Package ‘aplore3’

October 20, 2016

Title Datasets from Hosmer, Lemeshow and Sturdivant, “Applied Logistic Regression” (3rd Ed., 2013)

Version 0.9


URL https://github.com/lbraglia/aplore3

BugReports https://github.com/lbraglia/aplore3/issues

Depends R (>= 3.1.1)

License GPL-3

LazyData true

VignetteBuilder knitr

Suggests knitr, MASS, vcdExtra, nnet, survival, pROC

RoxygenNote 5.0.1

NeedsCompilation no

Author Luca Braglia [aut, cre]

Maintainer Luca Braglia <lbraglia@gmail.com>

Repository CRAN

Date/Publication 2016-10-20 09:50:53

R topics documented:
aplore3 ................................................................. 2
aps ............................................................... 3
burn1000 .......................................................... 4
burn13m .......................................................... 5
burn_eval_1 ...................................................... 6
burn_eval_2 ...................................................... 6
chdage .......................................................... 7
glow11m .......................................................... 8

Description

This package is an unofficial companion to the textbook "Applied Logistic Regression" by D.W. Hosmer, S. Lemeshow and R.X. Sturdivant (3rd ed., 2013).

Details

It includes all the datasets used in the book, both for easy reproducibility and algorithms benchmarking purposes.

Some analysis proposed in the text are reproduced in the examples, in order to provide data testing and code demos at the same time.

The vignette includes all the examples (with graphics too); therefore is organized per-dataset.

Datasets and variables have lower-case name with respect to the original sources. Categorical data were packaged as factor.

Regarding data coding, help pages list the internal/factor representation of the data (eg 1: No, 2: Yes), not the original one (eg 0: No, 1: Yes). This is intended to allow easier/safer recoding based on as.integer, especially for multinomial variables.

Source

Description

aps dataset.

Usage

aps

Format

A data.frame with 508 rows and 11 variables:

- **id**: Identification Code (1 - 508)
- **place**: Placement (1: Outpatient, 2: Day Treatment, 3: Intermediate Residential, 4: Residential)
- **place3**: Placement Combined (1: Outpatient or Day Treatment, 2: Intermediate Residential, 3: Residential)
- **age**: Age at Admission (Years)
- **race**: Race (1: White, 2: Non-white)
- **gender**: Gender (1: Female, 2: Male)
- **neuro**: Neuropsychiatric Disturbance (1: None, 2: Mild, 3: Moderate, 4: Severe)
- **emot**: Emotional Disturbance (1: Not Severe, 2: Severe)
- **danger**: Danger to Others (1: Unlikely, 2: Possible, 3: Probable, 4: Likely)
- **elope**: Elopement Risk (1: No Risk, 2: At Risk)
- **los**: Length of Hospitalization (Days)
- **behav**: Behavioral Symptoms Score (0 - 9)
- **custd**: State Custody (1: No, 2: Yes)
- **viol**: History of Violence (1: No, 2: Yes)

Source


Examples

```r
head(aps, n = 10)
summary(aps)
```

```r
## Table 8.2 p. 274
library(nnet)
```
modt8.2 <- multinom(place3 ~ viol, data = eps)
summary(modt8.2)
exp(coef(modt8.2)[, "violYes"])
t(exp(confint(modt8.2)["violYes", ]))
## To test differences between b_2 and b_1 we need the estimated variance
## covariance matrix for the fitted model (Table 8.3 p. 274).
v cov(modt8.2) # 'raw'
## To have exactly the same output as the text we need to rearrange just a
## minimum
VarCovM <- vcov(modt8.2)[c(2, 1, 4, 3), c(2, 1, 4, 3)]
VarCovM[upper.tri(VarCovM)] <- NA
VarCovM
## Testing against null model.
modt8.2Null <- multinom(place3 ~ 1, data = eps)
anova(modt8.2, modt8.2Null, test = "Chisq")

---

### burn1000

**burn1000 data**

**Description**

burn1000 dataset.

**Usage**

burn1000

**Format**

A data.frame with 1000 rows and 9 variables:

- **id** Identification code (1 - 1000)
- **facility** Burn facility (1 - 40)
- **death** Hospital discharge status (1: Alive, 2: Dead)
- **age** Age at admission (Years)
- **gender** Gender (1: Female, 2: Male)
- **race** Race (1: Non-White, 2: White)
- **tbsa** Total burn surface area (0 - 100%)
- **inh_inj** Burn involved inhalation injury (1: No, 2: Yes)
- **flame** Flame involved in burn injury (1: No, 2: Yes)

**Source**

Examples

```r
head(burn1000, n = 10)
summary(burn1000)

## Table 3.15 p. 80
summary(mod3.15 <- glm(death ~ tbsa + inh_inj + age + gender + flame + race,
                        family = binomial, data = burn1000 ))
```

Description

`burn13m` dataset.

Usage

`burn13m`

Format

A data.frame with 388 rows and 11 variables: the covariate are the same as those from `burn1000` with the addition of

- **pair** Pair Identification Code (1-119)
- **pairid** Subject Identification Code within pair (1-4)

Source


Examples

```r
head(burn13m, n = 10)
summary(burn13m)
```
burn_eval_1  BURN_EVAL_1 data

Description
burn_eval_1 dataset.

Usage
burn_eval_1

Format
A data.frame with 500 rows and 9 variables: the covariate are the same as those from burn1000.

Source

Examples
head(burn_eval_1, n = 10)
summary(burn_eval_1)

burn_eval_2  BURN_EVAL_2 data

Description
burn_eval_2 dataset.

Usage
burn_eval_2

Format
A data.frame with 500 rows and 9 variables: the covariate are the same as those from burn1000.

Source

Examples
head(burn_eval_2, n = 10)
summary(burn_eval_2)
**CHDAGE data**

**Description**

chdage dataset.

**Usage**

chdage

**Format**

A data.frame with 100 rows and 4 variables:

- **id**: Identification code (1 - 100)
- **age**: Age (Years)
- **chd**: Presence of CHD (1: No, 2: Yes)

**Source**


**Examples**

```r
head(chdage, n = 10)
summary(chdage)

## Figure 1.1 p. 5
plot(as.integer(chd) - 1 ~ age,
    pch = 20,
    main = "Figure 1.1 p. 5",
    ylab = "Coronary heart disease",
    xlab = "Age (years)",
    data = chdage)

## Table 1.2
with(chdage, addmargins(table(agegrp)))
with(chdage, addmargins(table(agegrp, chd)))
(Means <- with(chdage, tapply(as.integer(chd) - 1, list(agegrp), mean)))

## Figure 1.2 p. 6
midPoints <- c(24.5, seq(32, 57, 5), 64.5)
plot(midPoints, Means, pch = 20,
     ylab = "Coronary heart disease (mean)",
     xlab = "Age (years)", ylim = 0:1,
     main = "Figure 1.2 p. 6")
```
lines(midPoints, Means)

## Table 1.3
summary( mod1.3 <- glm( chd ~ age, family = binomial, data = chdage ))

## Table 1.4
vcov(mod1.3)

## Computing OddsRatio and confidence intervals for age ...
exp(coef(mod1.3))[-1]
exp(confint(mod1.3))[-1, ,]

---

glow11m    GLOW11M data

Description
glow11m dataset.

Usage
glow11m

Format
A data.frame with 238 rows and 16 variables: the covariate are the same as those from glow500 with the addition of
	pair  Pair Identification Code (1-119)

Source

Examples
head(glow11m, n = 10)
summary(glow11m)

## Table 7.2 p. 252
library(survival)
mod7.2 <- clogit(as.numeric(fracture) ~ height + weight + bmi +
priorfrac + premeno + momfrac + armassist + raterisk +
strata(pair), data = glow11m)
summary(mod7.2)
**Description**

glow500 dataset.

**Usage**

glow500

**Format**

A data.frame with 500 rows and 15 variables:

- **sub_id** Identification Code (1 - n)
- **site_id** Study Site (1 - 6)
- **phy_id** Physician ID code (128 unique codes)
- **priorfrac** History of Prior Fracture (1: No, 2: Yes)
- **age** Age at Enrollment (Years)
- **weight** Weight at enrollment (Kilograms)
- **height** Height at enrollment (Centimeters)
- **bmi** Body Mass Index (Kg/m^2)
- **premeno** Menopause before age 45 (1: No, 2: Yes)
- **momfrac** Mother had hip fracture (1: No, 2: Yes)
- **armassist** Arms are needed to stand from a chair (1: No, 2: Yes)
- **smoke** Former or current smoker (1: No, 2: Yes)
- **raterisk** Self-reported risk of fracture (1: Less than others of the same age, 2: Same as others of the same age, 3: Greater than others of the same age)
- **fracscore** Fracture Risk Score (Composite Risk Score)
- **fracture** Any fracture in first year (1: No, 2: Yes)

**Source**

Examples

```r
head(glow500, n = 10)
summary(glow500)

## Table 2.2 p. 39
summary(mod2.2 <- glm(fracture ~ age + weight + priorfrac +
                      premeno + raterisk,
                      family = binomial,
                      data = glow500))

## Table 2.3 p. 40
summary(mod2.3 <- update(mod2.2, . ~ . - weight - premeno))

## Table 2.4 p. 44
vcov(mod2.3)

## Table 3.6 p. 58
contrasts(glow500$raterisk)

## Contrasts: Table 3.8 and 3.9 p. 60
contrasts(glow500$raterisk) <- matrix(c(-1,-1,1,0,0,1), byrow= TRUE, ncol = 2)
summary(mod3.9 <- glm(fracture ~ raterisk, family = binomial,
                      data = glow500))
# cleaning modified dataset ...
rm(glow500)

## Table 5.1 pg 160 - Hosmer-Lemeshow test (with vcdExtra package)
mod4.16 <- glm(fracture ~ age * priorfrac + height + momfrac * armassist +
               I(as.integer(raterisk) == 3),
               family = binomial,
               data = glow500)
library(vcdExtra)
summary(HLtest(mod4.16))

## Table 5.3 p. 171 - Classification table
glow500$pred4.16 <- predict(mod4.16, type = "response")
with(glow500, addmargins(table( pred4.16 > 0.5, fracture)))

## Sensitivity, specificity, ROC (using pROC)
library(pROC)

## Figure 5.3 p. 177 - ROC curve (using pROC package)
print(roc4.16 <- roc(fracture ~ pred4.16, data = glow500))
plot(roc4.16, main = "Figure 5.3 p. 177")

## Table 5.8 p. 175
vars <- c("thresholds","sensitivities","specificities")
tab5.8 <- data.frame(roc4.16[vars])
## Now, for printing/comparison purposes, steps below in order to find
## threshold values most similar to those in the table
findIndex <- function(x, y) which.min( (x-y)^2 )
cutPoints <- seq(0.05, 0.75, by = 0.05)
```
tableIndex <- mapply(findIndex, y = cutPoints, MoreArgs = list(x = roc4.16$thresholds))
## And finally, let's print a reasonable approximation of table 5.8
writelines("\nTable 5.8 p. 175\n")
tab5.8[tableIndex, ]

## Figure 5.1 p. 175
plot(specificities ~ thresholds, xlim = c(0, 1), type = "l",
    xlab = "Probability cutoff", ylab = "Sensitivity/specificity",
    ylim = c(0, 1), data = tab5.8, main = "Figure 5.1 p. 175")
with(tab5.8, lines(thresholds, sensitivities, col = "red"))
legend(x = 0.75, y = 0.55, legend = c("Sensitivity", "Specificity"),
    lty = 1, col = c("red", "black"))
abline(h = c(0, 1), col = "grey80", lty = "dotted")

glow_bonemed

### Description

glow_bonemed dataset.

### Usage

glow_bonemed

### Format

A data.frame with 500 rows and 18 variables: the covariate are the same as those from glow500 with the addition of

- **bonemed** Bone medications at enrollment (1: No, 2: Yes)
- **bonemed_fu** Bone medications at follow-up (1: No, 2: Yes)
- **bonetreat** Bone medications both at enrollment and follow-up (1: No, 2: Yes)

### Source


### Examples

head(glow_bonemed, n = 10)

summary(glow_bonemed)
### GLOW_MIS_COMP data

**Description**

glow_mis_comp dataset.

**Usage**

`glow_mis_comp`

**Format**

A data.frame with 500 rows and 10 variables: the covariate are the same as those from `glow500`, without `bmi`, `premeno`, `armassist`, `smoke` and `fracscore`.

**Source**


**Examples**

```r
head(glow_mis_comp, n = 10)
summary(glow_mis_comp)
```

### GLOW_MIS_WMISSING data

**Description**

glow_mis_wmissing dataset.

**Usage**

`glow_mis_wmissing`

**Format**

A data.frame with 500 rows and 10 variables: the covariate are the same as those from `glow500`, without `bmi`, `premeno`, `armassist`, `smoke` and `fracscore`.

**Source**

**glow_rand**

**Examples**

```r
head(glow_mis_wmissing, n = 10)
summary(glow_mis_wmissing)
```

---

**glow_rand**

*GLOW RAND data*

**Description**

glow_rand dataset.

**Usage**

```r
glow_rand
```

**Format**

A data.frame with 500 rows and 15 variables: the covariate are the same as those from `glow500`.

**Source**


**Examples**

```r
head(glow_rand, n = 10)
summary(glow_rand)
```

---

**icu**

*ICU data*

**Description**

icu dataset.

**Usage**

```r
icu
```
Format

A data.frame with 200 rows and 21 variables:

- **id** Identification code (ID Number)
- **sta** Vital Status at hospital discharge (1: Lived, 2: Died)
- **age** Age (Years)
- **gender** Gender (1: Male, 2: Female)
- **race** Race (1: White, 2: Black, 3: Other)
- **ser** Service at ICU admission (1: Medical, 2: Surgical)
- **can** Cancer part of present problem (1: No, 2: Yes)
- **crn** History of chronic renal failure (1: No, 2: Yes)
- **inf** Infection probable at ICU admission (1: No, 2: Yes)
- **cpr** CPR prior to ICU admission (1: No, 2: Yes)
- **sys** Systolic blood pressure at ICU admission (mm Hg)
- **hra** Heart rate at ICU admission (Beats/min)
- **pre** Previous admission to an ICU within 6 months (1: No, 2: Yes)
- **type** Type of admission (1: Elective, 2: Emergency)
- **fra** Long bone, multiple, neck, single area, or hip fracture (1: No, 2: Yes)
- **po2** PO2 from initial blood gases (1: > 60, 2: <= 60)
- **ph** PH from initial blood gases (1: >= 7.25, 2: < 7.25)
- **pco** PCO2 from initial blood gases (1: <= 45, 2: > 45)
- **bic** Bicarbonate from initial blood gases (1: >= 18, 2: < 18)
- **cre** Creatinine from initial blood gases (1: <= 2.0, 2: > 2.0)
- **loc** Level of consciousness at ICU admission (1: No coma or deep stupor, 2: Deep stupor, 3: Coma)

Source


Examples

```r
head(icu, n = 10)
summary(icu)
```
Description

lowbwt dataset.

Usage

lowbwt

Format

A data.frame with 189 rows and 11 variables:

- **id** Identification Code
- **low** Low birth weight (1: $\geq 2500$, 2: < 2500 g)
- **age** Age of mother (Years)
- **lwt** Weight of mother at last menstrual period (Pounds)
- **race** Race (1: White, 2: Black, 3: Other)
- **smoke** Smoking status during pregnancy (1: No, 2: Yes)
- **ptl** History of premature labor (1: None, 2: One, 3: Two, etc)
- **ht** History of hypertension (1: No, 2: Yes)
- **ui** Presence of Uterine irritability (1: No, 2: Yes)
- **ftv** Number of physician visits during the first trimester (1: None, 2: One, 3: Two, etc)
- **bwt** Recorded birth weight (Grams)

Source


Examples

```r
head(lowbwt, n = 10)
summary(lowbwt)
```
myopia

Description
myopia dataset.

Usage
myopia

Format
A data.frame with 618 rows and 18 variables:
id Subject identifier (1-1503)
studyyear Year subject entered the study (Year)
myopic Myopia within the first five years of follow up (1: No, 2: Yes)
age Age at first visit (Years)
gender Gender (1: Male, 2: Female)
spheq Spherical Equivalent Refraction (diopter)
al Axial Length (mm)
acd Anterior Chamber Depth (mm)
lt Lens Thickness (mm)
vcd Vitreous Chamber Depth (mm)
sporthr How many hours per week outside of school the child spent engaging in sports/outdoor activities (Hours per week)
readhr How many hours per week outside of school the child spent reading for pleasure (Hours per week)
comphr How many hours per week outside of school the child spent playing video/computer games or working on the computer (Hours per week)
studyhr How many hours per week outside of school the child spent reading or studying for school assignments (Hours per week)
tvhr How many hours per week outside of school the child spent watching television (Hours per week)
diopterhr Composite of near-work activities (Hours per week)
mommy Was the subject’s mother myopic? (1: No, 2: Yes)
dadmy Was the subject’s father myopic? (1: No, 2: Yes)

Source
nhanes

Examples

  head(myopia, n = 10)
  summary(myopia)

nhanes  

NHANES data

Description

  nhanes dataset.

Usage

  nhanes

Format

  A data.frame with 6482 rows and 21 variables:

  id  Identification Code (1 - 6482)
  gender Gender (1: Male, 2: Female)
  age  Age at Screening (Years)
  marstat Marital Status (1: Married, 2: Widowed, 3: Divorced, 4: Separated, 5: Never Married, 6: Living Together)
  samplewt Statistical Weight (4084.478 - 153810.3)
  psu  Pseudo-PSU (1, 2)
  strata Pseudo-Stratum (1 - 15)
  tchol Total Cholesterol (mg/dL)
  hdl HDL-Cholesterol (mg/dL)
  sysbp Systolic Blood Pressure (mm Hg)
  dbp Diastolic Blood Pressure (mm Hg)
  wt  Weight (kg)
  ht  Standing Height (cm)
  bmi  Body mass Index (Kg/m^2)
  vigwrk Vigorous Work Activity (1: Yes, 2: No)
  modwrk  Moderate Work Activity (1: Yes, 2: No)
  wlkblk  Walk or Bicycle (1: Yes, 2: No)
  vigrecexr Vigorous Recreational Activities (1: Yes, 2: No)
  modrecexr  Moderate Recreational Activities (1: Yes, 2: No)
  sedmin Minutes of Sedentary Activity per Week (1: Yes, 2: No)
  obese BMI>35 (1: No, 2: Yes)
polypharm

Source

Examples
head(nhanes, n = 10)
summary(nhanes)

polypharm

POLYPHARM data

Description
polypharm dataset.

Usage
polypharm

Format
A data.frame with 3500 rows and 14 variables:

`id` Subject ID (1 - 500)
`polypharmacy` Outcome; taking drugs from more than three different classes (1: No, 2: Yes)
`mhv4` Number of outpatient Mental Health Visits (1: none, 2: one to five, 3: six to fourteen, 4: greater than 14)
`inptmhv3` Number of inpatient Mental Health Visits (1: none, 2: one, 3: more than one)
`year` Year (2002 to 2008)
`group` Group (1: Covered Families and Children - CFC, 2: Aged, Blind or Disabled - ABD, 3: Foster Care - FOS)
`urban` Location (1: Urban, 2: Rural)
`comorbid` Comorbidity (1: No, 2: Yes)
`anyprim` Any primary diagnosis (bipolar, depression, etc.) (1: No, 2: Yes)
`numprim` Number of primary diagnosis (1: none, 2: one, 3: more than one)
`gender` Gender (1: Female, 2: Male)
`race` Race (1: White, 2: Black, 3: Other)
`ethnic` Ethnic category (1: Non-Hispanic, 2: Hispanic)
`age` Age (Years and months, two decimal places)

Source
Examples

```
head(polypharm, n = 10)
summary(polypharm)
```

Description

scale_example dataset.

Usage

scale_example

Format

A data.frame with 500 rows and 2 variables:

- **y**: a dicotomic variable (say 1: No, 2: Yes)
- **x**: a numeric variable

Source


Examples

```
head(scale_example, n = 10)
summary(scale_example)
```
Index

*Topic datasets
  aps, 3
  burn1000, 4
  burn13m, 5
  burn_eval_1, 6
  burn_eval_2, 6
  chdage, 7
  glow11m, 8
  glow500, 9
  glow_bonemed, 11
  glow_mis_comp, 12
  glow_mis_wmissing, 12
  glow_rand, 13
  icu, 13
  lowbwt, 15
  myopia, 16
  nhanes, 17
  polypharm, 18
  scale_example, 19

aplore3, 2
aplore3-package (aplore3), 2
aps, 3

burn1000, 4, 5, 6
burn13m, 5
burn_eval_1, 6
burn_eval_2, 6

chdage, 7

glow11m, 8

glow500, 8, 9, 11–13

glow_bonemed, 11

glow_mis_comp, 12

glow_mis_wmissing, 12

glow_rand, 13

icu, 13

lowbwt, 15