



# SME0820

## Análise de Regressão

### 1º/2014

#### **Mário de Castro**

mcastro@icmc.usp.br

Sala 4-240, telefone 3373- 6631

#### **Cibele Russo**

cibele@icmc.usp.br

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

Sala 3-113, telefone 3373-6606

**Site da disciplina: CoteiaWIKI:**

<http://wiki.icmc.usp.br/index.php/SME0820>

# Objetivos do curso

Introduzir as principais técnicas de  
Análise de Regressão

Trabalhar em aplicações de Análise de  
Regressão com e sem o auxílio de  
pacotes estatísticos

# Programa resumido da disciplina

1. Regressão Linear Simples.
2. Estudo de adequação do modelo regressão linear simples.
3. Regressão linear múltipla.
4. Estudo de adequação do modelo regressão linear múltipla.
5. Regressão polinomial.
6. Regressão com variáveis Indicadoras.
7. Métodos de Diagnóstico.
8. Construção de modelos.
9. Validação do modelo de regressão
10. Tópicos adicionais.

# Critério de Avaliação

05/05/2014: Prova 1 (P<sub>1</sub>)

25/06/2014: Prova 2 (P<sub>2</sub>)

11 e 16/06/2014: Seminário (S)

Individual **ou**  
em dupla

$$\text{Média Final} = \text{MF} = \frac{4 P_1 + 5 P_2 + S}{10}$$

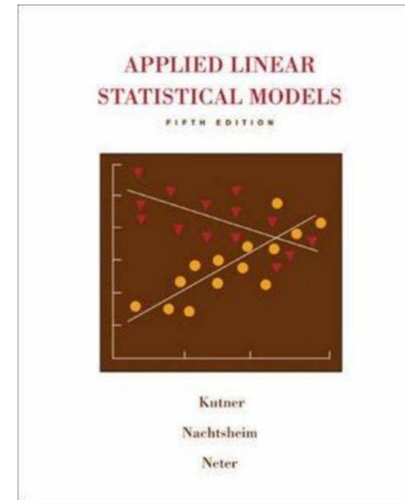
**Não haverá prova substitutiva.**

Critério de Recuperação: ver JúpiterWeb.

# Bibliografia principal

KUTNER, M. H., NACHTSHEIM, C. J., NETER, J. & LI, W. *Applied Linear Statistical Models*. 5<sup>th</sup> edition. McGraw-Hill Irwin, New York. 2005.

(disponível na biblioteca)



DRAPER, N. R.; SMITH, H. *Applied Regression Analysis*, 3<sup>rd</sup> edition, John Wiley & Sons. New York, 1998.

(disponível na biblioteca)



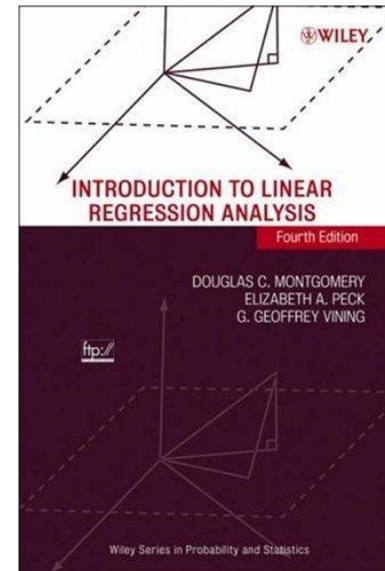
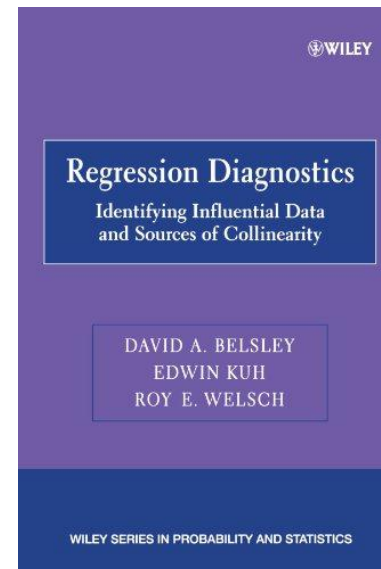
# Bibliografia complementar

Montgomery, D. C., Peck, E. A. and Vining, G. G., Introduction to Linear Regression Analysis, 4th ed. New York. John Wiley, 2006.

(disponível na biblioteca)

Belsley, D. A., Kuh, E., Welsch, R. E. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity, John Wiley, New York, 2004.

(disponível na biblioteca)

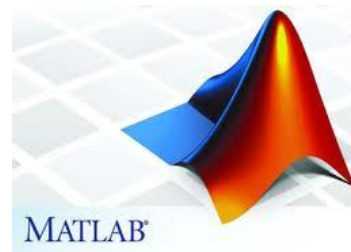


# Apoio computacional

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

Excel, Portal Action ([www.portalaction.com.br/](http://www.portalaction.com.br/))

Minitab, S-PLUS, SAS, SPSS, Statistica, Stata, Matlab, Scilab, entre outros.



# Importante: E-mail

- E-mail é um canal oficial de comunicação da docente com os alunos.
- Checar e-mail **cadastrado no JupiterWeb** regularmente (pelo menos uma vez ao dia).
- Ao enviar e-mail para a docente, identificar o código da disciplina no assunto SME0820.
- É importante como forma de **documentar ocorrências**  
Por exemplo: envio de listas, ausências por problemas de saúde, avisos sobre listas de exercícios, aulas.

# Algumas regras – em aulas

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

# Algumas regras de conduta

- Os alunos devem desligar seus telefones celulares durante aula e prova.
- Não é permitido o uso de notebook durante as aulas, exceto em possíveis aulas práticas.
- Qualquer fraude em prova implicará a reprovação direta do(s) envolvido(s).
- Falsidade ideológica na realização da prova (ou assinatura de presença) implicará abertura de processo disciplinar.

# Algumas regras – em provas

Em cada prova será permitido o uso de

- Lápis, caneta, lapiseira, borracha, régua
- Calculadora **própria**
- Uma folha A4 com anotações próprias (não será permitido xerox) identificada com o nome do aluno
- Tabelas de distribuições de probabilidade

**Não é permitido o uso de telefone celular**

# Orientações de estudo

- Manter **frequência de estudo** do início ao final do semestre. O ideal é estudar o conteúdo de cada aula logo após ter sido ministrada, isso facilita o entendimento das aulas seguintes e diminui a quantidade de conteúdos para os dias anteriores à prova.
- Estudar primeiramente o conteúdo dado em aula, buscando apoio em referências bibliográficas, em seguida resolver exemplos e exercícios. Não deixar para a véspera da prova.
- Buscar referências bibliográficas logo no início do semestre, dando preferência às referências principais e complementares adotadas.

# O que é Estatística?

**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/maths/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)

Arquivo resumido com alguns tópicos destacados

[http://wiki.icmc.usp.br/images/2/20/Hist%C3%B3ria\\_da\\_Estat%C3%ADstica\\_-\\_Resumida.PDF](http://wiki.icmc.usp.br/images/2/20/Hist%C3%B3ria_da_Estat%C3%ADstica_-_Resumida.PDF)

# Algumas das grandes áreas da Estatística

Probabilidades e Processos Estocásticos

Planejamento de Experimentos

Amostragem

Modelos de Regressão

Séries Temporais

Métodos Bayesianos

Análise de Sobrevida

# Análise de Regressão

O primeiro uso da palavra **regressão** surgiu no contexto estatístico por meio do **antropologista e meteorologista Sir Francis Galton** por volta de 1885, embora Legendre e Gauss já tivessem publicado trabalhos utilizando o método dos mínimos quadrados no início do século XIX.

Originalmente o termo utilizado por Galton foi **reversão**.

Os primeiros trabalhos de Galton buscavam **explicar a altura dos filhos a partir da altura de seus pais**.

O termo “Análise de Correlação” também é atribuído a Galton.

# O que é Análise de Regressão?

Análise de regressão é uma  
**técnica estatística**  
para investigar e modelar  
a relação entre variáveis.

(Montgomery, Peck and Vining, 2006, pág. 1)

# Exemplo

Deseja-se investigar a relação entre

- o tempo que um indivíduo leva para reagir a um certo estímulo (em décimos de segundos) e
- características como sexo, idade (em anos), acuidade visual (em %).

(Bussab, W., Análise de variância e regressão, 1986, Atual Editora)

# Exemplo

| Indivíduo | tempo | sexo | Idade | Acuidade_<br>visual |
|-----------|-------|------|-------|---------------------|
| 1         | 96    | M    | 20    | 90                  |
| 2         | 92    | F    | 20    | 100                 |
| 3         | 106   | M    | 20    | 80                  |
| 4         | 100   | F    | 20    | 90                  |
| 5         | 98    | F    | 25    | 100                 |
| 6         | 104   | M    | 25    | 90                  |
| 7         | 110   | M    | 25    | 80                  |
| 8         | 101   | F    | 25    | 90                  |
| 9         | 116   | F    | 30    | 70                  |
| 10        | 106   | M    | 30    | 90                  |
| 11        | 109   | M    | 30    | 90                  |
| 12        | 100   | F    | 30    | 80                  |
| 13        | 112   | F    | 35    | 90                  |
| 14        | 105   | F    | 35    | 80                  |
| 15        | 118   | M    | 35    | 70                  |
| 16        | 108   | M    | 35    | 90                  |
| 17        | 113   | F    | 40    | 90                  |
| 18        | 112   | F    | 40    | 90                  |
| 19        | 127   | M    | 40    | 60                  |
| 20        | 117   | M    | 40    | 80                  |

# Exemplo

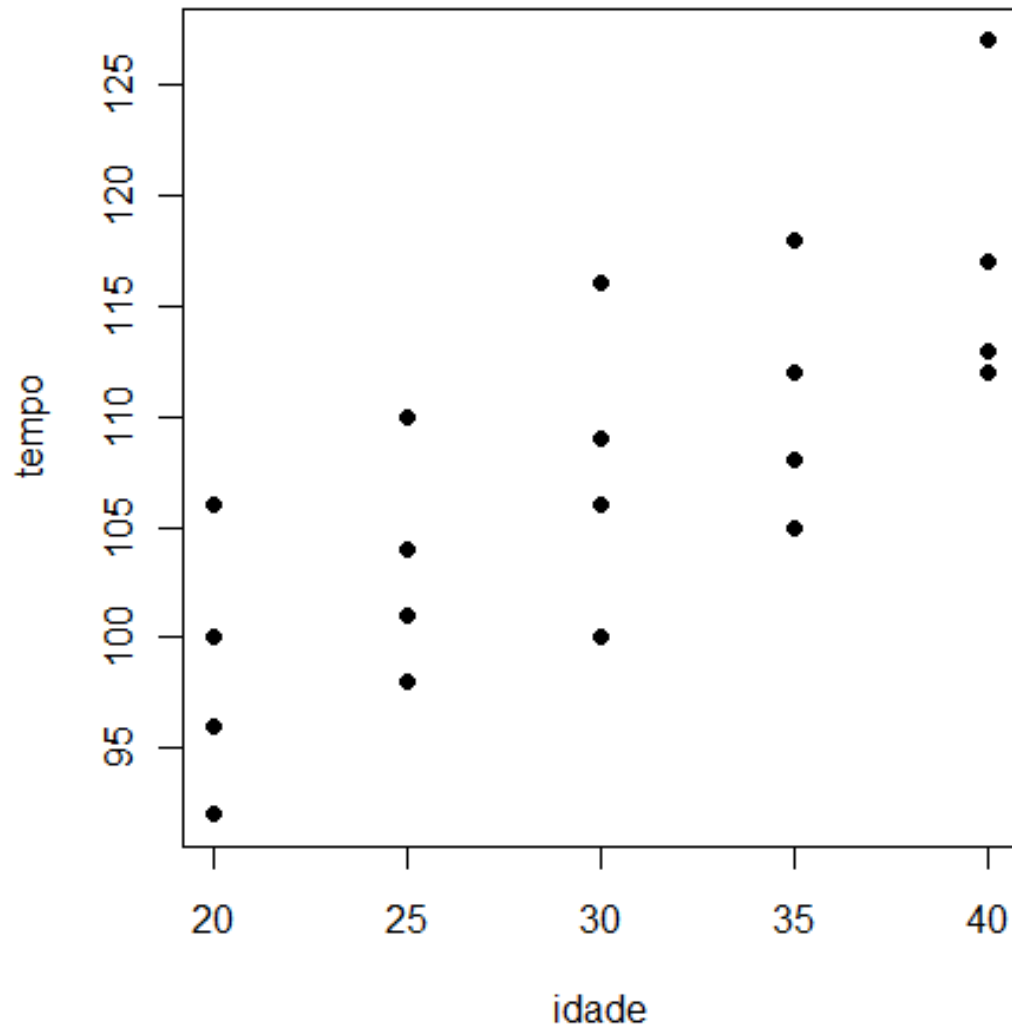
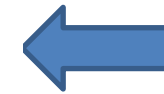
Inicialmente vamos estudar a relação entre

- Y: tempo
- X: idade

sem considerar as outras variáveis.

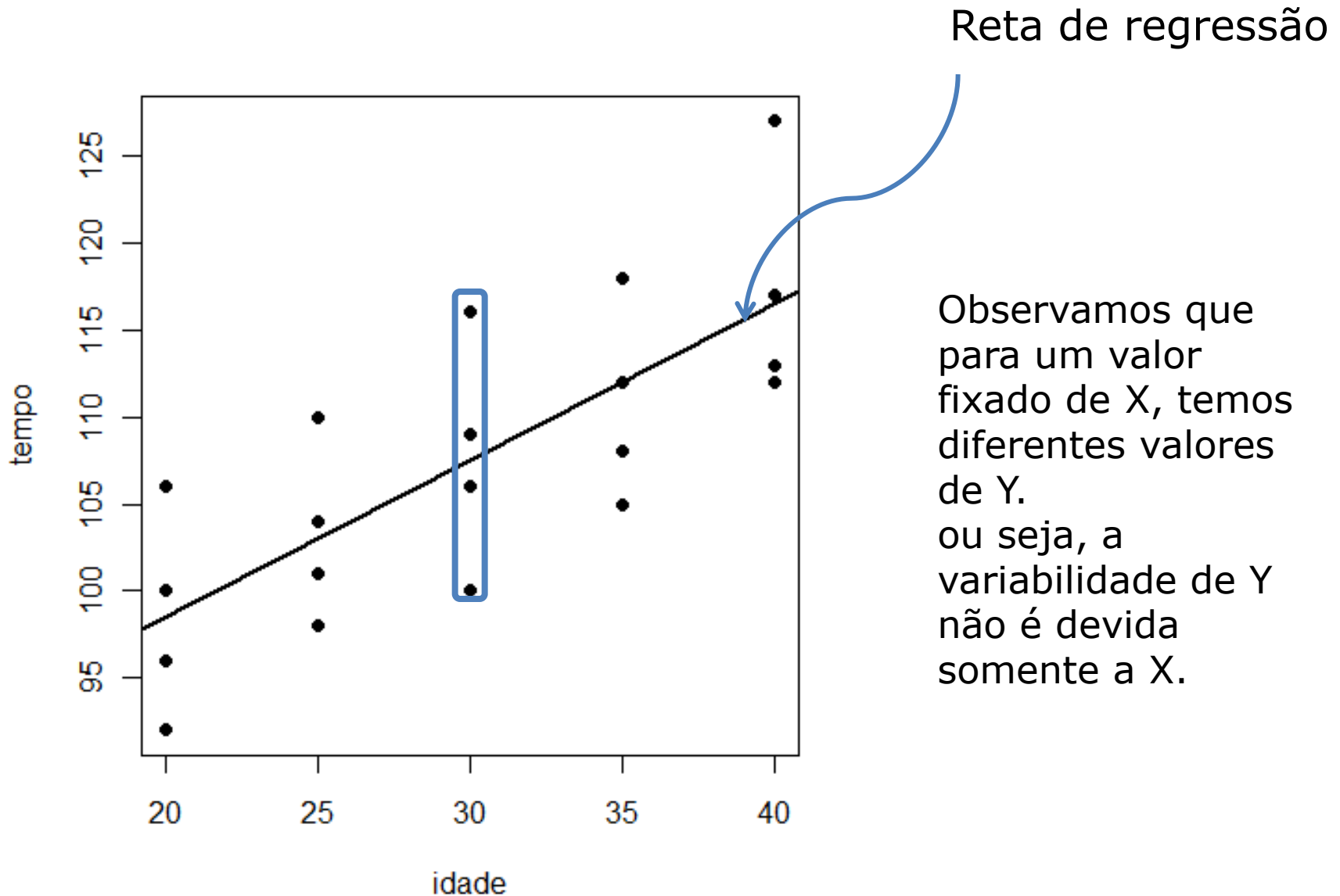
Inicialmente podemos construir o gráfico de dispersão entre a idade (X) e o tempo (Y).

# Exemplo



Pode-se notar uma tendência de aumento no tempo de reação com o aumento da idade.

# Exemplo



# Exemplo

Algumas das questões que desejamos responder:

- É possível explicar a variabilidade de  $Y$  utilizando a variabilidade de  $X$ ?
- É razoável assumir que a relação entre  $X$  e  $Y$  é linear?
- É razoável assumir que o erro cometido ao explicar variabilidade de  $Y$  utilizando a variabilidade de  $X$  tem uma certa distribuição de probabilidade?
- Existem observações atípicas ou que influenciam de forma desproporcional o modelo ajustado?
- No caso de  $p$  covariáveis  $X_1, \dots, X_p$ , quantas e quais variáveis devemos utilizar para explicar a variabilidade de  $Y$ ?