

ME613 – Análise de regressão

Segundo Semestre de 2016

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 e-mail deverá ser utilizado somente para: solicitação de agendamento de atendimento, justificativa de ausência em um atendimento agendado e para o envio de eventuais correções relativas ao site do curso, incluindo os materiais disponibilizados. Dúvidas serão sanadas somente durante o atendimento e durante as aulas.

Aulas: Segundas, sala CB09 e Quartas, sala CB02, das 14h00-16h00.

Atendimento :

- (Professor) Terça-feiras, das 13h às 14h, na sala do Professor.
- O(a) aluno(a) deverá enviar um e-mail (para o supracitado endereço) com 24 horas de antecedência, solicitando o agendamento de atendimento para o dia posterior. Por exemplo, se ele(a) quiser atendimento para o dia 06/09, deverá enviar um e-mail no dia 05/09 até as 13h00. 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.

Página na internet do curso: http://www.ime.unicamp.br/~cnaber/Material_ME613_2S_2016.html

Auxiliar didático (PAD): Rodolfo Lindemute

Monitoria (atendimento pelo auxiliar didático): A confirmar.

1. Programa

Resumo:

Apresentar os principais aspectos dos modelos de regressão lineares, com ênfase no uso da distribuição normal para modelar a variável resposta. 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 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.

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.

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
 $H_0: C\beta = M$ vs $H_1: C\beta \neq M$ (MRLM).
- 6.4 Aplicações.

7. Mecanismo de validação/comparação de modelos

7.1 Análise residual.

7.2. Comparação e seleção de modelos.

7.3 Aplicações.

8. Tópicos adicionais de modelagem e transformações de variáveis

8.1 Transformações nas variáveis explicativas.

8.2. Transformações na variável resposta.

8.4 Multicolineariedade

8.5 Modelos polinomiais

8.6 Teste para falta de ajuste

8.7 Introdução à regressão segmentada

8.8 Aplicações.

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.

2. Bibliografia

- **Básica**

- Draper, N. R. and Smith, H. (1998). **Applied regression analysis, third edition. New York, NY: John Wiley & Sons, disponível na biblioteca.**

- Faraway, J. J. (2014). **Linear Models with R, Second Edition, Chapman & Hall/CRC Texts in Statistical Science, disponível na biblioteca do IEL.**
 - Fox, J. (2008). **Applied regression analysis and generalized linear models, second 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.**
 - 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.**
- **Complementar**
 - Magnus, J. R. and Neudecker, H. (1998). **Matrix differential calculus with applications in Statistics and econometrics, first edition, Chichester, John Wiley, disponível na biblioteca**
 - 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.

3. Critérios de avaliação

Metodologia de avaliação: três provas e um trabalho.

- Listas de exercícios serão disponibilizadas periodicamente. Recomenda-se resolvê-las na íntegra.
- O trabalho consistirá em análises de conjuntos de dados, através de metodologias vistas em sala e/ou a serem pesquisadas pelos alunos, conforme sugestão do professor. Os conjuntos de dados poderão estar presentes (eventualmente de modo antecipado), também, nas listas de exercícios. O trabalho deverá ser entregue na forma de relatório o qual deverá incluir uma descrição do conjunto de dados e análises descritivas, além das análises inferenciais. O(s) pacote(s) computacional(ais) a ser(em) usado(s) nas análises é(são) de livre escolha. Futuramente mais detalhes acerca dos trabalhos (data de entrega, formato, número de páginas, conjuntos de dados etc) serão divulgados na página do curso e discutidos em sala. Recomenda-se, o quanto antes, que os alunos busquem formar os grupos.

- Média Global (MG):
 - NP_i: nota da i-ésima prova, i=1,2,3. Se $\min(\text{NP}_1, \text{NP}_2, \text{NP}_3) \geq 5,0$, então $\text{MGP} = 0,5 \cdot \text{NP}(1) + 0,5 \cdot \text{NP}(2)$, em que NP(1) é o máximo(NP₁,NP₂,NP₃) e NP(2) é a segunda maior nota. Caso contrário $\text{MGP} = (1/3) \cdot \text{NP}_1 + (1/3) \cdot \text{NP}_2 + (1/3) \cdot \text{NP}_3$ (MGP: média global parcial).
 - Se o(a) aluno(a) entregar o trabalho completo, então **$\text{MG} = 0,7 \cdot \text{MGP} + 0,3 \cdot \text{NT}$** , caso contrário **$0,6 \cdot (0,7 \cdot \text{MGP} + 0,3 \cdot \text{NT})$** em que NT é a nota do trabalho.
 - Se $\text{MG} \geq 7,0$ o(a) aluno(a) estará aprovado(a). Se $2,5 \leq \text{MG} < 7,0$, terá de fazer EXAME. Se $\text{MG} < 2,5$, estará reprovado. 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.

- Média Final (MF):
 - Se $\text{MG} \geq 7,0$; $\text{MF} = \max(\text{MG}, \text{ME})$, se $2,5 \leq \text{MG} < 7,0$, $\text{MF} = \text{ME}$, em que,
 - $\text{ME} = 0,5 \cdot \text{MG} + 0,5 \cdot \text{NE}$; NE: nota do exame.
 - Se $\text{ME} \geq 5,0$, o(a) aluno(a) estará aprovado(a), caso contrário, estará reprovado(a).

- Frequência mínima para aprovação é de 75%