

Analyse de survie (Master ISN)

G. Babykina

1 Quelques commandes utiles, packages nécessaires

```
> library(survival)
> library(survminer) # ! Super pour les graphs, ne pas hésiter à s'y mettre si besoin
```

2 Données sur recidivism

For illustration consider data from an experimental study of recidivism of 432 male prisoners, who were observed for a year after being released from prison (Rossi et al., 1980). The variables included in the data are described in Table 1.

TABLE 1 – Variables disponibles pour les données sur recidivisme

Var	Description
week	week of first arrest after release, or censoring time.
arrest	the event indicator, equal to 1 for those arrested during the period of the study and 0 for those who were not arrested.
fin	a factor, with levels yes if the individual received financial aid after release from prison, and no if he did not ; financial aid was a randomly assigned factor manipulated by the researchers.
age	in years at the time of release.
race	a factor with levels black and other.
wexp	a factor with levels yes if the individual had full-time work experience prior to incarceration and no if he did not.
mar	a factor with levels married if the individual was married at the time of release and not married if he was not.
paro	a factor coded yes if the individual was released on parole and no if he was not.
prio	number of prior convictions.
educ	education, a categorical variable coded numerically, with codes 2 (grade 6 or less), 3 (grades 6 through 9), 4 (grades 10 and 11), 5 (grade 12), or 6 (some post-secondary).
emp1 – emp52	factors coded yes if the individual was employed in the corresponding week of the study and no otherwise.

1. Lire les données dans le `data.frame`, se familiariser avec les données.

```
> data = read.table("data_Rossi.txt")
```

2. Analyser les données. Quels sont les facteurs qui influence le recidivisme ?

3 Données sur les divorces

The unit of observation is the couple and the event of interest is divorce, with interview and widowhood treated as censoring events. We have three fixed covariates : education of the husband and two indicators of the couple's ethnicity : whether the husband is black and whether the couple is mixed. The variables are described in Table 2.

TABLE 2 – Variables disponibles pour les données sur les divorces

Var	Description
id	a couple number.
heduc	education of the husband, coded 0 = less than 12 years, 1 = 12 to 15 years, and 2 = 16 or more years.
heblack	coded 1 if the husband is black and 0 otherwise.
mixed	coded 1 if the husband and wife have different ethnicity (defined as black or other), 0 otherwise.
years	duration of marriage, from the date of wedding to divorce or censoring (due to widowhood or interview).
div	the indicator for divorce (coded 1 for divorce and 0 otherwise)

1. Lire les données dans le `data.frame`, se familiariser avec les données.

```
> divorce = read.table(file = "divorce.raw.txt", sep = "", header = FALSE)
> names(divorce) = c("id", "heduc", "heblack", "mixed", "years",
+   "div")
```

2. Analyser les données. Qui divorce ?

4 Un exemple d'analyse (données Rossi)

```
> # Lecture de données
> data= read.table("data_Rossi.txt")
> # Description de survie (Kaplan-Meier)
> plot(survfit(Surv(week, arrest)~fin, data=data), ylim=c(0.7, 1), col=c("blue", "red"))
> plot(survfit(Surv(week, arrest)~mar, data=data), ylim=c(0.7, 1), col=c("blue", "red"))
> ...
> # Modèle de Cox semi-paramétrique
> model=coxph(Surv(week, arrest)~fin+age+race+wexp+mar+paro+prio, data=data)
> # Sélection automatique des variables
> model1=step(model, trace=FALSE)
> summary(model1)
> # Survie estimée
> plot(survfit(model1), ylim=c(0.7, 1))
> # Analyse des résidus/vérification d'hypothèse PH
> cox.zph(model1)
> plot(model1$residuals)
> abline(h=0)
> par(mfrow=c(2,2))
> for(i in 1:4){
+ plot(cox.zph(model1)[i], main=rownames((cox.zph(model1))$table)[i])
+ abline(h=0, col="red")
+ }
```