



GET00211 - Modelos Lineares 2

Rafael Erbisti

Instituto de Matemática e Estatística
Universidade Federal Fluminense

Aula 01

Organização do Curso

- Professor: Rafael Erbisti
- E-mail: rerbisti@id.uff.br
- Ambiente virtual: google Classroom
- Encontros não presenciais: google meet (link do Classroom)
- Horário de atendimento: quinta-feira, das 15 às 16 horas, com agendamento prévio (sala 505, 5º andar, bloco H)



Avaliações

- Avaliação 1: 02/10
- Avaliação 2: 02/12
- VR: 09/12
- VS: 16/12

Critério de avaliação

$$MF = (0,5 \times \text{Avaliação 1}) + (0,5 \times \text{Avaliação 2})$$

- $MF \geq 6,0 \longrightarrow$ aprovado
- $MF < 4,0 \longrightarrow$ reprovado
- $4,0 \leq MF < 6,0 \longrightarrow$ VS e, neste caso, se $VS \geq 6,0 \longrightarrow$ aprovado



- ① DOBSON, A.J.; E BARNETT, A.G. An introduction to generalized linear models. London: Chapman & Hall, 2018.
- ② CORDEIRO, G. M.; DEMÉTRIO, C. G. B.; E MORAL, R. A. Modelos Lineares Generalizados e Aplicações. São Paulo: Blucher, 2024.
- ③ AGRESTI, A. Foundations of linear and generalized linear models. Wiley, 2015. Wiley Series in Probability and Statistics.

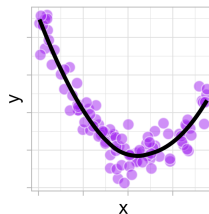
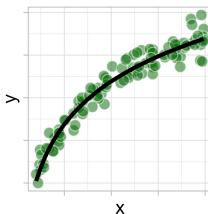
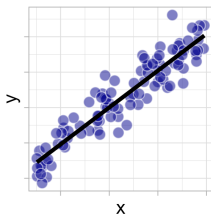


- Uma das preocupações estatísticas ao analisar dados é criar modelos que explicitem estruturas do fenômeno em observação.
- Em muitas situações, duas ou mais variáveis estão relacionadas, surgindo a necessidade de determinar e entender a natureza dessa relação.
- Com isso, um problema comum em Estatística é o estudo da relação entre as variáveis.
- Isto é, procura-se estabelecer uma função entre as variáveis.



Conceitos gerais

- Em geral, a relação não é perfeita: os pontos não se situam exatamente sobre a função que relaciona as variáveis.
- Mesmo se existir uma relação exata entre variáveis, como temperatura e pressão, flutuações em torno da curva aparecerão devido a erros de medição.
- Frequentemente, o tipo de curva a ser ajustada é sugerido por evidência empírica ou por argumentos teóricos.



O modelo estatístico a ser adotado depende de vários fatores, como a natureza das variáveis e a presença ou não de relações lineares entre elas.

Componentes do Modelo

- **Modelo estatístico:** modelo matemático que contém os componentes determinístico e aleatório, utilizados para descrever informações.
- **Sinal:** descrição matemática das principais características de um conjunto de dados; corresponde ao **componente determinístico**, não aleatório.
- **Ruído:** engloba todas as características não explicadas pelo modelo; corresponde ao **componente aleatório**.

A descrição matemática do sinal envolve quantidades que sintetizam a informação, chamadas **parâmetros**.



Conceitos gerais

Passos para formular um modelo estatístico:

- Postular uma equação envolvendo o **sinal** e o **ruído**;
- Associar uma **distribuição de probabilidade** descrevendo a parte aleatória.

Análise de regressão

Metodologia estatística que estuda (modela) a relação entre duas ou mais variáveis.

De forma geral, a análise de regressão passa por duas etapas:

- **Especificação**: descobrir entre os possíveis modelos qual é o mais adequado para o desfecho de interesse.
- **Determinação**: uma vez definido o modelo, estimar os parâmetros da equação.



Modelo de regressão linear

- No modelo de regressão a variável de interesse, variável resposta ou variável dependente é denotada por Y .
- São consideradas estocásticas, variáveis aleatórias.
- Enquanto que as variáveis explicativas, regressoras ou independentes são denotadas por X_1, \dots, X_p .
- Os valores de X_1, \dots, X_p são fixos, isto é, não são variáveis aleatórias.
- Quando existe uma relação linear entre a variável resposta e as regressoras o modelo é denominado **modelo de regressão linear**.



Modelo de regressão linear

- Usualmente, utilizamos modelos lineares da forma:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \sim \text{NMV}(\mathbf{0}, \sigma^2 \mathbf{I})$$

para tratar observações contínuas.

- Neste curso, estudaremos modelos que permitem uma análise semelhante àquela vista em modelos lineares normais.
- Mais especificamente, consideraremos que:
 - ① As variáveis resposta podem seguir uma distribuição diferente da normal.
 - ② A relação entre a variável resposta e as covariáveis (variáveis independentes) não precisa ser linear.
 - ③ Muitas propriedades dos modelos normais são estendidas para uma classe maior de distribuições, a **família exponencial**.



Variáveis qualitativas

- **Classificação ordinal:** existe uma ordenação natural entre as categorias. *Exemplo: jovem, meia idade, idoso; classes de pressão sanguínea (70–90, 91–110, >110mm/Hg).*
- **Classificação nominal:** as categorias são permutáveis e não possuem estrutura de ordenação. *Exemplo: vermelho, verde, amarelo.*

Variáveis quantitativas

- **Variável contínua:** pode assumir infinitos valores em uma escala contínua (sem interrupções). *Exemplo: renda, tempo de vida de um equipamento, peso, altura.*
- **Variável discreta:** assume valores em um conjunto enumerável, geralmente associados a contagens. *Exemplo: casos de dengue, quantidade de indivíduos com determinada característica.*



- Os métodos de análise estatística dependem da escala de medida das variáveis resposta e explicativa.
- A classe dos **Modelos Lineares Generalizados (GLM)** inclui casos em que as variáveis resposta são:
 - Contagens;
 - Proporções ou razões de contagens;
 - Quantidades positivas com distribuições assimétricas;
 - Múltiplas categorias de resposta.



Etapas do processo de modelagem

O processo de modelagem neste curso envolve os seguintes passos:

- 1 **Especificação do modelo:** equação que liga variável resposta e explicativa + distribuição da variável resposta.
- 2 **Estimação dos parâmetros** do modelo.
- 3 **Verificação da adequabilidade do modelo:** o modelo se ajusta bem aos dados?
- 4 **Interpretação dos resultados.**
- 5 **Tomada de decisão.**

Capítulo 2 — A. Dobson

- Exemplos de especificação de modelos para problemas aplicados
- Pontos importantes na modelagem estatística
- Notação e representação matricial de modelos a partir de exemplos

