

The doBy package

by Søren Højsgaard

Abstract The doBy is one of several general utility packages on CRAN. An abstract of less than 150 words.

Introduction

The doBy package (Højsgaard and Halekoh, 2020) appeared on CRAN (?) in 2006 and, much to our surprise, the package is still being used. The package originally grew out of a need to calculate groupwise summary statistics (much in the spirit of PROC SUMMARY of the SAS system, (SAS Institute Inc., 2020)). The name comes from doing some computations when data is stratified by the value of some variables. Today the package contains many different utilities. In this paper we focus 1) on these “doing by” functions, 2) on functions related to linear estimates and contrasts and 3) on some of the miscellaneous functions in the package.

A working dataset

```
data(CO2)
CO2 <- transform(CO2, Treat=Treatment, Treatment=NULL)
levels(CO2$Treat) <- c("nchl", "chl")
levels(CO2$Type) <- c("Que", "Mis")
CO2 <- subset(CO2, Plant %in% c("Qn1", "Qc1", "Mn1", "Mc1"))
airquality <- subset(airquality, Month %in% c(5,6))

mtcars <- within(mtcars, {
  vs <- factor(vs, labels = c("V", "S"))
  am <- factor(am, labels = c("auto", "man"))
})
mtcars$drat <- mtcars$disp <- mtcars$wt <- mtcars$carb <- mtcars$gear <- NULL
mtcars <- subset(mtcars, cyl < 8)
mtcars %>% head

#>           mpg  cyl  hp  qsec vs   am
#> Mazda RX4    21.0   6 110 16.46 V  man
#> Mazda RX4 Wag 21.0   6 110 17.02 V  man
#> Datsun 710    22.8   4  93 18.61 S  man
#> Hornet 4 Drive 21.4   6 110 19.44 S auto
#> Valiant      18.1   6 105 20.22 S auto
#> Merc 240D    24.4   4  62 20.00 S auto
```

Functions related to groupwise computations

summaryBy

```
library(doBy)
summaryBy(cbind(mpg, qsec) ~ cyl + vs, data=mtcars)

#>   cyl vs mpg.mean qsec.mean
#> 1    4 V   26.00    16.70
#> 2    4 S   26.73    19.38
#> 3    6 V   20.57    16.33
#> 4    6 S   19.12    19.21

summaryBy(list(c("mpg", "qsec"), c("cyl", "vs")), data=mtcars)

#>   cyl vs mpg.mean qsec.mean
#> 1    4 V   26.00    16.70
#> 2    4 S   26.73    19.38
#> 3    6 V   20.57    16.33
#> 4    6 S   19.12    19.21

summaryBy(. ~ cyl + vs, data=mtcars)
```

```
#>   cyl vs mpg.mean hp.mean qsec.mean
#> 1   4 V   26.00    91.0    16.70
#> 2   4 S   26.73    81.8    19.38
#> 3   6 V   20.57   131.7    16.33
#> 4   6 S   19.12   115.2    19.21

summaryBy(. ~ ., data=mtcars)

#>   vs   am mpg.mean cyl.mean hp.mean qsec.mean
#> 1 V man   21.93    5.500  121.50    16.42
#> 2 S auto   20.74    5.143  102.14    19.97
#> 3 S man   28.37    4.000   80.57    18.70

ss <- splitBy(~ vs, data=mtcars)
ss

#>   listentry vs
#> 1         V V
#> 2         S S

ss$V

#>           mpg cyl  hp  qsec vs  am
#> Mazda RX4    21.0   6 110 16.46 V man
#> Mazda RX4 Wag 21.0   6 110 17.02 V man
#> Porsche 914-2 26.0   4  91 16.70 V man
#> Ferrari Dino  19.7   6 175 15.50 V man
```

Functions related linear estimates and contrasts

```
esticon()
linest()
LSmeans()
```

Miscellaneous functions

Summary

This file is only a basic article template. For full details of *The R Journal* style and information on how to prepare your article for submission, see the [Instructions for Authors](#).

Bibliography

S. Højsgaard and U. Halekoh. *doBy: Groupwise Statistics, LSmeans, Linear Contrasts, Utilities*, 2020. URL <http://people.math.aau.dk/~sorenh/software/doBy/>. R package version 4.6.6. [p1]

SAS Institute Inc. *Base SAS 9.4 Procedures Guide, Seventh Edition*, April 2020. [p1]

Søren Højsgaard
 Department of Mathematical Sciences, Aalborg University, Denmark
 line 1
 line 2
sorenh@math.aau.dk