getdata4.Rmd

Göran Broström

2022-02-09

Introduction

Created 6 February 2022 after consultations.

** Run this (and only this) before running "paper4.Rmd"!**

We start with *skum::persons* and age is to begin with in days, with stillbirths *removed* and born alive but dead on birthday at age 0.25 days.

For use of *eha::make.communal*, we need to temporarily change time unit to years.

Further, we eventually skip the use of eha::make.communal in favor of *match*. Reason: eha::make.communal assumes age is measured in *years*. But this is no promise yet.

Read skum::pers

```
per <- pers[(pers$frsbobtyp == 2) & (pers$df == 0), ]
per$frsbobtyp <- per$df <- per$f2id <- per$m2id <- NULL
per$frsbobdatmin <- per$frsbobdatmax <- per$utdatmax <- per$utdatmin <- NULL
per <- per[(per$foddat > 18950000), ]
dn <- per$doddat < 1
per$doddat[dn] <- per$foddat[dn] + 40000
## Fix foddat:
yr <- per$foddat %% 10000 == 0
per$foddat[yr] <- per$foddat[yr] + 101 # As early as possible
mon <- per$foddat %% 100 == 0
per$foddat[mon] <- per$foddat[mon] + 1 # Ditto
## Fix doddat:
yr <- per$doddat %% 10000 == 0
per$doddat[yr] <- per$doddat[yr] + 1231 # As late as possible
mon <- per$doddat %% 100 == 0
per$doddat[mon] <- per$doddat[mon] + 28 # Ditto (guard for February ...)
##
per$birthdate <- as.Date(as.character(per$foddat), format = "%Y%m%d")
per$deathdate <- as.Date(as.character(per$doddat), format = "%Y%m%d")
nabir <- is.na(per$birthdate)
per$foddat[nabir] <- per$foddat[nabir] - 2
per$doddat[nabir] <- per$doddat[nabir] - 2
##
per$birthdate <- as.Date(as.character(per$foddat), format = "%Y%m%d")
```

```
per$exit <- as.numeric(difftime(per$deathdate, per$birthdate))
per$deathdate <- as.Date(as.character(per$doddat), format = "%Y%m%d")
## Fix doddat:
per$doddat[is.na(per$deathdate)] <- per$doddat[is.na(per$deathdate)] - 1
per$deathdate <- as.Date(as.character(per$doddat), format = "%Y%m%d")
##
per$exit <- as.numeric(difftime(per$deathdate, per$birthdate, units = "days"))
oj <- per$exit < 0
per$doddat[oj] <- per$doddat[oj] + 10000
per$deathdate <-  as.Date(as.character(per$doddat), format = "%Y%m%d")
per$exit <- as.numeric(difftime(per$deathdate, per$birthdate, units = "days"))
per$event <- per$exit <= 365
per$exit[per$exit == 0] <- 0.25
per$exit[per$exit > 365] <- 365
```

Now we need *nofrs* for newborn, found in the *observation* file(s), so

```
bobs <- obs[obs$starttyp == 2 & obs$id %in% per$id, ]
indx <- match(per$id, bobs$id)
per$socStatus <- bobs$socStatus[indx]
per$socBranch <- bobs$socBranch[indx]
per$hisclass <- bobs$hisclass[indx]
per$nofrs <- bobs$nofrs[indx]
per <- per[!is.na(per$nofrs), ]
per$subreg <- factor(per$nofrs)
levels(per$subreg) <- c(rep("ume", 2),
                        rep("bjur", 4),
                        rep("sten", 3))
##per <- per[(per$subreg != "sten"), ] # NOTE: Now Stensele station is included
per$subreg <- factor(per$subreg)
per$enter <- 0
per$sex <- factor(per$kon, labels = c("boy", "girl"))
per$birthdate <- toTime(per$birthdate)
per$illeg <- as.numeric(per$ab == 2)
per$parity <- per$paritet_g
per$paritet_g <- per$paritet_k <- per$doddat <- per$frsbosdat <- NULL
per$fid <- per$fboort <- per$ab <- per$fyrkfd <- per$fyrkind <- NULL
per$region <- per$kon <- per$fodhfrs <- per$frsbobdat <- per$frsbostyp <- NULL
per$fbonmn <- per$fboind <- per$fyrkrel <- per$fyrketxt <- NULL
per$foddat <- as.Date(as.character(per$foddat), format = "%Y%m%d")
## NEW:
per$socBranch <- as.character(per$socBranch)
per$socBranch[is.na(per$socBranch)] <- "none"
per$socBranch <- factor(per$socBranch)
per$socStatus <- as.character(per$socStatus)
per$socStatus[is.na(per$socStatus)] <- "low"
per$socStatus <- factor(per$socStatus)
per <- per[, c("id", "sex", "birthdate", "mid", "foddat", "parity", "illeg",
               "subreg", "socBranch", "socStatus", "hisclass", "enter", "exit", "event")]
#saveRDS(per, file = "data/infdat3.rds")
```

Temperature data

We have three stations, *Umeå*, *Bjuröklubb*, and *Stensele*.

```
source("R/tempdat.R")
temp_start <- 1894
temp_end <- 1951
temp_ivl <- c(temp_start, temp_end)
##
umetemp <- tempdat("ume", temp_ivl)
bjurtemp <- tempdat("bjur", temp_ivl)

##Changed 5 feb 2022:
##stentemp <- tempdat("sten", temp_ivl) # Added 27 Jan 2022
source("R/normal.R") # normal = Norsjö - Malå!!
stentemp <- normal(temp_ivl)
## End change

saveRDS(umetemp, file = "data/umetemp.rds")
saveRDS(bjurtemp, file = "data/bjurtemp.rds")
saveRDS(stentemp, file = "data/stentemp.rds")  ## Added 27 Jan2022
```

Put on temperature data

Here we must change time unit to *year*.

```
per$enter <- per$enter / 365
per$exit <- per$exit / 365
## per$birthdate <- toTime(per$birthdate) #already done!
temp_start <- 1894
temp_end <- 1951
temp_ivl <- c(temp_start, temp_end)
##

lagg <- 0 / 365 ##3 / 365  ## NOTE!!
tempvars <-  c("heat", "heat.1", "extemp", "extemp.1", "cold", "cold.1", "emintemp", "emeantemp", "emax
               "week", "year")
## Ume:
bume <- per[per$subreg == "ume", ]
bume <- make.communal(bume, umetemp[, tempvars],
                      start = temp_start, period = 1/52, lag = lagg)

### Ske:
bskel <- per[per$subreg == "bjur", ]
bskel <- make.communal(bskel, bjurtemp[, tempvars],
                       start = temp_start, period = 1/52, lag = lagg)
```

```
## Warning in make.communal(bskel, bjurtemp[, tempvars], start = temp_start, :
## Spells are cut
```

```
### Sten:
bsten <- per[per$subreg == "sten", ]
bsten <- make.communal(bsten, stentemp[, tempvars],
                       start = temp_start, period = 1/52, lag = lagg)


bume$region <- "ume"
bskel$region <- "ske"
bsten$region <- "sten"
infdat <- rbind(bume, bskel, bsten)
infdat$birthmonth <- ceiling((infdat$birthdate - floor(infdat$birthdate)) * 12)
infdat$period <- cut(infdat$birthdate, c(1895, 1914, 1935, 1951), dig.lab = 5)
infdat <- infdat[order(infdat$id, infdat$enter), c("subreg", "id", "sex", "enter", "exit", "event",
                                                   "birthmonth", "period", "socBranch", "socStatus", "il
                                                   "parity", "hisclass", tempvars)]
## New addition: (28 Jan 2022)
##infdat <- rc(infdat) # rc from 'skum'
##n <- NROW(infdat)

saveRDS(infdat, file = "data/infdat4.rds") # NOTE: Name change!!!

## Wait with this:
## tabdat4 <- toTpch(Surv(enter, exit, event) ~ subreg + sex + birthmonth + socBranch + illeg + parity
```

Note the name change of the data file: *infdat3.rds*! And the the time lag is 0.