

## **Lego III: Métodos Estatísticos para Análise de Dados Categóricos (3 créditos)**

**Prof. Carlos Antonio Costa Ribeiro**

**Horário: Quarta-feira, das 13 às 16 horas**

**Consultas: A combinar com o professor**

---

### **Objetivos**

Os métodos para a análise de dados categóricos ocupam lugar central nas análises estatísticas em sociologia e ciência política. Há uma variedade enorme de técnicas e métodos usados nessas disciplinas, que são baseados nas análises de categorias, como por exemplo, classe social, gênero, partidos políticos, transições educacionais, número de filhos nascidos vivos etc.

Este curso pretende introduzir os alunos nessas técnicas de análises estatísticas de dados categóricos. Espera-se que, ao concluir com sucesso o curso, os alunos sejam capazes de entender métodos básicos e alguns relativamente sofisticados. Embora o curso não seja exaustivo sobre o tema, pretende encorajar os alunos a continuar a estudar métodos estatísticos mais avançados para análise de dados categóricos.

### **Pré-requisitos**

Cursos de introdução a estatística e de análise de regressão linear (Lego I e Lego II). Conhecimento de conceitos básicos de amostragem e inferência; distribuições binomial e normal, e qui-quadrado; técnicas básicas para a análise de tabelas de contingência  $2 \times 2$ . É fundamental o conhecimento de princípios básicos de análise de regressão linear (múltipla).

### **Presença nas aulas**

As aulas e discussões em sala de aula são indispensáveis para o entendimento da matéria do curso. Falta até mesmo a uma única aula pode ser prejudicial para o aprendizado dos tópicos seguintes do curso.

## Exercícios

O entendimento dos tópicos do curso depende em grande medida dos exercícios que os alunos devem fazer fora da sala de aula. A cada semana será distribuído um conjunto de exercícios (contendo de 2 a 4 questões), que deve ser entregue na aula subsequente.

## Programas de computador

Os programas de computador utilizados no curso são: Excel, Lem (programa gratuito) e STATA.

## Trabalho final

Um trabalho final, para ser feito fora da sala de aula, será distribuído na última semana de aula. A nota final será estabelecida a partir desse trabalho e dos exercícios feitos ao longo do curso.

## Consultas

Os alunos são encorajados a discutir os exercícios entre si. O professor estará disponível para consultas em horários determinados de atendimento.

## Textos de Referência

FIENBERG, Stephen E. (1994), *The Analysis of Cross-Classified Categorical Data* (3ª ed.). Cambridge, Mass., MIT Press.

AGRESTI, Alan. (1996), *An Introduction to Categorical Data Analysis*. New York, Wiley Series in Probability and Statistics.

POWERS, Daniel A. e XIE, Yü. (2008), *Statistical Methods for Categorical Data Analysis*. New York, Academic Press.

KNOKE, David e BURKE, Peter J. (1980), *Log-linear Models*. Califórnia, Sage Publications.

HOUT, Michael. (1983), *Mobility Tables*. Califórnia, Sage Publications.

LONG, J. Scott e FREESE, Jeremy (2001), *Regression Models for Categorical Dependent Variables Using Stata*. Texas, Stata Press Corporation.

LONG, J. Scott. (1997), *Regression Models for Categorical and Limited Dependent Variables*. Vol. 7 of *Advanced Quantitative Techniques in the Social Sciences*.

Thousand Oaks, California, Sage Publications.

SILVA, Nelson do Valle. (1988), *Análise de Dados Categóricos*. Rio de Janeiro, Vértice.

GELMAN, Andrew e HILL, Jennifer. (2007), *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York, Cambridge University Press.

LONG, J. Scott. (2009), *The Workflow of Data Analysis Using Stata*. Texas, Stata Press Corporation.

RABE-HESKETH, Sophia e SKRONDAL, Anders. (2008), *Multilevel and Longitudinal Modeling Using Stata*. Texas, Stata Press Corporation.

TREIMAN, Donald J. (2009), *Quantitative Data Analysis: Doing Social Research to Test Ideas*. Califórnia, Jossey-Bass.

EFRON, Bradley e TIBSHIRANI, Robert J. (1993), *An Introduction to the Bootstrap*. Boca Raton, Fl., Chapman & Hall/CRC.

EVERITT, Brian S., LANDAU, Sabine e LEESE, Morven. (2001), *Cluster Analysis*. New York, Oxford University Press/Hodder Arnold Publication.

HAO, Lingxin e NAIMAN, Daniel Q. (2007), *Quantile Regression*. London, Sage.

O'DONNELL, Owen. *et alii* (eds.). (2007), *Analyzing Health Equity Using Household Survey Data*. Washington, DC, The World Bank.

LAREAU, Annette e CONLEY, Dalton. (2008), *Social Class: How Does It Work?*. New York, Russell Sage Foundation.

HENDRICKX, John. (2003), *The Analysis of Religious Assortative Marriage: An Application of Design Techniques for Categorical Models*. West Lafayette, IN, Purdue University Press.

HEDSTRÖM, Peter e BEARMAN, Peter. (2009), *The Oxford Handbook of Analytical Sociology*. Oxford, Oxford University Press.

## **Tópicos:**

### **1 – Revisão dos Modelos de Regressão Linear**

- Modelos de regressão
- Modelos de regressão linear
- Variáveis dependentes categóricas e contínuas

#### **Leitura:**

- Xie e Powels (capítulo 2).

### **2 - Análise de Tabelas de Contingência 2 x 2**

- Modelo de independência
- Medidas de associação
- Estatísticas de ajuste dos modelos
- Conceitos básicos para modelos log-lineares

#### **Leituras:**

- Fienberg (capítulos 1 e 2);
- Knoke e Burke (pp. 7-22, 30-33);
- Agresti (capítulo 2) e Silva (capítulos 2 e 3).

### **3 - Análises de Tabelas de Contingência com 3 Variáveis**

- Independência, associação, e modelos de não-interação
- Algoritmos iterativos de ajuste proporcional (estimativa iterativa dos valores esperados)
- Comparações hierárquicas

#### **Leituras:**

- Fienberg (capítulo 3);
- Agresti (capítulo 3).

### **4 - Dividindo Tabelas de Contingência**

- Modelos de independência em partes das tabelas
- Dividindo o Chi-quadrado

#### **Leitura:**

- Fienberg (capítulo 4).

## 5 - Análises de Tabelas de Contingência Multidimensionais

- Modelos de interação (*higher order*)
- Procedimentos para seleção de modelos
- Comparação hierárquica de modelos interligados

### Leituras:

- Fienberg (capítulo 5);
- Knoke e Burke (pp. 33-47).

## 6 - Análise de Tabelas de Mobilidade

- Mobilidade Total, fluxos de saída e de entrada
- Índice de Dissimilaridade
- Modelo de Mobilidade Quasi-perfeita
- Modelos de distância social
- Modelos Topológicos
- Modelos de associação escalar

### Leituras:

- Hout;
- Powers e Xie (capítulo 4).

## 7 - Modelos Logit e Probit para Variáveis Binárias

- Relação entre o modelo log-lineares e o modelo logit

### Leituras:

- Fienberg (capítulos 6 e 7);
- Knoke e Burke (pp. 24-20);
- Powers e Xie (capítulo 3);
- Agresti (pp. 162-167);
- Long e Freese (capítulo 4).

## 8 - Modelos para Variáveis Dependentes Ordinais

- Modelos Logit e Probit dados ordinais

### Leituras:

- Long e Freese (capítulo 5);
- Powers e Xie (capítulo 6).

## 9 - Modelos para Variáveis Dependentes Nominais

- Modelo Logit Multinomial
- O Modelo Logit Multinomial Básico
- O Modelo Logit Multinomial Condicional

### Leituras:

- Long e Freese, capítulo 6;
- Powers e Xie, capítulo 7.

## 10 - Modelos para Variáveis Dependentes de Contagem

- Regressão de Poisson
- Regressão Binomial negativa

### Leitura:

- Long e Freese (capítulo 7).

## 11 - Modelos Estatísticos para Razões

- Log-rate models
- Discrete time hazard models

### Leitura:

- Powers e Xie (capítulo 5).