

ME613 – Análise de regressão

Primeiro Semestre de 2019

Programa

Professor: Caio L. N. Azevedo

Sala: 210 IMECC

e-mail: cnaber@ime.unicamp.br

- Preferencialmente, procurar o Professor na supracitada sala, dentro do horário de atendimento. Não serão dirimidas dúvidas via e-mail.
- O ensino aberto (moodle) será usado somente para enviar e-mail (avisos) aos alunos. Caso o(a) aluno(a) queira se comunicar via e-mail, faça-o através do supracitado endereço (não enviar e-mails, posts etc, através do ensino aberto (moodle)).
- O e-mail deverá ser utilizado somente para: solicitação de agendamento de atendimento, justificativa de ausência em um atendimento agendado, justificativa de ausência em aula e o envio de eventuais correções relativas ao site do curso, incluindo os materiais disponibilizados e, eventualmente, para comunicações específicas, como aquela destacada abaixo sobre a prova substitutiva. Dúvidas serão sanadas somente durante o atendimento e durante as aulas. Colocar nome e RA no corpo do e-mail (começo), bem como “ME 613” no título do e-mail. Mensagens fora destes padrões serão ignorados.

Página na internet do curso: http://www.ime.unicamp.br/~cnaber/Material_ME613_1S_2019.html

Aulas: Segundas - sala CB14 e Quartas - sala CB12, das 16h00-18h00.

Atendimento :

- (Professor) Segundas-feiras, das 13h às 14h, sala 210 IMECC.
- O(a) aluno(a) deverá enviar um e-mail (para o supracitado endereço) em determinado dia, solicitando o agendamento de atendimento para o dia posterior. Por exemplo, se ele(a) quiser atendimento para o dia 25/03, deverá enviar um e-mail no dia 24/03 até as 18h00. O(a) aluno(a) que não comparecer a um atendimento agendado e não justificar devidamente (por e-mail) o motivo de sua ausência, não poderá mais solicitar agendamento de atendimento.

Monitoria: Não haverá

1. Programa

Resumo:

Apresentar os principais aspectos dos modelos de regressão lineares, com ênfase no uso da distribuição normal para modelar a variável resposta. Serão discutidos, dentro do contexto/limitações da disciplina, aspectos relativos a resolução de problemas (científicos/consultoria), análise de dados, e como reportar resultados de análise estatística. Diversos exemplos relativos ao preditor linear serão discutidos. Propriedades dos estimadores, intervalos de confiança, testes de hipóteses e avaliação da qualidade do ajuste de modelos serão apresentados. Serão discutidos e apresentados os principais resultados em termos do comportamento dos estimadores, intervalos de confiança e estatísticas do teste. Discutiremos as definições e propriedades das distribuições: normal multivariada, t de Student, F de Snedcor e Qui-quadrado, estas três últimas nas suas versões centrais e não centrais. Modelos que consideram transformações nas variáveis (resposta e explicativas) serão abordados. Estudar-se-á o ajuste dos modelos por mínimos quadrados (ordinários, ponderados e generalizados) e máxima verossimilhança. Metodologias de seleção de modelos (covariáveis) serão discutidas. As questões relativas à inferência (estimação, intervalo de confiança e testes de hipótese) serão apresentadas sob o enfoque frequentista. Apresentar-se-á uma introdução aos modelos de regressão não lineares. Serão apresentadas/discutidas análise de dados (aplicações) com o apoio computacional do pacote R.

1. Motivação

- 1.1 Exemplos motivadores e discussão sobre algumas ideias de modelagem.
- 1.2 Sobre análise de dados resolução de problemas científicos (pesquisa) e consultoria.

2. Revisão de Álgebra de Matrizes

- 2.1 Definições e propriedades básicas.

3. Distribuição normal multivariada

3.1 Apresentação.

3.2 Propriedades: distribuições marginais, condicionais e função característica.

4. Introdução aos modelos de regressão lineares

4.1 Modelo de regressão linear simples (MRLS).

4.2 Estimação dos parâmetros por mínimos quadrados ordinários e propriedades dos estimadores (MRLS).

4.3 Modelo de regressão linear múltiplo (MRLM).

4.4 Estimação dos parâmetros por mínimos quadrados ordinários e propriedades dos estimadores (MRLM).

4.5 Forma matricial do MRLM.

4.6 Estimação por mínimos quadrados ordinários.

4.7 Estimação por mínimos quadrados generalizados.

4.8 Aplicações e discussão sobre resolução de problemas/análise de dados.

5. Distribuições de probabilidade não centrais

5.1 Distribuição t de Student não central.

5.2 Distribuição Qui-quadrado não central.

5.3 Distribuição F de Snedcor não central.

6. Intervalos de Confiança e Testes de Hipótese

6.1 Distribuição de formas quadráticas Gaussianas.

6.2 Intervalos de confiança e testes de hipótese (MRLS).

6.3 Intervalos de confiança e testes de hipótese: Tabela ANOVA e Testes do tipo

$H_0: \mathbf{C}\boldsymbol{\beta} = \mathbf{M}$ vs $H_1: \mathbf{C}\boldsymbol{\beta} \neq \mathbf{M}$ (MRLM).

6.4 Aplicações e discussão sobre resolução de problemas/análise de dados .

7. Mecanismo de validação/comparação de modelos

7.1 Análise residual.

7.2. Comparação e seleção de modelos.

7.3 Aplicações e discussão sobre resolução de problemas/análise de dados.

8. Tópicos adicionais de modelagem e transformações de variáveis

8.1 Transformações nas variáveis explicativas.

8.2. Transformações na variável resposta.

8.3 Multicolinearidade

8.4 Modelos polinomiais

8.5 Teste para falta de ajuste

8.6 Introdução à regressão segmentada

8.7 Aplicações e discussão sobre resolução de problemas/análise de dados.

9. Introdução aos modelos de regressão não-lineares

9.1 Definição.

9.2 Estimação.

9.3 Intervalos de Confiança e Testes de hipótese.

9.4 Aplicações e discussão sobre resolução de problemas/análise de dados.

2. Bibliografia (em ordem alfabética)

- Caffo, B. (2019). **Regression models for data science in R** ([link](#)).
- Draper, N. R. and Smith, H. (1998). **Applied regression analysis, third edition**. New York, NY: John Wiley & Sons, ***disponível na biblioteca*** ([conjunto de dados](#)).
- Faraway, J. J. (2014). **Linear Models with R, Second Edition**, Chapman & Hall/CRC Texts in Statistical Science, ***disponível na biblioteca do IEL*** ([conjuntos de dados e códigos](#))
- Fox, J. (2016). **Applied regression analysis and generalized linear models, third edition**. Los Angeles, CA, Sage, ***disponível na biblioteca***.
- Kutner, M. H., Nachtsheim, C. J., Li, W. (2005). **Applied linear statistical models, fifth edition**. Boston, MA : McGraw-Hill Irwin, ***disponível na biblioteca*** ([conjunto de dados](#)) .
- Paula, G. A. (2013) **Modelos de regressão com apoio computacional** ([link](#))
- Rawlings, J. O, Pantula, J. O. and Dickey, D. A. (1998). **Applied regression analysis: a research tool, second edition**, New York, NY: Springer, ***disponível na biblioteca***.
- Sheater, S. (2009). **A Modern Approach to Regression with R**, Springer Texts in Statistics, ***disponível na biblioteca*** ([conjunto de dados e códigos](#)).
- Magnus, J. R. and Neudecker, H. (1998). **Matrix differential calculus with applications in Statistics and econometrics**, first edition, Chichester, John Wiley, ***disponível na biblioteca***.
- Weisberg, S (2013). **Applied Linear Regression**, Wiley Series, fourth editon, ***disponível na biblioteca*** ([conjuntos de dados](#)).
- Notas de aula e materiais adicionais a serem disponibilizados na página do curso.
- Referências adicionais a serem divulgadas, eventualmente, ao longo do curso, no site.

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](#).

3. Critérios de avaliação

Metodologia de avaliação: duas provas, um seminário e um trabalho, uma prova substitutiva e um exame (estes dois últimos, se forem necessários).

- Listas de exercícios serão disponibilizadas periodicamente. Recomenda-se resolvê-las na íntegra.
- O trabalho consistirá em análises de conjuntos de dados (disponíveis no site, nas listas e, eventualmente, a serem coletados pelos alunos), através de metodologias vistas em sala e/ou a serem pesquisadas pelos alunos, conforme sugestão do professor e eventualmente, exercícios envolvendo desenvolvimentos metodológicos. O(s) problema(s) em questão serão disponibilizados, ao longo do semestre, nas listas de exercícios. Assim, os alunos que forem resolvendo as listas, à medida em que elas forem sendo entregues, terão grande parte do trabalho concluído, com antecedência. O trabalho deverá ser feito em equipes de pelo menos dois e no máximo quatro componentes. Os membros de cada equipe serão inteiramente responsáveis pela elaboração e entrega do trabalho. O trabalho deverá ser entregue impresso (pessoalmente), em data e local a serem anunciados. Essa data 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).
- Os seminários deverão ser apresentados (de forma oral), com auxílio de slides (a escolha do programa é livre), por grupo de um ou dois alunos (todos devem apresentar uma parte do seminário). Deverá versar sobre algum dos temas disponíveis no site do curso ou a ser sugerido pela equipe (deve ser enviado um e-mail com o tópico sugerido e aguardar o parecer do professor). A definição dos temas e das equipes deve ser feito até o dia 30/04, as 18h00. Um dos membros da equipe deverá enviar um e-mail com o nome dos membros e o tópico escolhido/sugerido (naturalmente bem antes da data/horário acima, para que se tenha tempo hábil para registrar tal informação e/ou julgar a sugestão de tópico) . Obviamente, o mesmo tema não poderá ser apresentado por mais de uma equipe. No seminário deverá ser apresentado: um exemplo motivador, a metodologia escolhida, em que ele se diferencia (vantagens e desvantagens) em relação ao modelo de regressão normal linear homocedástico e a correta análise (comparando as duas metodologias) do exemplo motivador. Os slides da apresentação deverão ser enviados, por e-mail, até 2h antes da respectiva apresentação. Quando todas as equipes/temas estiverem definidos, as datas e horários das apresentações serão selecionadas.

- Média das provas (MP) = $0,5*NP1 + 0,5*NP2$, em que NP_i : nota da i-ésima prova, $i=1,2$ (provas regulares).
- Média Global (MG), calculada da seguinte forma:
- Se o(a) aluno(a) entregar o trabalho completo e apresentar o seminário e assistir a todos os outros seminários, $MG = 0,50*MP + 0,30*NT + 0,20*NS$, em que NT = nota do trabalho, NS = nota do seminário.
 - Se o(a) aluno(a) não entregar o trabalho completo e/ou não apresentar o seminário e/ou não assistir a todos os outros seminários $MG = 0,5*(0,50*MP + 0,30*NT + 0,20*NS)$. **Se o trabalho não for entregue, então, NT= 0. Se o seminário não for apresentado, então, NS = 0.**
 - Se $MG \geq 7,0$ o(a) aluno(a) estará aprovado(a), se $MG < 2,5$, estará reprovado(a), caso contrário, terá de fazer EXAME. OBS: o(a) aluno(a) aprovado(a) poderá fazer o EXAME, para melhorar sua nota. Entretanto, ele (a) deverá comunicar sua decisão até uma semana (sete dias) antes, pessoalmente, na sala do Professor.
 - A prova substitutiva (PS) será feita mediante solicitação, via e-mail, (até o dia 27/06/2019, as 18h00), o qual deve conter: nome completo, RA, o pedido, nome da disciplina e a(s) justificativa(s) da(s) ausência(s) na(s) prova(s), quando pertinente. Uma vez solicitada, caso o aluno não realize a PS, sua nota, nela, será igual a zero.
 - Caso o aluno tenha feito as duas provas regulares (PR), **a nota da PS substituirá a menor delas, se $MG < 7,0$ (mesmo que $MG < 2,5$). Ou seja, neste caso, a MG poderá diminuir. Se $MG \geq 7,0$, a nota da PS só substituirá a menor das notas das PR, caso esta seja menor do que aquela. Caso contrário, a MG permanecerá inalterada.** Se o aluno não tiver feito pelo menos uma delas (PR), a nota da PS será usada da seguinte forma:
 - Se o aluno tiver faltado a somente uma prova regular, a nota da PS substituirá a nota da prova não realizada.
 - Se o aluno faltar as duas provas regulares, a nota da PS substituirá a nota de uma das provas regulares, enquanto que a nota do EXAME substituirá a outra. Neste caso, a nota do EXAME também servirá como a própria.
 - Provas não realizadas equivalem a ter, nessas provas, nota zero.
- Média Final (MF) (em caso de Exame):
 - $MF = 0,5*MG + 0,5*NE$; NE: nota do exame.

- Se $MF \geq 5,0$, o aluno(a) estará aprovado(a), caso contrário, estará reprovado(a).
- Caso o aluno tenha sido aprovado com a MG e tenha feito exame, sua média final será $MF = \max(MG, MF)$.
- A frequência mínima para aprovação é de 75%.
- Caso o(a) aluno(a) seja reprovado(a) por faltas e tiver de fazer EXAME, sua nota (no EXAME), será zero.
- Sobre abono de faltas veja ([link](#)).