

Modelos de regressão para dados correlacionados

Cibele Russo

cibele@icmc.usp.br

ICMC USP

Mini-curso oferecido no
Workshop on Probabilistic and Statistical Methods

28 a 30 de janeiro de 2013

Conteúdo do curso

- Modelos de regressão
- Modelos lineares com efeitos mistos
- Modelo marginal \times modelo hierárquico
- Análise de diagnóstico
- Modelos não lineares com efeitos mistos

Principais referências bibliográficas

- Pinheiro and Bates (2000) 'Mixed-effects Models in S and S-PLUS', Springer.
- Searle, S. R., Casella, G. McCulloch, C. E. 2006, Variance Components. Wiley Series in Probability and Statistics.
- Verbeke G. and Molenberghs G. (2000) 'Linear mixed models for longitudinal data,' Springer Series in Statistics, Springer-Verlag, New-York.
- Mini-curso de Geert Molenberghs 'Models for Longitudinal and Incomplete Data' na ESALQ USP em 2010.
- Mini-curso de Dimitris Rizopoulos 'An Introduction to Joint Models for Longitudinal & Survival Data, with Applications in R' (Erasmus University Medical Center) no 27th International Workshop on Statistical Modelling, Prague.
- ... outras

Conteúdo da aula

- Introdução ao R
- Motivação
- Modelos de regressão
- Dados correlacionados
 - ▶ Dados longitudinais
 - ▶ Medidas repetidas

Objetivos da aula

Após esta aula, espera-se que os participantes sejam capazes de

- Entender a utilidade de modelos de regressão
- Identificar a presença de correlação entre observações
- Acessar conjuntos de dados disponíveis no pacote R

Apoio computacional: R



O que é R?

R é uma linguagem e um ambiente para o desenvolvimento computacional e gráfico de ferramentas estatísticas.

É um software livre e de código aberto.

Download em www.r-project.org

Apoio computacional: R

Instruções:

Serão disponibilizados comandos em R, indicados pelo símbolo ">", que deverá ser excluído no momento da execução dos comandos.

Exercício: Instalar os pacotes que serão utilizados no curso.

Abra o R e execute o comando abaixo (apenas deixe de selecionar ">")

```
> install.packages(c("stats", "nlme", "Hmisc", "lattice"))
```

Obs: Embora esteja no comando acima, "stats" já vem instalada R.



Certifique-se que os comandos foram passados corretamente para o R. Se preferir, utilize o

bloco de notas ou um editor para R, como o Tinn-R.

Apoio computacional: R

Algumas funções(bibliotecas) úteis em R para modelos de regressão e modelos de regressão com efeitos mistos:

- `lm` (stats): ajusta modelos lineares
- `nls` (stats): ajusta modelos não lineares
- `lme` (nlme): ajusta modelos lineares com efeitos mistos
- `nlme` (nlme): ajusta modelos não lineares com efeitos mistos

Para mais informações, utilize comandos do tipo:

```
> ?lm
```

```
> ?lme
```

Modelos de regressão

Análise de regressão é uma ferramenta estatística que investiga a relação entre duas ou mais variáveis.

Modelo de regressão é um modelo matemático que envolve relações entre as variáveis, parâmetros de interesse e componentes aleatórias.

Muitas vezes, existe correlação entre medidas de uma certa variável. Essa correlação deve ser levada em conta pelo modelo de regressão.

Motivação: Dados correlacionados (Machines)

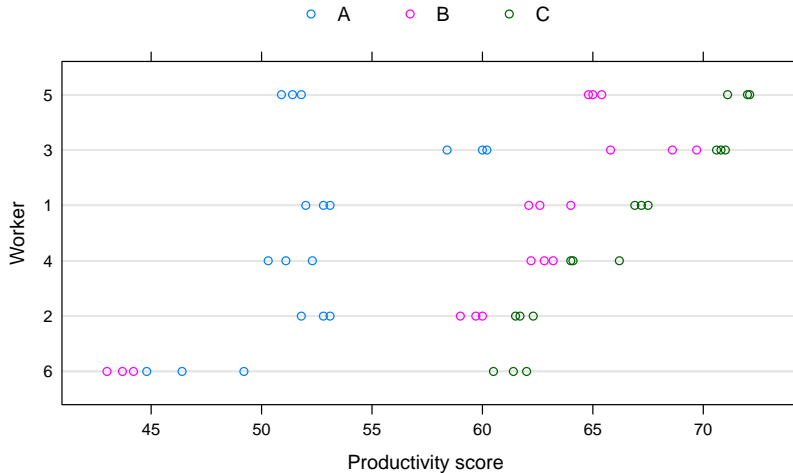
Com o objetivo de verificar a produtividade obtida com a utilização de três tipos de máquinas em um processo industrial, um experimento foi realizado com seis trabalhadores, escolhidos ao acaso entre os funcionários de uma fábrica.

Cada funcionário utilizou cada máquina três vezes (réplicas), e foi obtido um índice de produtividade que levava em conta o número e a qualidade dos componentes produzidos.

Esse conjunto de dados está disponível no pacote R com o nome `Machines{nlme}`.

(Milliken, G. A. and Johnson, D. E. (1992), *Analysis of Messy Data, Volume I: Designed Experiments*, Chapman and Hall, London.)

Motivação: Dados de produtividade



Motivação: Dados de produtividade industrial

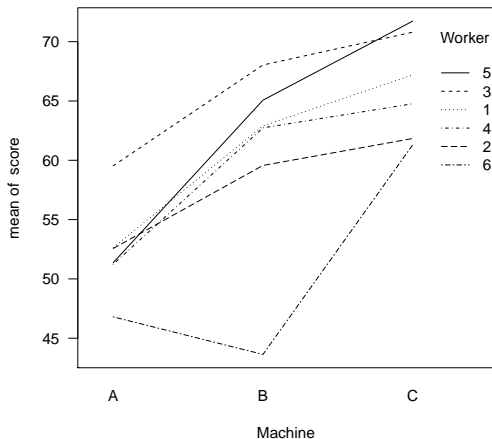


Gráfico de interação score \times máquina

Como acessar os dados no R: Machines{nlme}

```
> library(nlme)
> data(Machines)
> names(Machines)
> plot(Machines)
```

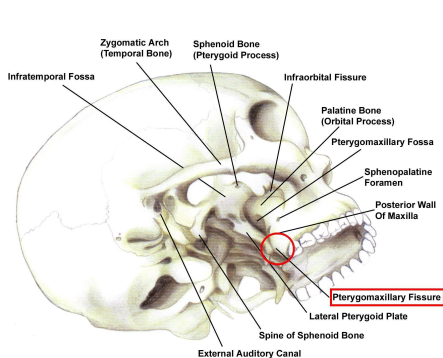
 Certifique-se que os comandos foram passados corretamente para o R.

Motivação: Orthodont (dados ortodônticos)

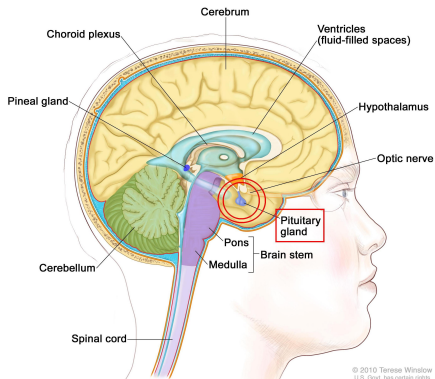
Com o objetivo de estudar a variabilidade da distância do centro da pituitária à fissura pterio maxilar em crianças, um experimento contou com 27 crianças (16 meninos e 11 meninas) de idades entre 8 e 14 anos. A cada dois anos foi feita uma medição de tal distância.

(Potthoff, R. F. and Roy, S. N. (1964), "A generalized multivariate analysis of variance model useful especially for growth curve problems", *Biometrika*, 51, 313-326.)

Motivação: Orthodont (dados ortodônticos)



Osteology of the base of the skull and the pterygomaxillary fossa.

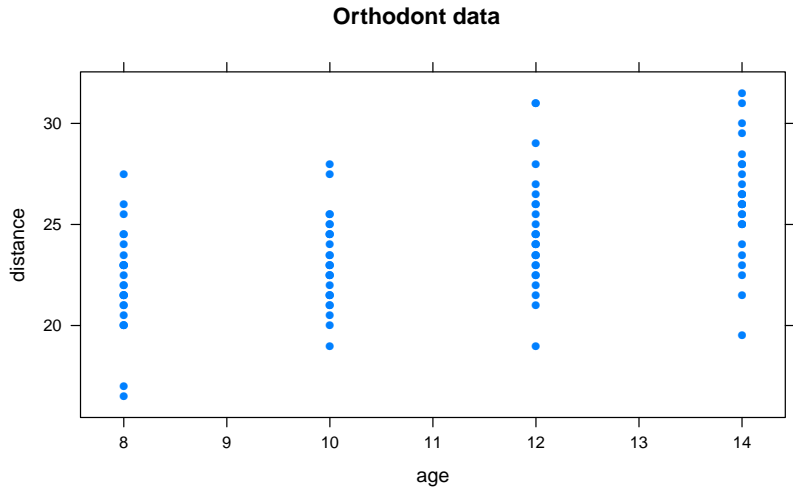


© 2010 Terese Winslow
U.S. Govt. has certain rights

Fonte das figuras: <http://diseases.blogcu.com/skull-base-anatomy/6229186>

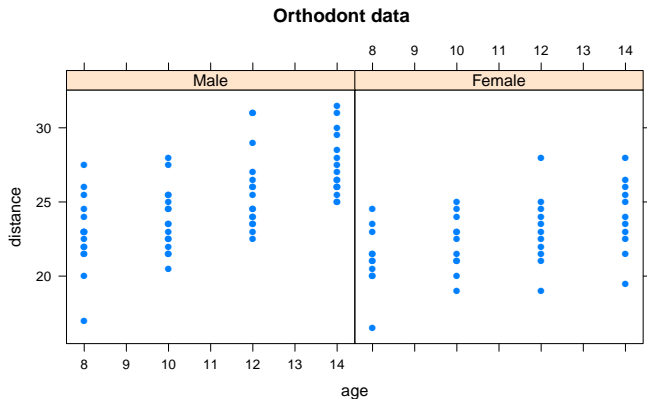
<http://www.uchospitals.edu/online-library/content=CDR579645>

Motivação: Orthodont (dados ortodônticos)



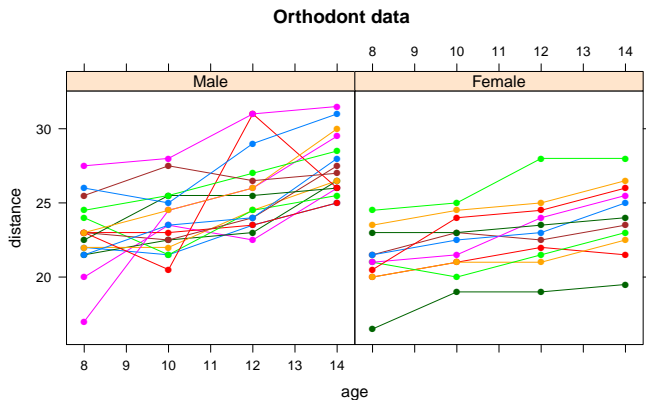
Motivação: Orthodont (dados ortodônticos)

Estudar a relação entre a **distância** do centro da pituitária à fissura pterio maxilar em crianças, a **idade** e o **gênero**.



Motivação: Orthodont (dados ortodônticos)

Estudar a relação entre a **distância** do centro da pituitária à fissura perio maxilar em crianças, a **idade** e o **gênero** levando em consideração a correlação entre as observações.



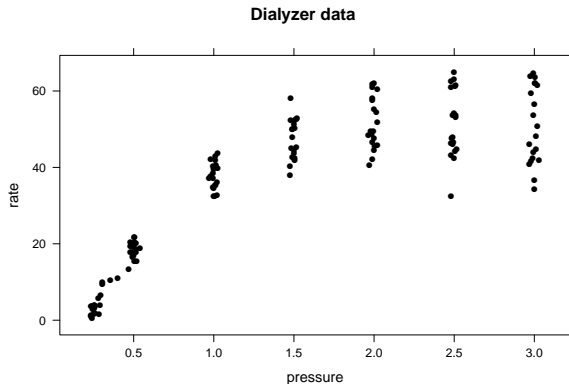
Como acessar os dados no R: Orthodont {nlme}

```
> library(nlme)
> data(Orthodont)
> names(Orthodont)
> plot(Orthodont)
```

Motivação: Dados correlacionados

Exemplo: Dados de hemodiálise (Vonesh & Carter, 1992)

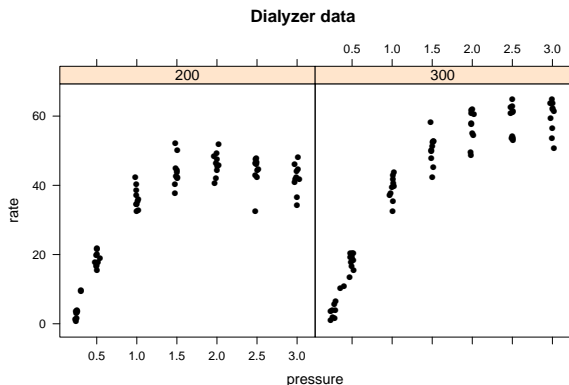
Estudar a relação entre a taxa de ultrafiltração (UFR) e a pressão da transmembrana (TMP) em pacientes em hemodiálise levando em consideração a presença de grupos.



Motivação: Dados correlacionados

Exemplo: Dados de hemodiálise (Vonesh & Carter, 1992)

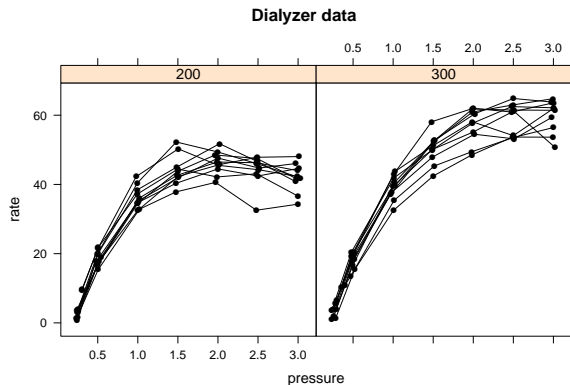
Estudar a relação entre a taxa de ultrafiltração (UFR) e a pressão da transmembrana (TMP) em pacientes em hemodiálise.



Motivação: Dados correlacionados

Exemplo: Dados de hemodiálise (Vonesh & Carter, 1992)

Estudar a relação entre a taxa de ultrafiltração (UFR) e a pressão da transmembrana (TMP) em pacientes em hemodiálise levando em consideração a correlação entre as observações.



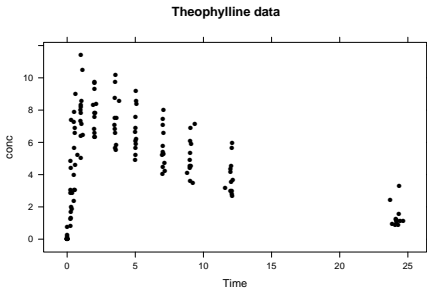
Como acessar os dados no R: Dialyzer

```
> library(nlme)
> data(Dialyzer)
> names(Dialyzer)
> plot(Dialyzer)
```

Motivação: Dados correlacionados

Exemplo: Dados farmacocinéticos de theophylline

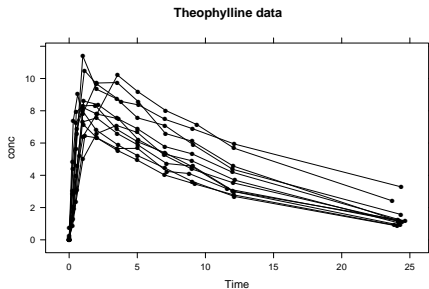
A concentração da substância anti asmática theophylline no sangue é medida após a administração oral da substância.



Motivação: Dados correlacionados

Exemplo: Dados farmacocinéticos de theophylline

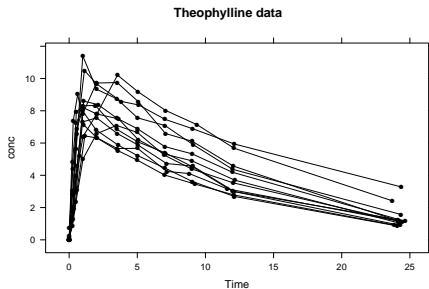
A concentração da substância anti asmática theophylline no sangue é medida após a administração oral da substância levando em consideração a correlação existente entre as observações.



Motivação: Dados correlacionados

Exemplo: Dados farmacocinéticos de theophylline

A concentração da substância anti asmática theophylline no sangue é medida após a administração oral da substância levando em consideração a correlação existente entre as observações.



Como acessar os dados no R: Theophylline

```
> library(nlme)
> data(Theoph)
> names(Theoph)
> plot(Theoph)
```

Próxima aula

Modelos de regressão