrito (salvo quando se tratar de questões relativas à análise de dados) ou digitado (editor de textos à escolha). O trabalho deve ser enviado por e-mail (por um dos membros da equipe, quando for o caso, com cópia para os demais), até a data limite de entrega. O formato é livre, mas as resoluções das questões têm de ser legíveis e organizadas, e o trabalho deve ser entregue com capa indicando os nomes e os RA's dos integrantes das equipes. O(s) pacote(s) computacional(ais) a ser(em) usado(s) nas análises é(são) de livre-escolha. A data de entrega será definida com, no máximo, 10 dias de antecedência. Portanto, é de suma importância que os alunos resolvam as listas com antecedência (além claro, do óbvio ganho no aprendizado que terão).

- (c) Provas: As questões das provas serão disponibilizadas (no moodle) em torno das 16h00, no site do curso, nos dias indicados (ver Cronograma). As provas deverão ser resolvidas individualmente e enviadas, por e-mail, dentro de um prazo de 48 horas, contados a partir da disponibilização da prova. Os aspectos relativos a resolução, formatação e envio são aqueles descritos acima, para o Trabalho.
- (d) O professor não é, direta ou indiretamente, responsável por questões relativas ao funcionamento dos sistemas computacionais, de e-mail, de impressão etc. Favor dirigir-se aos respectivos responsáveis.

2. Média Global (MG):

- (a) Se o aluno apresentar seminário e entregar o trabalho completo: $MG = 0,55 \times MP + 0,25 \times NT + 0,20 \times NS$, em que $MP = 0,5 \times NP_1 + 0,5 \times NP_2$, NP_i : é a nota da i -ésima prova ($i=1, 2$), NT: nota do trabalho, NS : nota do seminário.
- (b) Se o aluno não apresentar seminário e/ou não entregar o trabalho completo $MG = 0,5 \times (0,55 \times MP + 0,25 \times NT + 0,20 \times NS)$.
- (c) Conceito Final
 - i. Alunos de Pós-Graduação: A se $MG \in [8,5; 10,0]$, B se $MG \in [7,0; 8,4]$, C se $MG \in [5,0; 6,9]$, D se $MG \in [0,0; 4,9]$.
 - ii. Alunos de Graduação: será a própria nota (MG)
 - iii. Para TODOS os alunos, se $MG \geq 6,0$, estará aprovado, se $MG < 2,5$, estará reprovado se $2,5 \leq MG < 6,0$, terá de fazer EXAME.
 - iv. OBS: o aluno aprovado poderá fazer o EXAME, para melhorar sua nota. Entretanto, ele deverá comunicar sua decisão até uma semana (sete dias) antes, por e-mail.
 - v. Média Final (MF) (em caso de Exame): $MF = 0,5 \times MG + 0,5 \times NE$; NE: nota do exame. Se $MF \geq 5,0$, o aluno estará aprovado, caso contrário, estará

reprovado. Caso o aluno tenha sido aprovado com a MG e tenha feito exame, sua média final será $MF = \max(MG, MF)$

(d) Não será cobrada frequência. Contudo, é atitude de necessário respeito a participação nos seminários dos demais colegas/duplas, salvos em caso de pertinente impossibilidade.

(e) Recomenda-se a leitura (relação com PROVAS SUBSTITUTIVAS): link

3. Não será cobrada frequência.

Suporte médico/psicológico:

1. CECOM: Centro de Saúde da Comunidade link

2. SAPPE: Serviço de Assistência Psicológica e Psiquiátrica ao Estudante link 1, link 2