

labestData: Biblioteca de dados para aprendizado de Estatística

Prof. Walmes Zeviani

Programa de Educação Tutorial
PET Estatística

pet.estadistica.ufpr@gmail.com

Universidade Federal do Paraná

O que é o *labestData*?

- ▶ Pacote para o R com mais de 450 conjuntos de dados documentados.
- ▶ 19 obras nacionais completas incluídas no pacote.
- ▶ Branch master: 0.0-17.458 de 2016-06-27.
- ▶ Branch devel: 0.1-1.462 de 2016-08-29.



Por que desenvolver pacote com conjuntos de dados?

O problema

- ▶ Analisar de dados é uma arte dominada com o tempo.
- ▶ Praticar é fundamental.
- ▶ Digitar os dados não é estimulante.

Os benefícios

- ▶ Dados acessíveis e documentados → mais prática.
- ▶ Dados de livros: documentação + didático + com análise.
- ▶ Vários livros do mesmo assunto: diversidade de contexto.

Objetivos

1. Contribuir a comunidade por reunir, organizar, manter e disponibilizar conjuntos de dados para o ensino-apredizado de Estatística.
2. Capacitar os membros do PET Estatística a desenvolver pacotes R e trabalhar com versionamento Git em equipe.

Obras completas no pacote - fase I



Figura 1: As 18 obras completas contidas no pacote *labesData* que foram incorporadas na fase I do projeto.

Obras em desenvolvimento no pacote - fase II



Figura 2: As obras que estão sendo adicionadas ao pacote *labesData*. Algumas já estão disponíveis, porém nos ramos que serão incorporados ao master.

Organização dos conjuntos de dados

Nomeação

BanzattoQd3.4.1
PaulaEg1.12.2
BarbinEx8
PimentelPg269

Documentação

Contexto e descrição de todas as variáveis.

Exemplos

Análise descritiva dos dados.

Classificação

Keywords para os temas:

DIC	PSS	RL
DBC	GE	RM
DQL	BAF	RP
FAT2	AnaFat	contagem
FAT3	AnaDisc	binomial
FATADI	AnaClust	sensorial
FRAC	ACP	
PS	ACC	

Interface Shiny

`library(labestData)`

`labestDataView()`



labestData: Biblioteca de dados para aprendizado de Estatística
PET-Estatística UFPR - Versão 0.1.1.462

Escolha a(s) obra(s) Keyword(s) Dados disponíveis Download: Execução R:

Escolha uma obra Todas Escolha um dataset ▼  Stop

 

O **labestData** é um projeto coletivo do **PET-Estatística UFPR** que visa contribuir com o Departamento de Estatística, e a comunidade científica em geral, por reunir, organizar, manter e disponibilizar conjuntos de dados na forma de um pacote R de tal forma que possam ser usados para o ensino de Estatística. Todo o projeto é desenvolvido sob versionamento **Git** e mantido nos nos serviços de hospedagem remota **GitLab do C3SL** e **GitHub**.

Figura 3: Tela inicial da aplicação Shiny no pacote *labesData*.

labestData: Biblioteca de dados para aprendizado de Estatística

PET-Estatística UFPR - Versão 0.1.1.462

Escolha a(s) obra(s)

Escolha uma obra

Todas
Banzatto
Barbin
Charnet
Costa
Demetrio
Dias
Diniz

Keyword(s)

Todas

AAS
ACC
Afd
agrupamento
amostra
AnaClust
AnaComp
AnaDias

Dados disponíveis

Escolha um dataset

BanzattoQd1.2.3
BanzattoQd3.2.1
BanzattoQd3.4.1
BanzattoQd3.6.1
BanzattoQd3.7.1
BanzattoQd4.5.2
BanzattoQd4.7.1
BanzattoQd5.2.1

Download:



Execução R:

Stop

PET-Estatística UFPR que visa contribuir com o conhecimento científico em geral, por reunir, em pacotes de dados na forma de um pacote R de tal maneira que o usuário de Estatística. Todo o projeto é mantido nos serviços de hospedagem

Figura 4: Listas para selecionar as tabelas de dados de acordo com a obra e tema.

Diâmetro à Altura do Peito de *Eucalyptus saligna*

Description

Valores de diâmetro à altura do peito (DAP, cm) de uma amostra aleatória de árvores de *Eucalyptus saligna* de um povoamento com 15 anos de idade.

Format

Um vetor numérico com 20 elementos.

Source

BANZATTO; KRONKA (2013), Quadro 1.2.3, pág. 3.

Examples

```
data(BanzattoQd1.2.3)

hist(BanzattoQd1.2.3, prob = TRUE,
     xlab = expression(Diâmetro-à-altura-do-peito-(cm)),
     ylab = "Densidade", main = NULL)
lines(density(BanzattoQd1.2.3), lwd = 2)
rug(BanzattoQd1.2.3)

boxplot(BanzattoQd1.2.3,
        ylab = expression(Diâmetro-à-altura-do-peito-(cm)))

mean(BanzattoQd1.2.3)
sd(BanzattoQd1.2.3)
fivenum(BanzattoQd1.2.3)
```

Figura 5: Exemplo da documentação de um conjunto de dados dentro do pacote.

Como instalar o *labestData*?

```
library(devtools)
# Do GitLab.
url <- "https://gitlab.c3sl.ufpr.br/pet-estatistica/labestData.git"
install_git(url = url, branch = "master")
# Do GitHub.
install_github(repo = "pet-estatistica/labestData",
               branch = "master")
```

Detalhes mínimos do desenvolvimento

- ▶ Desenvolvimento em 13 pessoas na fase I.
- ▶ Padronização da escrita/estilo de código: [Google R Style Guide](#) → [H. Wickham](#) → pacote **lintr**.
- ▶ Adoção de um *workflow Git*: milestones, branches, integração contínua, revisão, correção, merge, etc.
- ▶ Ciclo semanal de desenvolvimento com reuniões Scrum.

Considerações finais

- ▶ Usem o *labestData*.
- ▶ Para sugestões, colaboração ou relato de bugs:
<https://github.com/pet-estatistica/labestData>.

Considerações finais

- ▶ Usem o *labestData*.
- ▶ Para sugestões, colaboração ou relato de bugs:
<https://github.com/pet-estatistica/labestData>.

Obrigado

