

Creation pp statut

Exemple de création d'un fichier personnes-périodes et d'une variable de statut.

On utilise le fichier SHP_divorce.rda extrait du panel suisse des ménages. Ce fichier contient les variables suivantes:

- IDPERS: identificateur personnel
- SEX
- AGE99 -> AGE09: âge de 1999 à 2009
- CIVSTA99 -> CIVSTA09: état-civil de 1999 à 2009
- NBKID99 -> NBKID09: nombre d'enfants vivant avec la personne de 1999 à 2009

Ouverture et premier examen des données :

```
load("SHP_divorce.rda")
summary(Dataset)
```

Le but est d'étudier le divorce chez les personnes qui étaient mariées au début de l'étude (1999).

1) On crée un nouveau jeu de données ne contenant que les personnes mariées en 1999 :

```
Data_married <- subset(Dataset, Dataset$CIVSTA99=="married")
summary(Data_married)
```

2) On crée le fichier personnes-périodes correspondant :

Définition des variables temporelles:

```
AGE <-
c("AGE99", "AGE00", "AGE01", "AGE02", "AGE03", "AGE04", "AGE05", "AGE06", "AGE07",
  "AGE08", "AGE09")
```

```
CIVSTA <-
c("CIVSTA99", "CIVSTA00", "CIVSTA01", "CIVSTA02", "CIVSTA03", "CIVSTA04", "CIVSTA05",
  "CIVSTA06", "CIVSTA07", "CIVSTA08", "CIVSTA09")
```

```
NBKID <-
c("NBKID99", "NBKID00", "NBKID01", "NBKID02", "NBKID03", "NBKID04", "NBKID05",
  "NBKID06", "NBKID07", "NBKID08", "NBKID09")
```

Création du nouveau jeu de données personnes-périodes :

```
Data_married_pp <- reshape(Data_married, direction="long",
  varying=list(AGE, CIVSTA, NBKID))
```

Réordonnement selon les personnes :

```
ordre <- sort.list(Data_married_pp[,1])
Data_married_pp <- Data_married_pp[ordre,]
```

Changement des noms de variables :

```
names(Data_married_pp)[c(4)] <- c("AGE")
names(Data_married_pp)[c(5)] <- c("CIVSTA")
names(Data_married_pp)[c(6)] <- c("NBKID")
```

3) Ajout de deux variables représentant les temps de début et de fin de chaque année en mois à partir de janvier 1999 :

Création des vecteurs pour une personne :

```
t1 <- c(1,13,25,37,49,61,73,85,97,109,121)
t2 <- c(12,24,36,48,60,72,84,96,108,120,132)
```

Création des vecteurs pour toutes les personnes (utilisation du produit de Kronecker) :

```
nbpersonnes <- nrow(Data_married)
time_deb <- kronecker(rep(1,nbpersonnes),t1)
time_fin <- kronecker(rep(1,nbpersonnes),t2)
```

Ajout des deux nouvelles variables au fichier de données :

```
Data_married_pp <- data.frame(Data_married_pp,time_deb,time_fin)
```

4) Création de la variable "statut" qui doit être codée 1 pour les événements (divorces) et 0 pour les censures. Provisoirement, la variable statut pourra aussi prendre le code 2 pour les données à supprimer.

Attention: Il peut y avoir des données manquantes dans l'état-civil. Par ailleurs, il faut supprimer les états-civil qui ne nous intéressent pas (veuf, séparé, ...).

On repère dans la variable CIVSTA les lignes qui ne sont ni "married", ni "divorced" et on les remplace par des données manquantes :

```
temp <- which(Data_married_pp$CIVSTA!="married" &
Data_married_pp$CIVSTA!="divorced")
Data_married_pp$CIVSTA[temp] <- NA
```

Initialisation à zéro de la variable statut et ajout de celle-ci dans la base de données :

```
nblignes <- nrow(Data_married_pp)
statut <- rep(0,nblignes)
Data_married_pp <- data.frame(Data_married_pp,statut)
```

On recode la variable statut 1 pour les divorces :

```
temp <- which(Data_married_pp$CIVSTA=="divorced")
Data_married_pp$statut[temp] <- 1
```

On recode la variable statut 2 pour les données manquantes :

```
temp <- which(is.na(Data_married_pp$CIVSTA)==TRUE)
Data_married_pp$statut[temp] <- 2
```

Soit nbperiod le nombre de périodes d'observations de chaque personne (dans notre exemple: 11) et nbpersonnes le nombre de personnes :

```
nbperiod <-11
```

Pour chaque personne (boucle de 1 à nbpersonnes), on recode le statut à 2 les années suivant un divorce ou une donnée manquante :

```
ligne <- 0
for (i in 1:nbpersonnes) {
  test <- 0
  for (j in 1:nbperiod) {
    ligne <- ligne+1
    ligne2 <- ligne+nbperiod-j
    if (test==1) {
      Data_married_pp$statut[ligne:ligne2] <- 2
      ligne <- ligne2
      break
    }
    if (Data_married_pp$statut[ligne]==1 |
Data_married_pp$statut[ligne]==2) {
      test <- 1
    }
  }
}
```

5) Finalisation et sauvegarde du nouveau fichier.

Suppression de toutes les lignes pour lesquelles le statut est codé 2 :

```
Data_married_pp <- subset(Data_married_pp,Data_married_pp$statut<2)
```

Sauvegarde:

```
save(Data_married_pp,file="Data_married_pp.rda")
```

6) Analyse

```
library(survival)
attach(Data_married_pp)
```

Définition d'un objet survie

```
survie_divorce <- Surv(time_deb,time_fin,statut)
```

Courbe de survie de Kaplan-Meier

```
survie_KM <- survfit(survie_divorce~1)
summary(survie_KM)
```

Plot de la fonction de survie

```
plot(survie_KM)
```

Modèles de Cox

```
Cox_1 <- coxph(survie_divorce~as.numeric(SEX)+AGE+NBKID)
summary(Cox_1)
```