

Estatística Básica

Prof. Fernando de Souza Bastos
fernando.bastos@ufv.br

Instituto de Ciências Exatas e Tecnológicas
Universidade Federal de Viçosa
Campus UFV - Florestal

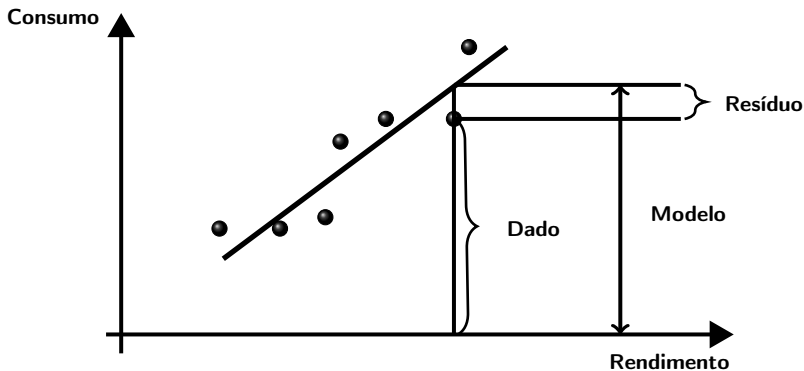


Sumário

- 1 Modelos
- 2 Técnicas Computacionais
- 3 Métodos Gráficos
- 4 Conjuntos de Dados

Fundamentalmente, quando se procede a uma análise de dados, busca-se alguma forma de regularidade ou padrão ou, ainda, um modelo, presente nas observações.

Figura: Relação entre consumo e rendimento.



Fonte: (MORETTIN; BUSSAB, 2017)

Imagine que estejamos estudando a relação entre rendimentos e gastos de consumo de um conjunto de indivíduos. Podemos obter um gráfico como o da Figura 1. O que se espera, intuitivamente, é que os gastos de um indivíduo estejam diretamente relacionados com os seus rendimentos, de modo que é razoável supor uma “relação linear” entre essas duas quantidades. Os pontos da Figura 1 não estão todos, evidentemente, sobre uma reta, essa seria o nosso padrão ou modelo. A diferença entre os dados e o modelo constitui os resíduos.

Modelos

De modo esquemático:

$$\text{Dados} = \text{Modelo} + \text{Resíduos}, \quad \text{ou, ainda,} \quad D = M + R \quad (1)$$

Modelos

De modo esquemático:

$$\text{Dados} = \text{Modelo} + \text{Resíduos}, \quad \text{ou, ainda,} \quad D = M + R \quad (1)$$

A parte M é também chamada parte suave (ou regular ou, ainda, previsível) dos dados, enquanto R é a parte aleatória.

Modelos

De modo esquemático:

$$\text{Dados} = \text{Modelo} + \text{Resíduos}, \quad \text{ou, ainda,} \quad D = M + R \quad (1)$$

A parte M é também chamada parte suave (ou regular ou, ainda, previsível) dos dados, enquanto R é a parte aleatória.

A parte R é tão importante quanto M , e a análise dos resíduos constitui uma parte fundamental de todo trabalho estatístico. Basicamente, são os resíduos que nos dizem se o modelo é adequado ou não para representar os dados.

Uma análise exploratória de dados busca, essencialmente, fornecer informações para estabelecer (1).

O desenvolvimento rápido e constante na área de computação foi acompanhado pela introdução de novas técnicas de análise de dados, notadamente de métodos gráficos e de métodos chamados de computação intensiva.

Técnicas Computacionais

Para a implementação dessas técnicas, foram desenvolvidos programas estatísticos, atualmente usados em larga escala tanto no meio acadêmico como em indústrias, bancos, órgãos de governo e etc. Um dos mais utilizados na atualidade é o software R.



Técnicas Computacionais

Por outro lado, os programas podem exigir maior ou menor experiência computacional dos usuários. Alguns operam com menus, e seu uso é mais simples. Outros requerem maior familiaridade com o computador e são baseados em linguagens próprias.

Técnicas Computacionais

Por outro lado, os programas podem exigir maior ou menor experiência computacional dos usuários. Alguns operam com menus, e seu uso é mais simples. Outros requerem maior familiaridade com o computador e são baseados em linguagens próprias.

Fiquem atentos, procurem aprender programação, não tenha medo do computador! Estamos a bordo de uma revolução tecnológica que já está transformando a forma como vivemos, trabalhamos e nos relacionamos e à medida que todas as empresas se tornam digitais, a demanda por profissionais de tecnologia aumenta mais rápido do que a disponibilidade de mão de obra qualificada!

Técnicas Computacionais

Além dos pacotes estatísticos, há outros pacotes de grande utilidade para realizar tarefas matemáticas e estatísticas. Dentre estes, mencionamos o Mathematica, o Maple, o Gauss, o Geogebra, o excel e o MatLab. Existe também o Wolfram Alpha e o Symbolab que são mecanismos de conhecimento computacional que funcionam online. Além disso, nesse curso, iremos utilizar o \LaTeX como editor de texto.



O Software R

Vamos utilizar no curso, preferencialmente, o software R, que pode ser obtido livremente no Comprehensive R Archive Network (CRAN). Como o curso é baseado no livro de Estatística Básica do Pedro Morettin e Wilton Bussab, há um repositório com vários scripts do livro nos endereços: www.ime.usp.br/~pam/EstBas.html e rpubs.com./EstatBasica.

Os métodos gráficos têm encontrado um uso cada vez maior devido ao seu forte apelo visual. Normalmente, é mais fácil para qualquer pessoa entender a mensagem de um gráfico do que aquela embutida em tabelas ou sumários numéricos.

Os gráficos são utilizados para:

- buscar padrões e relações;

Os gráficos são utilizados para:

- buscar padrões e relações;
- confirmar (ou não) certas expectativas que se tinha sobre os dados;

Os gráficos são utilizados para:

- buscar padrões e relações;
- confirmar (ou não) certas expectativas que se tinha sobre os dados;
- descobrir novos fenômenos;

Os gráficos são utilizados para:

- buscar padrões e relações;
- confirmar (ou não) certas expectativas que se tinha sobre os dados;
- descobrir novos fenômenos;
- confirmar (ou não) suposições feitas sobre os procedimentos estatísticos usados; e

Os gráficos são utilizados para:

- buscar padrões e relações;
- confirmar (ou não) certas expectativas que se tinha sobre os dados;
- descobrir novos fenômenos;
- confirmar (ou não) suposições feitas sobre os procedimentos estatísticos usados; e
- apresentar resultados de modo mais rápido e fácil.

Podemos usar métodos gráficos para plotar os dados originais ou outros dados derivados deles. Por exemplo, a investigação da relação entre as variáveis da Figura (1) pode ser feita por meio daquele diagrama de dispersão. Mas podemos também “ajustar” uma reta aos dados, calcular o desvio (resíduo) para cada observação e fazer um novo gráfico, de consumo contra resíduos, para avaliar a qualidade do ajuste.

Conjuntos de Dados

Na página da disciplina (<https://maf105.github.io/>) aparecem alguns conjuntos de dados que serão utilizados nos exemplos ou nos exercícios propostos. Aconselho a todos a reproduzir os exemplos, usando esses dados, bem como resolver os problemas, pois somente a efetiva manipulação de dados pode levar a um bom entendimento das técnicas apresentadas.

Os conjuntos de dados apresentados provêm de diferentes fontes, que são mencionadas em cada conjunto e depois explicitadas nas referências. Usaremos para as análises estatísticas o software R, calculadora científica e planilhas do Excel.

O novo petróleo!

Em 2006, Clive Robert Humby (Matemático britânico e empresário milionário da área de ciência de dados) cunhou a frase “Os dados são o novo petróleo” (ARTHUR, 2013). Michael Palmer expandiu a citação de Humby dizendo, como o petróleo, os dados são “valiosos, mas se não forem refinados, não podem realmente ser usados. Petróleo deve ser transformado em gás, plástico, produtos químicos, etc. para criar uma entidade valiosa que impulse atividades lucrativas; então, os dados devem ser desmembrados e analisados para que tenham valor” (PALMER, 2006; HAUPT, 2016; FIRICAN, 2019).

Mãos à obra!!!

Referências I

ARTHUR, Charles. Tech giants may be huge, but nothing matches big data. Página visitada em 11/06/2021. 2013.

Disponível em: <[https:](https://www.theguardian.com/technology/2013/aug/23/tech-giants-data)

[//www.theguardian.com/technology/2013/aug/23/tech-giants-data](https://www.theguardian.com/technology/2013/aug/23/tech-giants-data)>.

FIRICAN, por George. How Data Is (And Isn't) Like Oil. Página visitada em 11/06/2021. 2019. Disponível em:

<<https://tdwi.org/articles/2019/04/22/data-all-how-data-is-like-oil.aspx>>.

HAUPT, Michael. “Data is the New Oil” — A Ludicrous Proposition. Página visitada em 11/06/2021. 2016. Disponível em:

<<https://medium.com/project-2030/data-is-the-new-oil-a-ludicrous-proposition-1d91bba4f294>>.

Referências II

MORETTIN, Pedro Alberto; BUSSAB, Wilton Oliveira. **Estatística básica**. [S.l.]: Saraiva Educação, 2017.

PALMER, Michael. **Data is the New Oil**. Página visitada em 11/06/2021. 2006. Disponível em: <https://ana.blogs.com/maestros/2006/11/data_is_the_new.html>.