

## ME 720- Modelos Lineares Generalizados

Primeiro Semestre de 2016

**Professor:** Caio L. N. Azevedo

**Sala do professor:** 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** será usado somente para enviar e-mail aos alunos. Caso o(a) aluno(a) queira se comunicar via e-mail, faça-o através do supracitado endereço (não enviar e-mails através do ensino aberto).
- 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 constantes no site do curso, incluindo os materiais disponibilizados. Dúvidas serão sanadas somente durante o atendimento e durante as aulas.

**Atendimento :** Terças, das 13h às 14h (na supracitada sala).

- 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 15/03, deverá enviar um e-mail no dia 14/03 até as 13h00. O(a) aluno(a) que não comparecer à um atendimento agendado e não justificar devidamente (por e-mail) o motivo de sua ausência, não poderá mais solicitar agendamento de atendimento.

**Aulas:** Terças e Quintas Sala PB 11, das 16h00 às 18h00.

**Página do curso:** [http://www.ime.unicamp.br/~cnaber/Material\\_MLG\\_1S\\_2016.htm](http://www.ime.unicamp.br/~cnaber/Material_MLG_1S_2016.htm)

**Monitoria:** Não haverá

## **1. Programa**

Resumo:

As suposições de normalidade e homocedasticidade dos dados são reconhecidamente inválidas em diversos exemplos na literatura e situações reais. Ademais, estruturas (modelos) lineares, muitas vezes, não são apropriadas para modelar a média da variável resposta em função de covariáveis de interesse. Sendo assim, relaxar tais suposições é de extrema importância para que uma análise mais realista seja conduzida. A classe de modelos de regressão lineares generalizados (MLG), e outras classes dela derivadas, constituem alternativas de análise que contemplam tal relaxamento. Portanto, o estudo da classe de MLG proverá aos alunos ferramentas adicionais em termos de análise de dados. Sendo assim, nesta disciplina, pretendemos apresentar e discutir os principais modelos da classe dos lineares generalizados bem como mecanismos de inferência, estimação, testes de hipótese e validação de modelos sob a ótica frequentista. Estudaremos modelos para dados positivos assimétricos, para dados binários, para dados de contagem, para dados de contagem com sobredispersão, para dados de contagem com excesso de zeros. Também discutiremos, com menor ênfase, à respeito de modelos não lineares e modelos lineares mistos. Consideraremos o pacote R para concretização da análise de dados.

### 1. Introdução

- a. Motivação.
- b. Notação.
- c. Revisão sobre modelos de regressão normais lineares homocedásticos.
- d. Família exponencial.

### 2. Modelos Lineares Generalizados

- a. Definição.
- b. Função desvio.
- c. Estimação dos parâmetros.
- d. Teste de hipóteses.
- e. Técnicas de diagnóstico.
- f. Seleção de modelos.
- g. Aplicações.

### 3. Modelos para dados positivos assimétricos

- a. Modelos com resposta gama.
- b. Estimação, teste de hipóteses, técnicas de diagnóstico e seleção de modelos.
- c. Aplicações.

### 4. Modelos para dados binários

- a. Dados agrupados (Bernoulli).
- b. Dados não agrupados (binomial).
- c. Regressão logística linear.
- d. Estimação, teste de hipóteses, técnicas de diagnóstico e seleção de modelos.
- e. Aplicações.

### 5. Modelos para dados de contagem

- a. Modelos com resposta Poisson.
- b. Modelos com resposta binomial negativa (dados com sobredispersão).
- c. Estimação, teste de hipóteses, técnicas de diagnóstico e seleção de modelos.
- d. Aplicações.
- e. Modelos para dados com excesso de zeros.
- f. Estimação, teste de hipóteses, técnicas de diagnóstico e seleção de modelos.
- g. Aplicações.

### 6. Tópicos adicionais

- a. Modelos não lineares.
- b. Modelos lineares mistos.

## **2. Bibliografia**

- Paula, G. A. (2013). ***Modelos de Regressão: Com Apoio Computacional, Versão Preliminar, 428 p., IME-USP ([https://www.ime.usp.br/~giapaula/texto\\_2013.pdf](https://www.ime.usp.br/~giapaula/texto_2013.pdf)).***
- Atkinson, A. C. (1985). ***Plots, Transformations, and Regressions, Oxford Science Publications, Oxford.***
- Collett, D. (2003). ***Modelling Binary Data, second edition, Chapman and Hall, London, disponível na biblioteca.***
- Hosmer, D. W. e Lemeshow, S. (2000). ***Applied Logistic Regression, second edition, Wiley, New York, NY, disponível na biblioteca.***
- Lindsey, J. K. (1997). ***Applying Generalized Linear Models, first edition, Springer, New York, NY, disponível na biblioteca.***
- McCullagh, P. e Nelder, J. A. (1989). ***Generalized Linear Models, second Edition, Chapman and Hall, London, disponível na biblioteca.***
- McCulloch, C. E. e Searle, S. R. (2008). ***Generalized, Linear, and Mixed Models, second edition, Wiley, New York, disponível na biblioteca.***
- Montgomery, D. C.; Peck, E. A. e Vining, G. G. (2012). ***Introduction to Linear Regression Analysis, fifth edition. John Wiley, New York, 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.***
- Faraway, J. J. (2014). ***Linear Models with R, second Edition, Chapman and Hall, London, disponível na biblioteca (IEL).***
- Faraway, J. J. (2005). ***Extending the Linear Model with R: Generalized Linear, Mixed Effects and Nonparametric Regression Models, first edition, Chapman and Hall, London, disponível na biblioteca.***
- Notas de aula e material adicional 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: duas provas, dois trabalhos e um seminário.
- Periodicamente serão disponibilizadas listas de exercícios. Recomenda-se, fortemente, resolvê-las na íntegra.
- Cada 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. Os trabalhos deverão ser feitos e entregues em grupos de 2 a 4 alunos. Não serão aceitos trabalhos cujos grupos tenham um número fora desse intervalo. Recomenda-se, o quanto antes, que os alunos busquem formar os grupos.
- Os seminários serão individuais e consistirão na apresentação de assuntos, relativos à modelos de regressão, que não serão vistos durante curso. Em breve, tópicos a serem escolhidos serão divulgados, bem como informações adicionais (tempo e formato da apresentação etc). Caberá a cada aluno(a) pesquisar sobre o tema, embora possa recorrer ao Professor acerca de referências e eventuais dúvidas. O objetivo principal é o que o(a) aluno(a) apresente as principais ideias e utilidades da metodologia de análise escolhida, através da breve análise de um conjunto de dados reais.
- Média Global (MG):
  - Se o aluno(a) entregar os dois trabalhos (completos), apresentar seu seminário e assistir a todos os outros:  $MG = 0,65*MP + 0,25*MT + 0,10*NS$ , em que  $MP = 0,4*NP1 + 0,6*NP2$ ,  $NP_i$  é a nota da i-ésima prova,  $i=1,2$ ,  $MT = 0,5*NT1 + 0,5*NT2$ ,  $NT_i$  é a nota do i-ésimo trabalho,  $i=1,2$  e  $NS$  é a nota do seminário. Caso contrário, ou seja, se ele(a) não entregar pelo menos um dos trabalhos completo e/ou não apresentar seminário e/ou não assistir a todos os demais,  $MG = 0,6*(0,65*MP + 0,25*MT + 0,10*NS)$ .
  - Se  $MG \geq 7,0$ , o(a) aluno(a) estará aprovado(a). Se  $2,5 \leq MG < 7,0$ , terá de fazer EXAME. Se  $MG < 2,5$ , 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  $MG \geq 7,0$ ;  $MF = \text{máximo}(MG, ME)$ , se  $2,5 \leq MG < 7,0$ ,  $MF = ME$ , em que,  
 $ME = 0,5 \cdot MG + 0,5 \cdot NE$ ; NE: nota do exame.
  - Se  $ME \geq 5,0$ , o(a) aluno(a) estará aprovado(a), caso contrário, estará reprovado(a).
  
- A frequência mínima para aprovação é de 75%. Recomenda-se, fortemente, que o(a) aluno(a) monitore com atenção seu respectivo número de faltas. Tal informação também poderá ser obtida no horário de atendimento.