

## ME 714 – Análise de Dados Discretos

Primeiro Semestre de 2017

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

**Sala do professor:** 210 IMECC

**e-mail:** [cnaber@ime.unicamp.br](mailto: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, justificativa de ausência nas aulas 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.

### **Atendimento :**

- (Professor) Quartas-feiras, das 13h às 14h, na sala do Professor, mediante agendamento.
- 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 a um atendimento agendado e não justificar devidamente (por e-mail) o motivo de sua ausência, não poderá mais solicitar agendamento de atendimento.

**Aulas:** Segundas e Quartas, das 16h00 às 18h00, sala CB 09

**Página na internet do curso:** [http://www.ime.unicamp.br/~cnaber/Material\\_ADD\\_1S\\_2017.htm](http://www.ime.unicamp.br/~cnaber/Material_ADD_1S_2017.htm)

**Monitoria:** Não haverá

## **1. Programa**

Resumo:

Apresentar as principais metodologias estatísticas relativas à análise de dados discretos, incluindo alguns tópicos de análise de dados categorizados (categóricos). Revisão sobre variáveis aleatórias e vetores aleatórios discretos e funções geradoras (geratrizes). Testes de hipótese e medidas de associação relativas a tabelas de contingência. Modelos de regressão linear com variáveis categorizadas na matriz de planejamento. Modelos de regressão para tabelas de contingência. Modelos de regressão para respostas binárias e politômicas (nominais e ordinais). Modelos lineares generalizados. Modelo de regressão de Poisson. Métodos de estimação por máxima verossimilhança e mínimos quadrados. Validação, verificação da qualidade de ajuste e seleção de modelos. O suporte computacional será desenvolvido/apresentado na plataforma R (eventualmente o SAS pode ser utilizado).

### **1 Revisão sobre variáveis aleatórias e vetores aleatórios discretos**

- 1.1** Definição e funções geradoras (ou geratrizes) de momentos e de probabilidade.
- 1.2** Algumas distribuições discretas: Binomial, Multinomial, Poisson, Geométrica, Hipergeométrica, Binomial Negativa.

### **2 Tabelas de Contingências**

- 2.1** Características e análises de interesse.
- 2.2** Testes de homogeneidade e independência.
- 2.3** Teste exato de Fisher e teste de McNemar.
- 2.4** Testes de Mantel Haenszel e Teste de Pearson.
- 2.5** Medidas de associação.
- 2.6** Aplicações.

### **3 Revisão sobre modelos de regressão (normais) lineares**

**3.1** Definição e propriedades.

**3.2** Regressão linear com variáveis categorizadas (categóricas) na matriz de planejamento.

**3.3** Equivalência entre as parametrizações da matriz de planejamento com variáveis categorizadas.

**3.4** Aplicações.

### **4 Modelos de regressão para variáveis categorizadas (categóricas)**

**4.1** Modelos de regressão para tabelas de contingências.

**4.2** Modelos de regressão para variáveis binárias.

**4.3** Modelos de regressão para variáveis politômicas.

### **5 Modelos de regressão para outros tipos de variáveis discretas**

**5.1** Modelos lineares generalizados.

**5.2** Modelos de regressão de Poisson.

**5.3** Validação e seleção de modelos.

**5.4** Aplicações.

## **2. Bibliografia Básica**

- Agresti. A. (2012). *Categorical data analysis, terceira edição*. New York, John Wiley. **Disponível na biblioteca (primeira e segunda edições).**
- Agresti. A. (2007). *An introduction to categorical data analysis, segunda edição*. New Jersey, Wiley-Interscience. **Disponível na biblioteca.**

- Finney, D. J. (1971). ***Probit analysis, terceira edição***, Cambridge, University Press. **Disponível na biblioteca.**
- Hosmer, D. W. and Lemeshow, S. (2013). ***Applied logistic regression, terceira edição***, New York, John Wiley. **Disponível na biblioteca.**
- McCullagh, P. and Nelder, J. A. (1989). ***Generalized linear models, second edition***, Chapman & Hall/CRC. **Disponível na biblioteca.**
- Paulino, C. D. e Singer, J. M. (2006). ***Análise de Dados Categorizados***, Edgard Blücher. **Disponível na biblioteca**
- Paula, G. A. (2013). ***Modelos de regressão com apoio computacional, versão preliminar***. Disponível em: [http://www.ime.usp.br/~giapaula/texto\\_2013.pdf](http://www.ime.usp.br/~giapaula/texto_2013.pdf).
- Materiais extras a serem disponibilizados no site.

### **3. Critérios de avaliação**

- Metodologia de avaliação: três provas e um trabalho.
- Periodicamente serão disponibilizadas listas de exercícios. Recomenda-se, fortemente, 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, necessariamente, estarão presentes 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). Futuramente, todos os detalhes acerca do trabalho (data de entrega, conjuntos de dados etc) serão divulgados na página do curso e discutidos em sala. O trabalho deverá ser feito em grupos de 2 a 5 alunos. Recomenda-se que as equipes sejam constituídas o quanto antes. Não serão aceitos trabalhos entregues por equipes com um número de componentes fora desse intervalo. A partir da data de divulgação das questões do trabalho, dar-se-á, no máximo, duas semanas para sua entrega. Assim, recomenda-se, fortemente, que todos os exercícios das listas sejam resolvidos.

- Média Global (MG):
  - NP<sub>i</sub>: nota da i-ésima prova, i=1,2,3. Se  $\text{mínimo}(NP_1, NP_2, NP_3) \geq 6,0$ , então  $MGP = 0,5 \cdot NP(1) + 0,5 \cdot NP(2)$ , em que NP(1) é o máximo(NP<sub>1</sub>, NP<sub>2</sub>, NP<sub>3</sub>) e NP(2) é a segunda maior nota. Caso contrário  $MGP = (1/3) \cdot NP_1 + (1/3) \cdot NP_2 + (1/3) \cdot NP_3$  (MGP: média global parcial).
  - Se o(a) aluno(a) entregar o trabalho completo, então  **$MG = 0,7 \cdot MGP + 0,3 \cdot NT$** , caso contrário  **$0,6 \cdot (0,7 \cdot MGP + 0,3 \cdot NT)$**  em que NT é a nota do trabalho.
  - 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).
  
- O(a) aluno(a) que faltar a pelo menos uma prova fará uma única prova (“substitutiva”), em data a ser definida, cuja nota será alocada a cada uma das provas não realizadas.
- Frequência mínima para aprovação é de 75%