

SME0803

Análise Exploratória de Dados

1º/2010

Mário de Castro

mcastro@icmc.usp.br

6-204, ramal 9567 (Lab. de Estatística)

3-160, ramal 8174 (sala)

CoteiaWIKI: <http://wiki.icmc.usp.br/index.php/SME-803>

Programa resumido

1. Tipos de variáveis e organização de banco de dados.
2. Representação de variáveis qualitativas: tabelas de frequências, gráficos de barras e de setores.
3. Medidas descritivas de dados quantitativos: posição, dispersão, assimetria e concentração.
4. Representação de variáveis quantitativas: tabelas de frequências, histograma, gráfico de caixas, ramo-e-folhas e gráfico de pontos.
5. Tabelas de contingência e medidas de associação entre variáveis qualitativas.
6. Associação entre variáveis quantitativas e qualitativas.
7. Representação gráfica de dados multidimensionais.
8. Análise de agrupamentos.

Avaliação

1ª prova: 5/5/2010. 2ª prova: 30/6/2010.

Recuperação: data a combinar.

$$P = (P_1 + P_2) / 2.$$

Normas de recuperação: regras do SME (Júpiter).

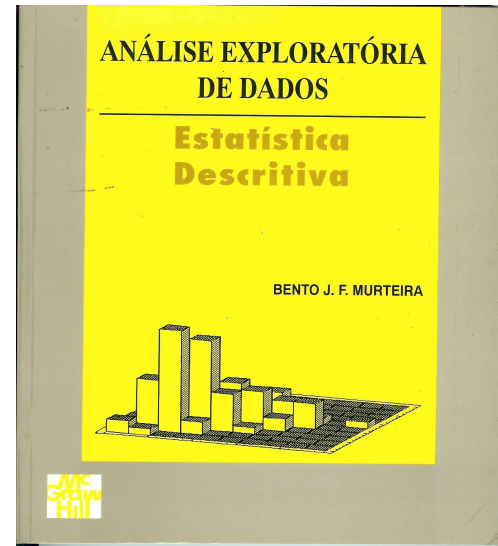
Prova substitutiva? Não.

Trabalhos para “ajudar na nota”? Não.

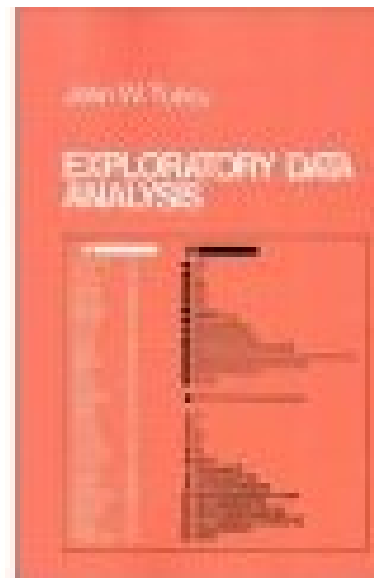
Listas de exercícios? Não.

Bibliografia

1. Murteira, B. J. F., Análise Exploratória de Dados. McGraw-Hill, Lisboa, 1993.



2. Tukey, J. W., Exploratory Data Analysis. Addison-Wesley, Reading-MA, 1977.

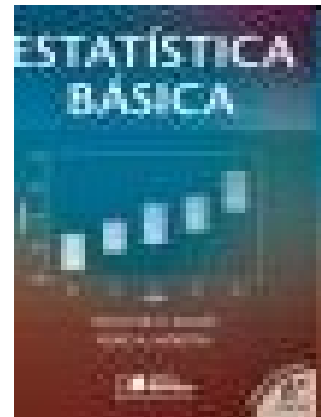
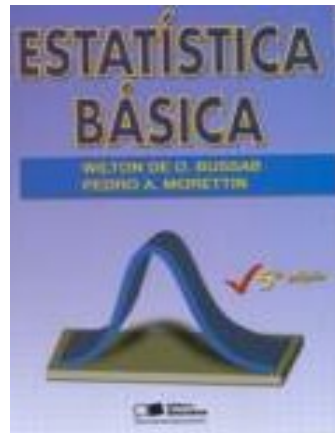


Bibliografia

3. Magalhães, M.N.; Lima, A.C.P. Noções de Probabilidade e Estatística, 6ª ed., EDUSP, São Paulo, 2004.



4. Bussab, W.O.; Morettin, P.A. Estatística Básica, 6ª ed., Saraiva, São Paulo, 2009.

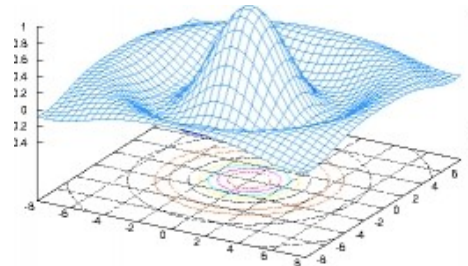
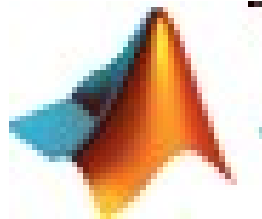
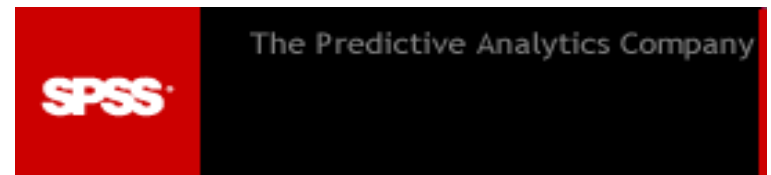


Apoio computacional

Excel.

R (www.R-project.org).

Minitab, S-PLUS, SAS, SPSS, Statistica, MatLab, Octave, Scilab, etc.



O que é Estatística ?

Para muitos, Estatística não passa de conjuntos de tabelas. Os estatísticos são pessoas que coletam esses dados.

A Estatística originou-se com a coleta de dados e a construção de tabelas para os governos.

A situação evoluiu e a coleta de dados representa somente um dos aspectos da Estatística.

Estatística é um conjunto de técnicas utilizadas para a coleta, organização, processamento e análise de dados.

A Estatística é uma ciência baseada na **Teoria da Probabilidade**, cujo objetivo principal é nos auxiliar a tomar decisões ou tirar conclusões em situações de incerteza, a partir de dados.

História da Estatística

Desenvolvimento da Estatística, principalmente no Brasil:

<http://www.redeabe.org.br/historia.htm>

Memória, J. M. P., Breve História da Estatística. EMBRAPA, Brasília, 2004.

http://www.embrapa.br/publicacoes/tecnico/folderTextoDiscussao/arquivos-pdf/Miolo_21.pdf/view

Materials for the History of Statistics:

<http://www.york.ac.uk/depts/maths/histstat/>

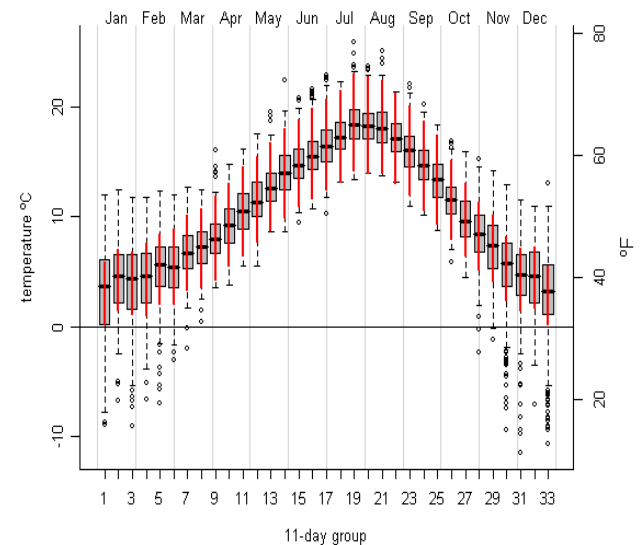
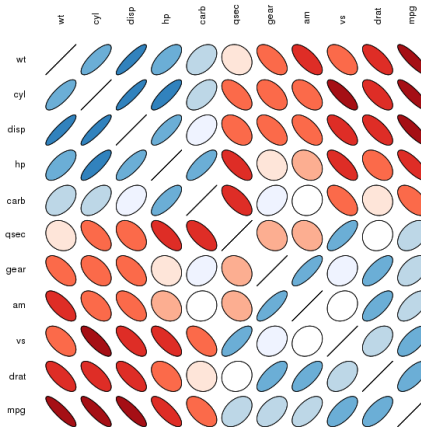
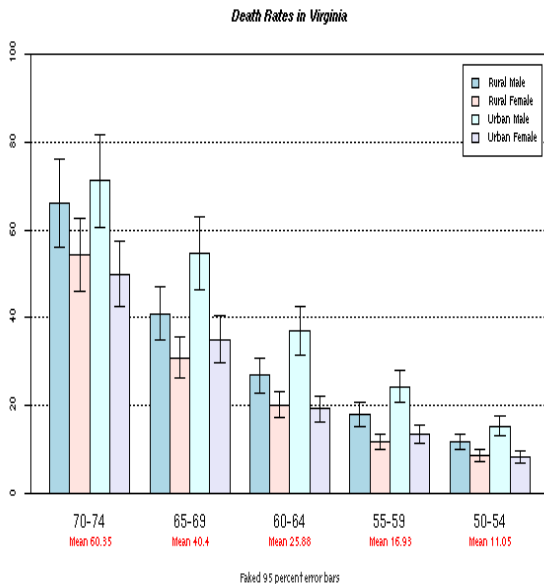
Análise exploratória de dados

Estatística Descritiva e Análise Exploratória

Etapas iniciais de muitos trabalhos estatísticos. Utilizadas para descrever e resumir os dados.

A disponibilidade de uma grande quantidade de dados e de métodos computacionais muito eficientes revigorou estas áreas da Estatística.

A análise exploratória de dados utiliza técnicas gráficas.



Análise exploratória de dados



John W. Tukey (1915 – 2000)

<http://www.york.ac.uk/depts/maths/histstat/lifework.htm#t>

“Exploratory data analysis is detective work – numerical detective work – or counting detective work – or graphical detective work”.

Muitas vezes o trabalho estatístico não se limita à exploração e descrição. É necessário **estender** as conclusões para além dos dados coletados.

Análise confirmatória é parte importante de muitos trabalhos. Técnicas de **inferência estatística** são utilizadas.

“Exploratory and confirmatory can – and should – proceed side by side”.