



SME0803

ANÁLISE EXPLORATÓRIA DE DADOS

1º/2012

**Cibele Russo**

cibele@icmc.usp.br

<http://www.icmc.usp.br/~cibele>

Sala 3-162, ramal 6618

**CoteiaWIKI:**

[http://wiki.icmc.usp.br/index.php/SME-803\(cibele\)](http://wiki.icmc.usp.br/index.php/SME-803(cibele))



## Aulas

Quintas-feiras das 21h às 22h40

Sextas-feiras das 19h às 20h40

## Horário de atendimento

Quintas-feiras das 16h às 17h ou mediante agendamento por e-mail.



# PRINCIPAIS OBJETIVOS DO CURSO

- Apresentar as principais técnicas estatísticas para descrição e exploração de dados
- Conhecer os principais pacotes estatísticos
- Interpretar resultados de análises descritivas e exploratórias



# 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.



# CRITÉRIO DE AVALIAÇÃO

$$\text{Média Final MF} = \frac{4 \text{ P1} + 4 \text{ P2} + \text{NE} + \text{NT}}{10}$$

P1: Prova 1: **27/04/2012** das 19h às 20h40

P2: Prova 2: **22/06/2012** das 19h às 20h40

NE: Nota de exercícios propostos em sala ou para casa

NT: Nota de trabalho prático (individual)

Prova Substitutiva: Somente com justificativa oficial, por exemplo atestado médico emitido pelo UBAS



# CRITÉRIO DE AVALIAÇÃO

Se  $3,0 \leq MF < 5,0$  e Frequência  $\geq 70\%$

O aluno **pode** fazer REC

Nova média final **NMF** e Nota da REC **NR**

$$NMF = \begin{cases} 5; & \text{se } 5 \leq NR \leq (10 - MF); \text{ ou} \\ (MF + NR)/2 & \text{se } NR > (10 - MF); \text{ ou} \\ MF & \text{se } NR < 5 \end{cases}$$

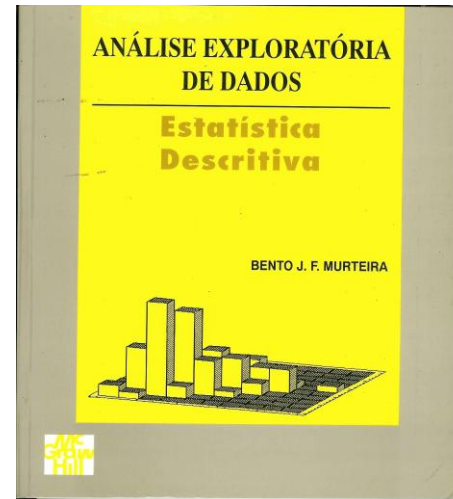
Data da REC: **12/07/2012** (Quinta-feira)

Preferencialmente às 19h

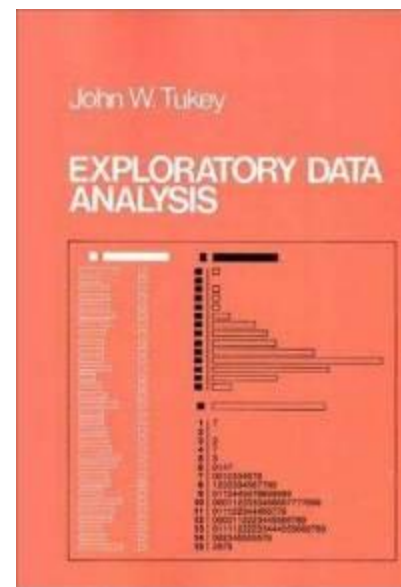


# BIBLIOGRAFIA PRINCIPAL

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

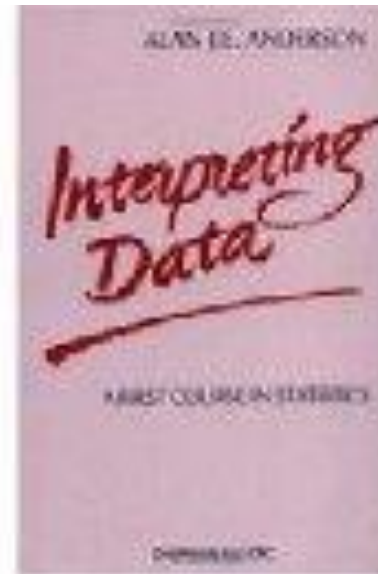


2. Tukey, J. W., Exploratory Data Analysis. Addison-Wesley, Reading-MA, 1977.  
(Disponível na biblioteca)



# BIBLIOGRAFIA COMPLEMENTAR

1. Anderson, A. J. B.,  
Interpreting Data: a First  
Course in Statistics.  
London:  
Chapman&Hall,1989.  
(Disponível na biblioteca)



2. Levine, D. M., Berenson, M. L., Stephan, D. Estatística: Teoria e Aplicações usando Microsoft Excel em Português. Rio de Janeiro:LTC, 1998.  
(Disponível na biblioteca)



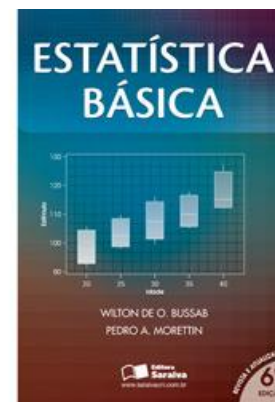
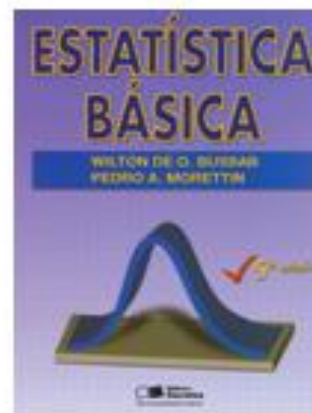
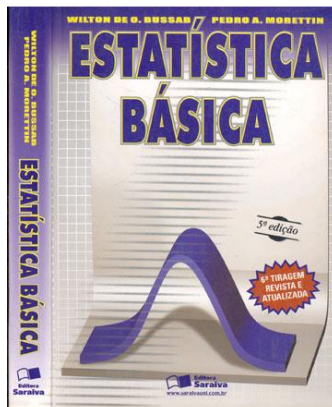


# BIBLIOGRAFIA COMPLEMENTAR

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



4. Bussab, W.O.; Morettin, P.A. Estatística Básica, 6ª ed., Saraiva, São Paulo, 2009. (Disponível na biblioteca)



# ALGUMAS REGRAS

- A USP exige no mínimo 70% de presença nas aulas
- A docente poderá fazer a chamada oralmente em qualquer instante do horário da aula, mesmo se os alunos já tiverem assinado a lista de presença, e eventuais ausências não serão abonadas
- Questionamentos a respeito da matéria poderão ser feitos a qualquer momento
- Conversas paralelas entre alunos não serão toleradas – os outros alunos têm direito a assistir a aula em silêncio

# ALGUMAS REGRAS

- Os alunos devem desligar seus telefones celulares antes a aula
- Não é permitido o uso de notebook durante as aulas, exceto em possíveis aulas práticas
- Ao enviar e-mail para a docente, incluir no assunto o código **SME0803**
- Eventuais dúvidas, consultar Normas da USP  
<http://www.usp.br/leginf/>



# ALGUMAS REGRAS – NAS PROVAS

- Será permitido o uso de **calculadora própria**
- Será permitido o uso de uma folha A4 com **anotações próprias** (não será permitido xerox)
- Pode ser solicitado apresentar o cartão USP para a realização da prova
- Fraude em prova implicará a reprovação direta do(s) envolvido(s)
- Falsidade ideológica na realização da prova (ou assinatura de presença) poderá implicar abertura de processo disciplinar



# IMPORTANTE

Canal oficial de comunicação

**E-mail cadastrado no JupiterWeb**

Ao enviar e-mail para a docente, escrever  
no Assunto **SME0803**

Os alunos devem trazer **calculadora** para  
todas as aulas

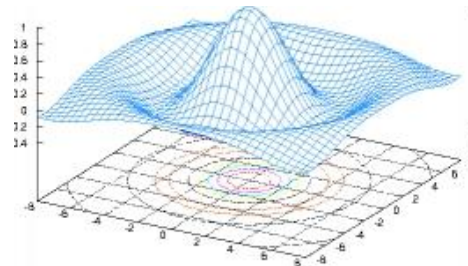
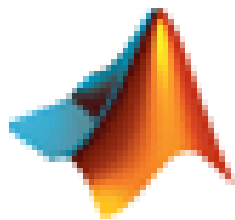
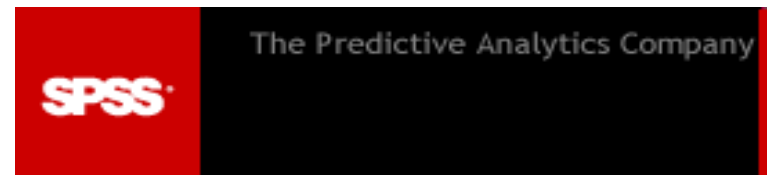


# APOIO COMPUTACIONAL

Excel.

R ([www.R-project.org](http://www.R-project.org)).

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



# EXERCÍCIO



Baixar e instalar o R a partir de [www.R-project.org](http://www.R-project.org)



## 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.





## EM OUTRAS PALAVRAS...

**Estatística** é um conjunto de técnicas e métodos de pesquisa que envolve o planejamento do experimento a ser realizado, a coleta qualificada dos dados, a inferência, o processamento, a análise e a disseminação das informações  
(ENCE - Escola Nacional de Ciências Estatísticas)



# 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](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/math/histstat/>

História da Estatística (por Prof. Gauss Cordeiro):

[http://www.des.uem.br/uploads/arquivos\\_professor/1125192908.pdf](http://www.des.uem.br/uploads/arquivos_professor/1125192908.pdf)



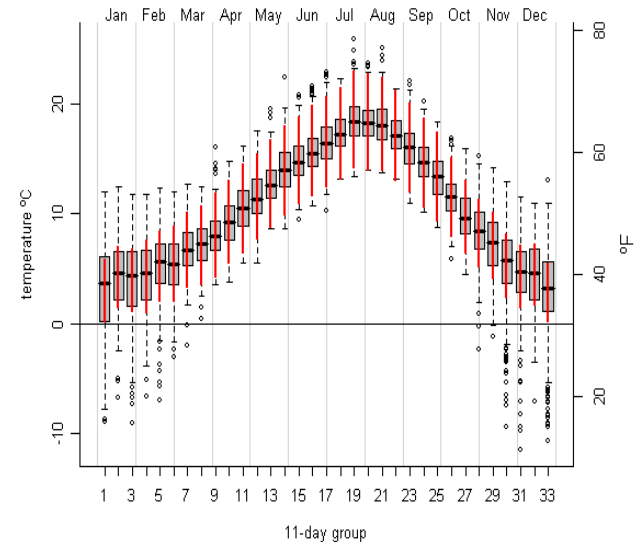
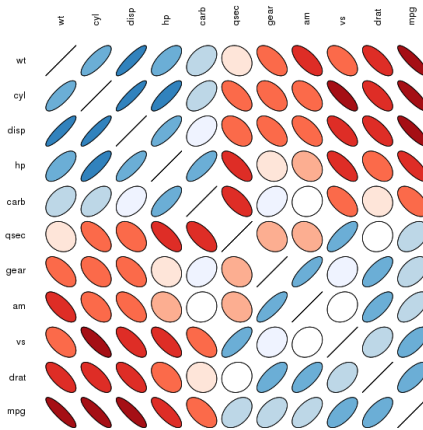
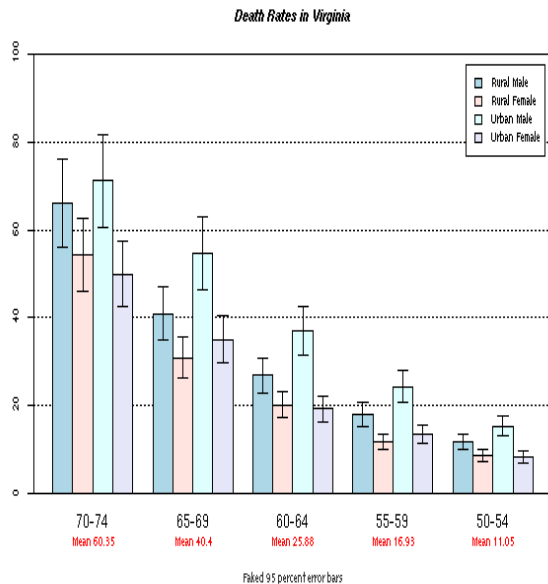
# ANÁLISE EXPLORATÓRIA DE DADOS

## Estatística Descritiva e Análise Exploratória

Etapas iniciais da **maioria** dos trabalhos estatísticos. Utilizadas para descrever, resumir e 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 ...

It can never be the whole story, but nothing else can serve as the foundation stone - as the first step.



## ANÁLISE CONFIRMATÓRIA DE DADOS

Apesar de ser uma etapa muito importante, o trabalho estatístico vai além da exploração e descrição de dados. 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”.



# VÍDEO: 200 PAÍSES, 200 ANOS, 4 MINUTOS

<http://www.gapminder.org/videos/200-years-that-changed-the-world-bbc/>

