

ME731 – Métodos em Análise Multivariada
Segundo Semestre de 2021
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) Dias e horários: Segundas e Quartas, das 16h00 as 18h00.
- (b) Os conteúdos da matéria serão ministrados e gravados durante os dias e horários das aulas, via Google Meet. As respectivas gravações serão regular e posteriormente disponibilizadas no Google Drive. Os respectivos links de acesso as gravações serão disponibilizados via moodle.
- (c) O(s) link(s) de acesso as aulas (convite) é (serão) enviado(s) no moodle (em princípio, será o mesmo link para todas as aulas).

2. Tira-dúvidas

- (a) Professor: Durante os dias e horários das aulas (Segundas e Quartas das 16h00 às 18h00), poderão ser dirimidas dúvidas, consoante o andamento das aulas.
- (b) Também poderá ser solicitado atendimento (das 13h30 as 14h00, às Segundas e Quartas), conforme descrito abaixo (o link da respectiva sala virtual é o mesmo das aulas regulares, veja item 1 (c)).
 - No dia anterior, enviar um e-mail (até as 18h00, nos moldes apresentados neste Programa, veja abaixo) solicitando agendamento. Por exemplo, se houver interesse em agendar atendimento para o dia 16/08, deve-se enviar e-mail no dia 15/08, até as 18h00.

3. Comunicação

- (a) Aluno-Professor:
- Via e-mail do professor (informado acima, não enviar e-mail via moodle).
 - Durante as aulas.
 - Deverá ser enviado e-mail ao Professor, somente: em caso de correções necessárias nos slides/programas em R/site do curso, para informar problemas com links de acesso/arquivos, para 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/atendimento) e solicitar atendimento. Utilize seu e-mail acadêmico. O título do e-mail deve conter o assunto seguido de (ME731) 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

Resumo: Apresentar e discutir as principais ideias relativas à análise multivariada de dados, através de algumas das principais metodologias. As metodologias apresentadas são: Distribuição normal multivariada; Análise de variância multivariada; Análise de componentes principais; Análise fatorial; Análise de correspondência; Análise de correlações canônicas; Análise discriminante. Muitas dessas técnicas baseiam-se na suposição de normalidade multivariada dos dados. No entanto, serão discutidas, muito brevemente, algumas extensões dessas metodologias a fim de relaxar tal suposição. 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 multivariados.

Conteúdo

1. Distribuição normal multivariada
 - (a) Caso bivariado.
 - (b) Caso multivariado, densidade e função de distribuição acumulada.
 - (c) Função geradora de momentos e função característica.
 - (d) Distribuições marginais e condicionais.
 - (e) Estimação e testes de hipótese
 - (f) Matriz de dados e verificação da normalidade multivariada dos dados.
2. Análise de variância multivariada.
 - (a) Modelo linear geral.
 - (b) Estimação.
 - (c) Testes de Hipótese.
 - (d) Verificação da qualidade de ajuste do modelo.
 - (e) Aplicações.
3. Análise de componentes principais.
 - (a) Introdução.
 - (b) Obtenção das componentes principais.
 - (c) Determinação do número das componentes e sua posterior utilização/interpretação.
 - (d) Aplicações.
4. Análise fatorial.
 - (a) Modelo de análise fatorial.
 - (b) Estimação das cargas fatoriais.
 - (c) Determinação do número de fatores.
 - (d) Predição dos fatores.
 - (e) Verificação da qualidade de ajuste do modelo.
 - (f) Comparação com a metodologia de componentes principais.
 - (g) Aplicações.

5. Análise de correspondência.
 - (a) Definição e paralelo com a metodologia de análise fatorial.
 - (b) Obtenção dos fatores e determinação da quantidade de fatores a serem utilizados.
 - (c) Gráficos de análise.
 - (d) Aplicações.
6. Análise de correlações canônicas.
 - (a) Definição e interpretação.
 - (b) Equações canônicas.
 - (c) Interpretações das equações canônicas.
 - (d) Aplicações.
7. Análise discriminante.
 - (a) Definição.
 - (b) Métodos de análise discriminante.
 - (c) Verificação da qualidade de classificação.
 - (d) Aplicações.

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

- Affi, A, May, S., Donatello, R., Clark, V. A. (2019). Practical Multivariate Analysis, fifth edition, Chapman and Hall/CRC. (disponível na biblioteca do IMECC)
- Anderson, T. W. (2003). An Introduction to Multivariate Statistical Analysis, third edition, Wiley. (disponível na biblioteca do IMECC)
- Azevedo, C. L. N. (2021). Notas de aula disponíveis no site do curso (e outras referências a serem disponibilizadas/divulgadas no próprio site).
- Brian, E., Hottorn, T. (2011). An introduction to applied multivariate analysis with R, 2011th edtion, New York, NY : Springer. (disponível na biblioteca do IMECC)
- Carroll, J. D., Green, P. E., Chatuverdi, A. (1997). Mathematical Tools for Applied Multivariate Analysis, first edition, Academic. (disponível na biblioteca do IMECC)
- Denis, D. J. (2020) Univariate, Bivariate, and Multivariate Statistics Using R: Quantitative Tools for Data Analysis and Data Science, first edition, Wiley Series.

- Everitt, B. S. (2009) *Multivariable Modeling and Multivariate Analysis for the Behavioral Sciences*, first edition, CRC Press. (disponível na biblioteca do IMECC)
- Everitt, B. S., Dunn, G. (2010) *Applied Multivariate Data Analysis*, second edition, Wiley. (disponível na biblioteca do IMECC)
- Everitt, B. S., Hothorn, T. (2011) *An Introduction to Applied Multivariate Analysis with R*, Springer. (disponível na biblioteca do IMECC)
- Härdle, W. K., Simar, L. (2019) *Applied Multivariate Statistical Analysis*, fifth edition, Springer, NY. (disponível na biblioteca do IMECC)
- Harville, D. A. (2008). *Matrix Algebra From a Statistician's Perspective*, second edition, Springer, NY. (disponível na biblioteca do IMECC)
- Housson, F., Lê, S., Pagês, J. (2017) *Exploratory Multivariate Analysis by Example Using R*, second edition, Chapman and Hall/CRC.
- Johnson, R. A., Wichern, D. W. (2018). *Applied Multivariate Statistical Analysis*, sixth edition, Upper Saddle River, NJ : Prentice-Hall. (disponível na biblioteca do IMECC)
- Mardia, K. V., Kent, J. T. and Bibby, J. M. (1980). *Multivariate Analysis*, first edition, London : Academic Press. (disponível na biblioteca do IMECC)
- Rencher, A. C. and Christensen, W. F. (2012). *Methods of Multivariate Analysis*, third edition, John Wiley & Sons, New Jersey, NJ. (disponível na biblioteca do IMECC)
- Searle, S. R., Khuri, A. I. (2017). *Matrix Algebra Useful for Statistics*, second edition, Wiley. (disponível na biblioteca do IMECC)
- Schumacker, R. E. (2015). *Using R With Multivariate Statistics*, first edition, SAGE.
- Vehkalahti, K., Everitt, B. S. (2018) *Multivariate Analysis for the Behavioral Sciences*, second edition, CRC Press.
- Zelterman, D. (2015). *Applied Multivariate Statistics with R*, first edition, Springer. (disponível na biblioteca do IMECC)

Materiais e referências adicionais serão, eventualmente, disponibilizados no site, inclusive sobre temas específicos.

Para acessar os livros em formato digital (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 e um exame (este último, se necessário).

- (a) 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 cujo número (mínimo e máximo) de integrantes será posteriormente informante. Ele poderá ser manuscrito (salvo quando se tratar de questões relativas à análise de dados) ou digitado (editor de textos de livre escolha). O trabalho deverá ser entregue via moodle, por somente um dos membros da equipe, quando for o caso, até a data limite de entrega. Maiores instruções sobre o modo de entrega serão divulgadas posteriormente. As resoluções (digitalização) 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).
- (b) Provas: As questões das provas serão disponibilizadas (no moodle) em torno das 16h00, no sistema moodle, nos dias indicados (ver Cronograma). As provas deverão ser resolvidas individualmente e enviadas, por e-mail, dentro de um prazo de 24 horas, contados a partir da disponibilização da prova. Os aspectos relativos a resolução, formatação e envio são aqueles descritos acima, para o Trabalho.
- (c) 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 entregar o trabalho completo: $MG = 0,70 \times MP + 0,30 \times NT$, 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.
- (b) Se o aluno não entregar o trabalho completo $MG = 0,5 \times (0,70 \times MP + 0,30 \times NT)$.
- (c) Conceito Final
 - i. Se $MG \geq 6,0$, estará aprovado, se $MG < 2,5$, estará automaticamente reprovado e se $2,5 \leq MG < 6,0$, terá de fazer EXAME.
 - ii. 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.

- iii. 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 de responsabilidade do aluno o acompanhamento do curso.
- (e) Recomenda-se a leitura (relação com PROVAS SUBSTITUTIVAS): [link](#)
- (f) Excepcionalmente, o EXAME poderá servir como avaliação substitutiva para pelo menos uma das Provas (I e II). Caso o aluno não tenha feito pelo menos uma delas, entrar em contato com o Professor, via e-mail, o quanto antes.

Suporte médico/psicológico:

1. CECOM: Centro de Saúde da Comunidade [link](#)
2. SAPPE: Serviço de Assistência Psicológica e Psiquiátrica ao Estudante [link](#).