Merging_data_sets.R

015004

Thu Feb 21 13:57:40 2019

```
#################
# Load packages #
#################

#install.packages("JM")
library(JM)
```

```
## Warning: package 'JM' was built under R version 3.5.2
```

```
## Loading required package: MASS
```

```
## Loading required package: nlme
```

```
## Loading required package: splines
```

```
## Loading required package: survival
```

```
## Warning: package 'survival' was built under R version 3.5.2
```

```
######################
# Merging data sets #
######################

## Example 1

dat1 = data.frame(Id = c(1:6), drug = c(rep("placebo", 3), rep("treatment", 3)))
dat2 = data.frame(Id = c(1:6), Country = c(rep("Netherlands", 2), rep("Belgium", 4)))
merge(dat1, dat2)
```

```
##   Id      drug     Country
## 1  1   placebo Netherlands
## 2  2   placebo Netherlands
## 3  3   placebo     Belgium
## 4  4 treatment     Belgium
## 5  5 treatment     Belgium
## 6  6 treatment     Belgium
```

```
dat1 = data.frame(Id = c(1:6), time = c(0, 0.2, 0.4, 0, 0.2, 1),
                  drug = c(rep("placebo", 3), rep("treatment", 3)))
dat2 = data.frame(Id = c(1:6), time = c(0, 0.2, 0.4, 0, 0.2, 1),
                  Country = c(rep("Netherlands", 2), rep("Belgium", 4)))
merge(dat1, dat2)
```

```
##   Id time      drug     Country
## 1  1  0.0   placebo Netherlands
## 2  2  0.2   placebo Netherlands
## 3  3  0.4   placebo     Belgium
## 4  4  0.0 treatment     Belgium
## 5  5  0.2 treatment     Belgium
## 6  6  1.0 treatment     Belgium
```

```r
merge(dat1, dat2, by = "Id")
```

```
##   Id time.x      drug time.y     Country
## 1  1    0.0   placebo    0.0 Netherlands
## 2  2    0.2   placebo    0.2 Netherlands
## 3  3    0.4   placebo    0.4     Belgium
## 4  4    0.0 treatment    0.0     Belgium
## 5  5    0.2 treatment    0.2     Belgium
## 6  6    1.0 treatment    1.0     Belgium
```

```r
## Example 2

dat1 = data.frame(Id = c(1:6), drug = c(rep("placebo", 3), rep("treatment", 3)))
dat2 = data.frame(Id = c(2, 4, 6), Country = c(rep("Netherlands", 2), rep("Belgium", 1)))
merge(dat1, dat2)
```

```
##   Id      drug     Country
## 1  2   placebo Netherlands
## 2  4 treatment Netherlands
## 3  6 treatment     Belgium
```

```r
merge(dat1, dat2, all = TRUE)
```

```
##   Id      drug     Country
## 1  1   placebo        <NA>
## 2  2   placebo Netherlands
## 3  3   placebo        <NA>
## 4  4 treatment Netherlands
## 5  5 treatment        <NA>
## 6  6 treatment     Belgium
```

```r
merge(dat1, dat2, all = FALSE)
```

```
##   Id      drug     Country
## 1  2   placebo Netherlands
## 2  4 treatment Netherlands
## 3  6 treatment     Belgium
```

```r
merge(dat1, dat2, all.x = TRUE)
```

```
##   Id      drug     Country
## 1  1   placebo        <NA>
## 2  2   placebo Netherlands
## 3  3   placebo        <NA>
## 4  4 treatment Netherlands
## 5  5 treatment        <NA>
## 6  6 treatment     Belgium
```

```r
merge(dat1, dat2, all.y = TRUE)
```

```
##   Id      drug     Country
## 1  2   placebo Netherlands
## 2  4 treatment Netherlands
## 3  6 treatment     Belgium
```

```r
## Example 3

dat1 = data.frame(Id = c(1,1,1,2,2,2), score = c(sample(1:20, 6)))
```

```r
dat2 = data.frame(Id = c(1,2), Country = c(rep("Netherlands", 1), rep("Belgium", 1)))
merge(dat1, dat2)
```

```
##   Id score     Country
## 1  1     8 Netherlands
## 2  1     7 Netherlands
## 3  1    14 Netherlands
## 4  2     2     Belgium
## 5  2     6     Belgium
## 6  2    17     Belgium
```

```r
merge(dat1, dat2, by = "Id")
```

```
##   Id score     Country
## 1  1     8 Netherlands
## 2  1     7 Netherlands
## 3  1    14 Netherlands
## 4  2     2     Belgium
## 5  2     6     Belgium
## 6  2    17     Belgium
```

```r
## Example 4

dat1 = data.frame(Id = c(1,1,1,2,2,2), score = c(sample(1:20, 6)))
dat2 = data.frame(IDs = c(1,2), Country = c(rep("Netherlands", 1), rep("Belgium", 1)))
merge(dat1, dat2)
```

```
##    Id score IDs     Country
## 1   1    13   1 Netherlands
## 2   1    15   1 Netherlands
## 3   1     2   1 Netherlands
## 4   2     6   1 Netherlands
## 5   2    20   1 Netherlands
## 6   2    18   1 Netherlands
## 7   1    13   2     Belgium
## 8   1    15   2     Belgium
## 9   1     2   2     Belgium
## 10  2     6   2     Belgium
## 11  2    20   2     Belgium
## 12  2    18   2     Belgium
```

```r
merge(dat1, dat2, by.x = c("Id"), by.y = c("IDs"))
```

```
##   Id score     Country
## 1  1    13 Netherlands
## 2  1    15 Netherlands
## 3  1     2 Netherlands
## 4  2     6     Belgium
## 5  2    20     Belgium
## 6  2    18     Belgium
```

```r
## Example 5

dat1 = data.frame(Id = c(1:6), drug = c(rep("placebo", 3),
                    rep("treatment", 3)), x1 = c(1:6), x2 = c(6:1))
dat2 = data.frame(Id = c(2, 4, 6), Country = c(rep("Netherlands", 2),
```

```
                    rep("Belgium", 1)), x1 = sample(1:10, 3), x2 = sample(1:10, 3))
merge(dat1, dat2)
```

```
## [1] Id      x1      x2      drug    Country
## <0 rows> (or 0-length row.names)
```

```
merge(dat1, dat2, by.x = c("Id"), by.y = c("Id"))
```

```
##   Id      drug x1.x x2.x     Country x1.y x2.y
## 1  2   placebo    2    5 Netherlands    1    3
## 2  4 treatment    4    3 Netherlands    5    2
## 3  6 treatment    6    1     Belgium    2    8
```

```
merge(dat1, dat2, by = c("Id"))
```

```
##   Id      drug x1.x x2.x     Country x1.y x2.y
## 1  2   placebo    2    5 Netherlands    1    3
## 2  4 treatment    4    3 Netherlands    5    2
## 3  6 treatment    6    1     Belgium    2    8
```