



Bacharelado em Estatística

UFJF – ICE - Departamento de Estatística

DISCIPLINA: Análise de Dados Categóricos (EST 043 – A) – 1º. Sem. 2020 – Sala 3504 – ICE
(3as e 5as de 08:00 às 10:00)

PROFESSOR: Ronaldo Rocha Bastos, Ph.D (University of Liverpool, UK)

OBJETIVO GERAL DA DISCIPLINA:

Em muitas situações nas ciências sociais aplicadas, na indústria, em medicina e biologia, na área financeira e marketing, entre outras, os dados coletados são categóricos ou categorizados (*qualitativos*) e não numéricos (*quantitativos*). Em geral, os pesquisadores destas áreas estão interessados em analisar as associações e relações entre pares ou entre grupos de variáveis deste tipo (*categóricas nominais* ou *ordinais*). Estes dados, geralmente podem ser apresentados na forma de tabelas de contingência (no mínimo bidimensionais), devendo ser analisados por técnicas e métodos apropriados. O objetivo deste curso é apresentar e discutir os procedimentos de análise mais importantes, desde análises exploratórias até análises confirmatórias.

Estas análises serão sempre realizadas tendo em vista¹:

- a) quais as questões substantivas que se deseja responder;
- b) qual o plano amostral utilizado; e qual a população objetivo, real ou conceitual, que originou a amostra;
- c) qual o modelo probabilístico mais adequado aos dados obtidos, com base nos itens anteriores e também no senso comum;
- d) quais os parâmetros do modelo probabilístico escolhido serão usados nas hipóteses
- e) qual a metodologia estatística utilizada para ajustar os modelos especificados
- f) comparação entre modelos concorrentes;
- g) tradução dos resultados obtidos em respostas aos questionamentos substantivos

REFERÊNCIAS BIBLIOGRÁFICAS:

- AGRESTI, A. 2007. *An Introduction to Categorical Data Analysis – 2nd Edition*. New York: Wiley.
- D'HAINAUT, L. 1992. *Conceitos e Métodos da Estatística*. vol. 2 – Duas ou Três Variáveis Segundo Duas ou Três Dimensões. Lisboa: Fundação Calouste Gulbenkian.
- EVERITT, B.S. 1992. *The analysis of contingency tables – 2ª ed.*. Monographs on statistics and applied probability 45. London: Chapman & Hall.
- GREENACRE, M. 2007. *Correspondence Analysis in Practice*, 2ª edição. Boca Raton: CRC Press.
- JOBSON, J.D. 1992. *Applied multivariate data analysis - Vol. 2: Categorical and Multivariate methods*. Springer texts in statistics. New York: Springer.
- MURTEIRA, B.J.P. 1993. *Análise Exploratória de Dados: Estatística Descritiva – Cap. 8* – Lisboa: McGraw-Hill.
- MURTEIRA, B.J.P. 1990. *Probabilidade e Estatística - Vol. 2 (cap. 16)*. Lisboa: McGraw-Hill.
- NISHISATO, S., 2007. *Multidimensional Nonlinear Descriptive Analysis*. Boca Raton: CRC Press.
- PAULINO, C.D. E SINGER 2006. *Análise de Dados Categorizados*. São Paulo: Blücher.

CALENDÁRIO:

Mar. 03, 05, 10, 12, 17, 19, 24, 26, 31; **Abr.** 02, 07, 09, 14, 16, 21, 23, 28, 30 (1o.TVC); **Mai.** 02, 05, 07, 12, 14, 19, 21, 26, 28; **Jun.** 02, 04, 09, 11, 16, 18, 23, 25, 30 (2o.TVC);

¹ Ver Paulino e Singer (2006) para um detalhamento desta forma de abordagem

PROGRAMA DO CURSO:

1ª Parte :

1. Escalas de mensuração, escalonamento, tipos de variáveis categóricas.
2. Introdução à análise de dados categóricos, objetivos e enfoque a serem utilizados. Tabelas de contingência: notação e exemplos. Revisão dos modelos utilizados para dados discretos univariados: Bernoulli e binomial; hipergeométrico; geométrico e binomial negativo; Poisson.
- 3 Tabelas de contingência bidimensionais e suas formas de análise: proporções, “odds ratios”. Tabelas 2 X 2 e Tabelas 1 X c. Os modelos probabilísticos adotados para as observações em tabelas de contingência: Produto de distribuições de Poisson, Multinomial, Hipergeométrica, Produto de distribuições Multinomiais. Variáveis resposta (desfecho) e variáveis explicativas (fatores). Graus de liberdade.
4. A estatística qui-quadrado de Pearson. Correção de Yates. Qui-quadrado de McNemar. Agregação de tabelas 2 x 2. Partição de tabelas 2 X c em tabelas 2 X 2. Análise de resíduos das células em tabelas de contingência.
5. Tabelas tridimensionais e multidimensionais. Independência mútua, independência parcial, independência condicional, efeitos de segunda ordem. Estatística de Cochran-Mantel-Haenszel. Estatística de Breslow-Day. Paradoxo de Simpson.
6. Breve introdução a Modelos Lineares Generalizados

2ª Parte:

7. Modelos loglineares para tabelas de contingência. Especificação, ajuste, análise e interpretação no contexto de MLG.
8. Variáveis ordinais: testes para tendência linear. Associação linear \times linear. Efeitos de linha e efeitos de coluna. Especificação, ajuste, análise, interpretação no contexto de MLG
9. Modelos logísticos lineares binários. Especificação, ajuste, análise, interpretação no contexto de MLG. Estratégias para modelos de classificação e modelos de previsão.
10. Modelos probit. Especificação, ajuste, análise, interpretação no contexto de MLG.
11. Modelos Poisson. Especificação, ajuste, análise, interpretação no contexto de MLG.
12. Análise de Correspondência Simples: princípios, notação algébrica, interpretação geométrica e análise gráfica.

AVALIAÇÃO:

Duas provas escritas, marcadas para os dias **30/04** (quinta-feira) – 1º. TVC (40%) e **30/06** (terça-feira) – 2º. TVC (40%), além de uma segunda chamada única (matéria toda) para aqueles que não puderem comparecer a um dos dois TVCs sem apresentação de justificativa, marcada para o dia **02/07** (quinta-feira). Não haverá **em hipótese alguma** prova substitutiva para qualquer avaliação. A **terceira nota** (20%) será correspondente à avaliação das apresentações/seminários e listas/ trabalhos apresentados pelos alunos, individualmente ou em grupo, conforme combinado durante o curso.

ATENDIMENTO:

O Prof. Ronaldo estará disponível para atendimento às quartas-feiras entre 14:30 e 17:30 (mediante marcação), ou em qualquer outro horário previamente combinado. Para entrar em contato usar o email: ronest@terra.com.br.

But Statistics must be made otherwise than to confirm preconceived ideas.

(Florence Nightingale; 1820-1910)

Un linguaggio diverso è una diversa visione della vita.

(Federico Fellini; 1920-1993)

Statistics is the new grammar.

(Clive Thompson. Revista *Wired*: 19/04/2010)