



Universidade de Brasília - IE - Departamento de Estatística
Professor George von Borries

1. Informações Gerais

- Disciplina: Análise de Dados Categorizados - EST0064 (SIGAA)
- Professor: George F. von Borries
- Horário: Segunda e Quarta - 10 às 11h.
- Local: Plataforma MS Teams.
- Pré-requisitos: Análise de Regressão Linear.
- Livro Texto: Agresti (2019) [2] e/ou Agresti (2013) [1].
- Horário de atendimento: Após as aulas, de 11 às 11h30.
- E-mail: gborries@unb.br (Favor colocar no título do e-mail: **DCAT**)¹
- Ambiente de aprendizagem: <http://www.aprender3.unb.br>.

1.1 Informações sobre o Curso Não Presencial

1. O curso será inteiramente não presencial.
2. As aulas ocorrerão através da Plataforma MS Teams da UnB (eventualmente no Google Meets em caso de problema com o MS Teams).
3. As notas de aula, em slides, serão disponibilizadas em arquivos PDF no site do curso. Listas de exercícios e material complementar também serão disponibilizados na área do curso no Aprender 3.
4. Atendimento de dúvidas será realizado no final de cada aula pelo professor.
5. As aulas não serão gravadas. Alunos que eventualmente realizarem gravação, devem fazer para uso pessoal e não disponibilizar em meios eletrônicos/mídias sociais.
6. Os alunos serão inscritos no site do curso (no Aprender 3) pelo professor.

¹E-mail deve ser utilizado apenas para questionamentos gerais. Para dúvidas gerais sobre a teoria e exercícios, favor utilizar os horários de atendimento.

2. Descrição Geral

O curso visa introduzir técnicas de análise de dados categorizados (com resposta binária, multicategóricas e contagens). O objetivo é a análise de Tabelas de Contigência e também na criação e avaliação de modelos com respostas categóricas. Na modelagem, as variáveis explicativas podem ser discretas e/ou contínuas.

2.1 Referências Básicas e Computação

O curso é baseado nos textos de Agresti [2, 1]. Agresti (2019) [2] é um livro de nível técnico mais simples e com apelo mais computacional através da apresentação de várias aplicações com base no software R. Agresti (2013) [1] é um livro mais avançado que exige melhor base matemática. O livro apresenta alguns resultados obtidos com SAS e R, e possui um website com códigos de vários exemplos utilizando R, SAS, STATA e SPSS

(<http://users.stat.ufl.edu/~aa/cda/cda.html>).

Este curso fará uso do software livre R, disponível em www.r-project.org ou <https://rstudio.cloud> e do software SAS, disponível através do link https://www.sas.com/pt_br/software/on-demand-for-academics.html.

O uso dos softwares R e SAS melhoram o ensino e aprendizagem de novas ideias e conceitos do curso. Uma boa referência para o uso do SAS em Dados Categorizados é o livro de Stokes (2012) [9].

2.2 Ambiente de Aprendizagem

O curso possui uma página no moodle, <http://www.aprender3.unb.br>. Nesta serão disponibilizadas informações gerais sobre o curso, material complementar, listas de exercícios, notas de provas e nota final.

2.3 Tópicos do Curso

1. **Conceitos Introdutórios:** dados com resposta categórica e principais distribuições. Estimativa por máxima verossimilhança. Intervalos de confiança para proporções. Testes de hipóteses.
2. **Tabelas de Contingência:** Estrutura, tipos de estudos, modelos amostrais, medidas de independência, diferença entre proporções, Razão de Chances e Risco Relativo. Tabelas tridimensionais e o Método de Cochran-Mantel-Haenszel para independência condicional. Homogeneidade. Testes de tendência em tabelas Bidimensionais. Representação gráfica.
3. **Noções de Modelos Lineares Generalizados:** componentes, GLM para dados binários, GLM para dados de contagens e proporções. Ajuste e Inferência.
4. **Regressão Logística com Respostas Binárias:** modelagem, inferência e testes de hipóteses. Seleção de variáveis e diagnóstico.

5. **Regressão Logística com Respostas Politémicas:** modelagem com respostas nominais e ordinais.
6. **Modelo Log-lineares:** modelagem, ajuste e comparação de modelos.
7. **Modelos para Dados Pareados:** Teste de McNemar para comparação de proporções marginais. Discussão de tópicos.

3. Atividades e Avaliação

A avaliação constará de duas provas com consulta e alguns exercícios a serem realizados em grupo. As provas serão entregues em aula de quinta-feira para entrega no dia seguinte. Exercícios em grupo serão solicitados com uma semana de antecedência.

O aluno que perder uma prova, poderá repor na semana seguinte. Os exercícios não terão reposição.

A Nota Final (NF) será calculada da seguinte forma:

$$NF = \frac{P1 + P2 + EG}{2.8}$$

sendo

$P1$ = Prova 1, $P2$ = Prova 2, EG = Exercícios em Grupo. Cada atividade vale 100 pontos.

3.1 Menção Final

As menções seguem os critérios definidos pela Universidade de Brasília conforme a seguinte escala: $[90, 100]$ = SS, $[70, 90)$ = MS, $[50, 70)$ = MM, $[30, 50)$ = MI, $(0, 30)$ = II, 0 = SR.

Importante: A aferição de frequência será feita através da entrega das atividades de avaliação.

Referências

- [1] AGRESTI, A. *Categorical Data Analysis*, 3ª ed. 2013.
- [2] AGRESTI, A. *An Introduction to Categorical Data Analysis*, 3ª ed. John Wiley & Sons, 2019.
- [3] BILDER, C. R., AND LOUGHIN, T. M. *Analysis of Categorical Data with R*. CRC Press, 2015.
- [4] CHRISTENSEN, R. *Log-Linear Models and Logistic Regression*. Springer, 1997.
- [5] FREEMAN., D. H. *Applied Categorical Data Analysis*. Marcel Dekker, Inc, 1987.
- [6] GIOLO, S. R. *Introdução à Análise de Dados Categóricos com Aplicações*. Blucher - Projeto Fisher, 2017.

- [7] HOSMER, D. W., LEMESHOW, S., AND STURDIVANT, R. X. *Applied Logistic Regression*, 3^a ed. John Wiley & Sons, 2013.
- [8] LACHIN, J. M. *Biostatistical Methods (The Assessment of Relative Risks)*. John Wiley & Sons, 2000.
- [9] STOKES, M. E., DAVIS, C. S., AND KOCH, G. G. *Categorical Data Analysis Using SAS*, third ed. SAS Institute, 2012.

Programa tentativo de aulas: Última atualização 17/08/2020.

Semana	Aula	Dia	Conteúdo
01	01	17/08	Apresentação do curso e Conceitos Introdutórios
	02	19/08	Conceitos Introdutórios
02	03	24/08	Conceitos Introdutórios
	04	26/08	Tabelas de Contingência
03	05	31/08	Tabelas de Contingência
	06	02/09	Tabelas de Contingência
04	-	07/09	Feriado: Independência
	07	09/09	Tabelas de Contingência
05	08	14/09	Noções de Modelos Lineares Generalizados
	09	16/09	Noções de Modelos Lineares Generalizados
06	10	21/09	Noções de Modelos Lineares Generalizados
	11	23/09	Noções de Modelos Lineares Generalizados
07	12	28/09	Regressão Logística com Respostas Binárias
	13	30/09	Regressão Logística com Respostas Binárias
08	14	05/10	Regressão Logística com Respostas Binárias
	15	07/10	Regressão Logística com Respostas Binárias
09	-	12/10	Feriado: Nossa Senhora
	16	14/10	Entrega da Prova 1
10	17	19/10	Regressão Logística com Respostas Politômicas
	18	21/10	Regressão Logística com Respostas Politômicas
11	19	26/10	Regressão Logística com Respostas Politômicas
	20	28/10	Regressão Logística com Respostas Politômicas
12	-	02/11	Feriado: Finados
	21	04/11	Regressão Logística com Respostas Politômicas
13	22	09/11	Modelos Log-lineares
	23	11/11	Modelos Log-lineares
14	24	16/11	Modelos Log-lineares
	25	18/11	Modelos Log-lineares
15	26	23/11	Modelos Log-lineares
	27	25/11	Modelos Log-lineares
16	28	30/11	Modelos para Dados Pareados
	29	02/06	Modelos para Dados Pareados
17	30	07/07	Modelos para Dados Pareados
	31	09/07	Modelos para Dados Pareados
18	32	14/12	Modelos para Dados Pareados
	33	16/12	Entrega da Prova 2

Importante: O programa de aulas é tentativo (inclusive data de provas) e será atualizado sempre que necessário.