Data_Manipulation.R

015004

Tue Jan 22 10:43:49 2019

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

library(JM)
```

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

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

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

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

```
#####################
# Data Manipulation #
#####################

## Round continuous variables
pbc2.id$years <- round(pbc2.id$years, digits = 2)
pbc2.id$age <- round(pbc2.id$age, digits = 2)
head(pbc2.id)
```

```
##   id years       status      drug   age    sex year ascites hepatomegaly
## 1  1  1.10         dead D-penicil 58.77 female    0     Yes          Yes
## 2  2 14.15        alive D-penicil 56.45 female    0      No          Yes
## 3  3  2.77         dead D-penicil 70.07   male    0      No           No
## 4  4  5.27         dead D-penicil 54.74 female    0      No          Yes
## 5  5  4.12 transplanted   placebo 38.11 female    0      No          Yes
## 6  6  6.85         dead   placebo 66.26 female    0      No          Yes
##   spiders                    edema serBilir serChol albumin alkaline  SGOT
## 1     Yes edema despite diuretics     14.5     261    2.60     1718 138.0
## 2     Yes                No edema      1.1     302    4.14     7395 113.5
## 3      No      edema no diuretics      1.4     176    3.48      516  96.1
## 4     Yes      edema no diuretics      1.8     244    2.54     6122  60.6
## 5     Yes                No edema      3.4     279    3.53      671 113.2
## 6      No                No edema      0.8     248    3.98      944  93.0
##   platelets prothrombin histologic status2
## 1       190        12.2          4       1
## 2       221        10.6          3       0
## 3       151        12.0          4       1
## 4       183        10.3          4       1
## 5       136        10.9          3       0
## 6        NA        11.0          3       1
```

```
## Set age40higher as factor
factor(pbc2.id$age40higher, levels = c(0:1),
       labels = c("low","high"))
```

```
## factor(0)
```

```
## Levels: low high
```

```
# why do I not get the new dichotomous variable?
pbc2.id$age40higher <- pbc2.id$age > 40
pbc2.id$age40higher <- as.numeric(pbc2.id$age40higher)
head(pbc2.id)
```

```
##   id years       status      drug   age    sex year ascites hepatomegaly
## 1  1  1.10         dead D-penicil 58.77 female    0     Yes          Yes
## 2  2 14.15        alive D-penicil 56.45 female    0      No          Yes
## 3  3  2.77         dead D-penicil 70.07   male    0      No           No
## 4  4  5.27         dead D-penicil 54.74 female    0      No          Yes
## 5  5  4.12 transplanted   placebo 38.11 female    0      No          Yes
## 6  6  6.85         dead   placebo 66.26 female    0      No          Yes
##   spiders                   edema serBilir serChol albumin alkaline  SGOT
## 1     Yes edema despite diuretics     14.5     261    2.60     1718 138.0
## 2     Yes                No edema      1.1     302    4.14     7395 113.5
## 3      No      edema no diuretics      1.4     176    3.48      516  96.1
## 4     Yes      edema no diuretics      1.8     244    2.54     6122  60.6
## 5     Yes                No edema      3.4     279    3.53      671 113.2
## 6      No                No edema      0.8     248    3.98      944  93.0
##   platelets prothrombin histologic status2 age40higher
## 1       190        12.2          4       1           1
## 2       221        10.6          3       0           1
## 3       151        12.0          4       1           1
## 4       183        10.3          4       1           1
## 5       136        10.9          3       0           0
## 6        NA        11.0          3       1           1
```

```
pbc2.id$age40higher <- factor(pbc2.id$age40higher, levels = c(0:1),
                              labels = c("low","high"))
head(pbc2.id)
```

```
##   id years       status      drug   age    sex year ascites hepatomegaly
## 1  1  1.10         dead D-penicil 58.77 female    0     Yes          Yes
## 2  2 14.15        alive D-penicil 56.45 female    0      No          Yes
## 3  3  2.77         dead D-penicil 70.07   male    0      No           No
## 4  4  5.27         dead D-penicil 54.74 female    0      No          Yes
## 5  5  4.12 transplanted   placebo 38.11 female    0      No          Yes
## 6  6  6.85         dead   placebo 66.26 female    0      No          Yes
##   spiders                   edema serBilir serChol albumin alkaline  SGOT
## 1     Yes edema despite diuretics     14.5     261    2.60     1718 138.0
## 2     Yes                No edema      1.1     302    4.14     7395 113.5
## 3      No      edema no diuretics      1.4     176    3.48      516  96.1
## 4     Yes      edema no diuretics      1.8     244    2.54     6122  60.6
## 5     Yes                No edema      3.4     279    3.53      671 113.2
## 6      No                No edema      0.8     248    3.98      944  93.0
##   platelets prothrombin histologic status2 age40higher
## 1       190        12.2          4       1        high
## 2       221        10.6          3       0        high
## 3       151        12.0          4       1        high
## 4       183        10.3          4       1        high
## 5       136        10.9          3       0         low
## 6        NA        11.0          3       1        high
```

```
## Transform variables
pbc2.id$ageST <- (pbc2.id$age-mean(pbc2.id$age))/(sd(pbc2.id$age))


## Wide/long format
head(pbc2)
```

```
##   id    years status      drug      age    sex      year ascites
## 1  1  1.09517   dead D-penicil 58.76684 female 0.0000000     Yes
## 2  1  1.09517   dead D-penicil 58.76684 female 0.5256817     Yes
## 3  2 14.15234  alive D-penicil 56.44782 female 0.0000000      No
## 4  2 14.15234  alive D-penicil 56.44782 female 0.4983025      No
## 5  2 14.15234  alive D-penicil 56.44782 female 0.9993429      No
## 6  2 14.15234  alive D-penicil 56.44782 female 2.1027270      No
##   hepatomegaly spiders                   edema serBilir serChol albumin
## 1          Yes     Yes edema despite diuretics     14.5     261    2.60
## 2          Yes     Yes edema despite diuretics     21.3      NA    2.94
## 3          Yes     Yes                No edema      1.1     302    4.14
## 4          Yes     Yes                No edema      0.8      NA    3.60
## 5          Yes     Yes                No edema      1.0      NA    3.55
## 6          Yes     Yes                No edema      1.9      NA    3.92
##   alkaline  SGOT platelets prothrombin histologic status2
## 1     1718 138.0       190        12.2          4       1
## 2     1612   6.2       183        11.2          4       1
## 3     7395 113.5       221        10.6          3       0
## 4     2107 139.5       188        11.0          3       0
## 5     1711 144.2       161        11.6          3       0
## 6     1365 144.2       122        10.6          3       0
```

```
vec <- table(pbc2$id)
vec2 <- sequence(vec)

pbc2$visit <- vec2

library(reshape2)
pbc2Wide <- reshape(pbc2, idvar = c("id"),
                    drop = c("years", "status", "drug", "year", "hepatomegaly",
                             "serChol", "spiders", "albumin", "alkaline",
                             "SGOT", "platelets", "prothrombin", "age", "sex",
                             "ascites", "edema", "histologic", "status2"),
                    timevar = "visit", direction = "wide")


head(pbc2Wide)
```

```
##    id serBilir.1 serBilir.2 serBilir.3 serBilir.4 serBilir.5 serBilir.6
## 1   1       14.5       21.3         NA         NA         NA         NA
## 3   2        1.1        0.8        1.0        1.9        2.6        3.6
## 12  3        1.4        1.1        1.5        1.8         NA         NA
## 16  4        1.8        1.6        1.7        3.2        3.7        4.0
## 23  5        3.4        1.9        2.5        5.7        5.2       19.0
## 29  6        0.8        0.8        0.9        0.7        0.6        0.7
##    serBilir.7 serBilir.8 serBilir.9 serBilir.10 serBilir.11 serBilir.12
## 1          NA         NA         NA          NA          NA          NA
```

```
## 3         4.2        3.6        4.6          NA          NA          NA
## 12         NA         NA         NA          NA          NA          NA
## 16        5.3         NA         NA          NA          NA          NA
## 23         NA         NA         NA          NA          NA          NA
## 29         NA         NA         NA          NA          NA          NA
##    serBilir.13 serBilir.14 serBilir.15 serBilir.16
## 1           NA          NA          NA          NA
## 3           NA          NA          NA          NA
## 12          NA          NA          NA          NA
## 16          NA          NA          NA          NA
## 23          NA          NA          NA          NA
## 29          NA          NA          NA          NA
```

```
pbc2Long <- reshape(pbc2Wide, idvar = c("id"), timevar = "visit",
                    varying = list(names(pbc2Wide)[2:17]),
                    v.names = "serBilir", direction = "long", times = 1:16)
head(pbc2Long)
```

```
##     id visit serBilir
## 1.1  1     1     14.5
## 2.1  2     1      1.1
## 3.1  3     1      1.4
## 4.1  4     1      1.8
## 5.1  5     1      3.4
## 6.1  6     1      0.8
```

```
pbc2Long <- pbc2Long[order(pbc2Long$id),]
head(pbc2Long)
```

```
##     id visit serBilir
## 1.1  1     1     14.5
## 1.2  1     2     21.3
## 1.3  1     3       NA
## 1.4  1     4       NA
## 1.5  1     5       NA
## 1.6  1     6       NA
```

```
## Missings
?mean
```

```
## starting httpd help server ...
```

```
##  done
```