



Données censurées, survie : vocabulaire et approches descriptives



- La durée jusqu'à survenue d'un évènement

Qu'est-ce qu'une donnée censurée ?

Introduction à la statistique avec R > Données censurées : introduction



- La durée jusqu'à survenue d'un évènement
- Les durées sont en général « censurées »

- La durée jusqu'à survenue d'un évènement
 - Les durées sont en général « censurées »
 - Que vaut-il mieux étudier ?
 - La durée jusqu'à survenue de l'évènement
 - Le pourcentage de survenue de l'évènement

- Survie (en fait pas toujours de la « survie » au sens propre du terme)

- Survie
- Les censures :
 - Les exclus vivants



- Survie
- Les censures :
 - Les exclus vivants
 - Les perdus de vue



- Survie
- Les censures :
 - Les exclus vivants
 - Les perdus de vue
- La fonction de survie $S(t)$ (pourcentage de « survivants » au cours du temps)

- Survie
- Les censures :
 - Les exclus vivants
 - Les perdus de vue
- La fonction de survie $S(t)$
- Le risque « instantané » de décès $h(t)$

La méthode de Kaplan-Meier

Introduction à la statistique avec R > Données censurées : introduction



Taux de rémission

1
0,8
0,6
0,4
0,2
0

A la 10^{ème} semaine : 8 rémissions sur 21 patients
→ Taux de rémission = $8/21 = 0,381$

Rémission en semaines



La méthode de Kaplan-Meier

Introduction à la statistique avec R > Données censurées : introduction



Taux de rémission

1

0,8

0,6

0,4

0,2

0

Rémission en semaines

A la fin de l'étude, plus de 40% des patients sont encore vivants



La méthode de Kaplan-Meier

Introduction à la statistique avec R > Données censurées : introduction



Taux de rémission

1

0,8

0,6

0,4

0,2

0

Tout les patients ne sont pas suivis sur la même durée. Ainsi, au fur et à mesure que l'essai se déroule, le numérateur et le dénominateur du pourcentage de survivants changent.

0

10

20

30

Rémission en semaines



```
> alc <- read.csv2("D:/MOOC/Data/alcool.csv")
```

```
> alc <- read.csv2("D:/MOOC/Data/alcool.csv")
> str(alc)
'data.frame':  125 obs. of  5 variables:
 $ t      : int  121 121 40 39 66 64 5 30 34 5 ...
 $ SEVRE  : int   0 0 0 0 0 0 1 0 0 0 ...
 $ AGE    : int  53 52 45 48 45 42 35 35 41 37 ...
 $ SEXE   : int   1 2 2 1 1 1 1 1 1 ...
 $ EDVNEG: int   0 0 0 1 0 0 0 0 0 0 ...
```

```
> alc <- read.csv2("D:/MOOC/Data/alcool.csv")
> str(alc)
'data.frame': 125 obs. of 5 variables:
 $ t : int 121 121 40 39 66 64 5 30 34 5 ...
 $ SEVRE : int 0 0 0 0 0 0 1 0 0 0 ...
 $ AGE : int 53 52 45 48 45 42 35 35 41 37 ...
 $ SEXE : int 1 2 2 1 1 1 1 1 1 ...
 $ EDVNEG: int 0 0 0 1 0 0 0 0 0 0 ...
```

```
> alc <- read.csv2("D:/MOOC/Data/alcool.csv")
> str(alc)
'data.frame':  125 obs. of  5 variables:
 $ t      : int  121 121 40 39 66 64 5 30 34 5 ...
 $ SEVRE  : int  0 0 0 0 0 0 1 0 0 0 ...
 $ AGE    : int  53 52 45 48 45 42 35 35 41 37 ...
 $ SEXE   : int  1 2 2 1 1 1 1 1 1 ...
 $ EDVNEG: int  0 0 0 1 0 0 0 0 0 0 ...
```

```
> alc <- read.csv2("D:/MOOC/Data/alcool.csv")
> str(alc)
'data.frame':  125 obs. of  5 variables:
 $ t      : int  121 121 40 39 66 64 5 30 34 5 ...
 $ SEVRE  : int   0 0 0 0 0 0 1 0 0 0 ...
 $ AGE    : int   53 52 45 48 45 42 35 35 41 37 ...
 $ SEXE   : int   1 2 2 1 1 1 1 1 1 ...
 $ EDVNEG: int   0 0 0 1 0 0 0 0 0 0 ...
```

```
> alc <- read.csv2("D:/MOOC/Data/alcool.csv")
> str(alc)
'data.frame':  125 obs. of  5 variables:
 $ t      : int  121 121 40 39 66 64 5 30 34 5 ...
 $ SEVRE  : int  0 0 0 0 0 0 1 0 0 0 ...
 $ AGE    : int  53 52 45 48 45 42 35 35 41 37 ...
 $ SEXE  : int  1 2 2 1 1 1 1 1 1 ...
 $ EDVNEG: int  0 0 0 1 0 0 0 0 0 0 ...
```

```
> alc <- read.csv2("D:/MOOC/Data/alcool.csv")
> str(alc)
'data.frame': 125 obs. of 5 variables:
 $ t      : int 121 121 40 39 66 64 5 30 34 5 ...
 $ SEVRE  : int 0 0 0 0 0 0 1 0 0 0 ...
 $ AGE    : int 53 52 45 48 45 42 35 35 41 37 ...
 $ SEXE   : int 1 2 2 1 1 1 1 1 1 ...
 $ EDVNEG: int 0 0 0 1 0 0 0 0 0 0 ...
```

```
> library(survival)
```

Variable censurée étudiée :

« Délais jusqu'à rechute de la maladie alcoolique »

La méthode de Kaplan-Meier

Introduction à la statistique avec R > Données censurées : introduction



```
> library(survival)
> plot(survfit(Surv(alc$t,alc$SEVRE)~1),
  main="Courbe de maintien dans l'abstinence")
```

La méthode de Kaplan-Meier

Introduction à la statistique avec R > Données censurées : introduction



```
> library(survival)
> plot(survfit(Surv(alc$t,alc$SEVRE)~1),
  main="Courbe de maintien dans l'abstinence")
```

La méthode de Kaplan-Meier

Introduction à la statistique avec R > Données censurées : introduction



```
> library(survival)
> plot(survfit(Surv(alc$t,alc$SEVRE)~1),
      main="Courbe de maintien dans l'abstinence")
```

La méthode de Kaplan-Meier

Introduction à la statistique avec R > Données censurées : introduction



```
> library(survival)
> plot(survfit(Surv(alc$t, alc$SEVRE)~1),
      main="Courbe de maintien dans l'abstinence")
```

La méthode de Kaplan-Meier

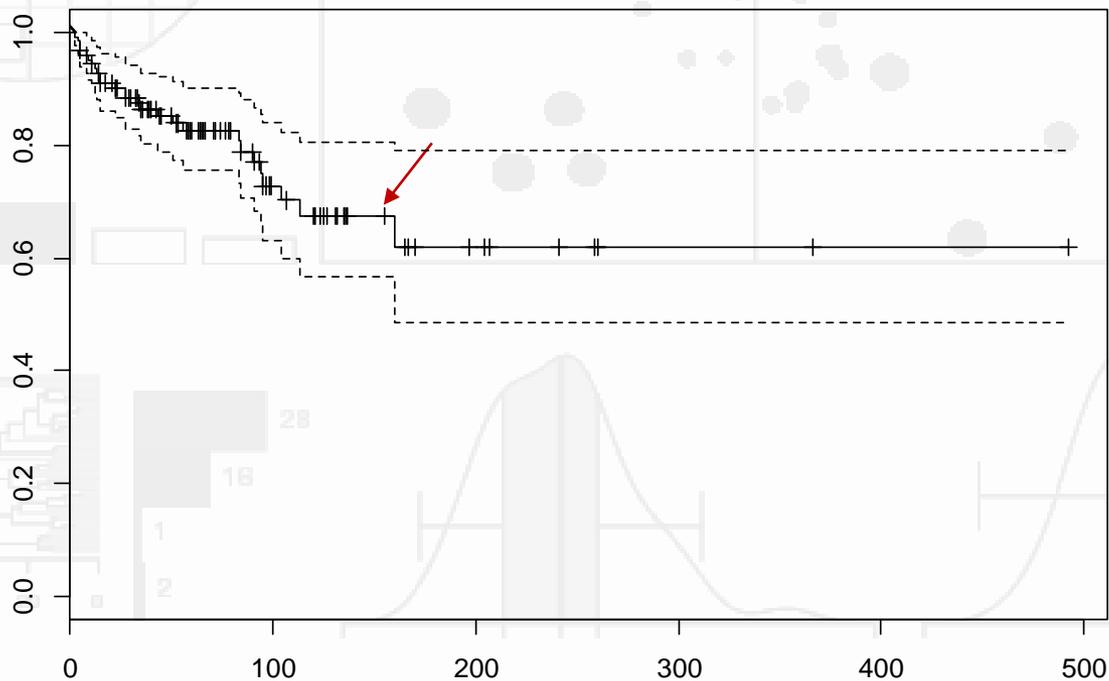
Introduction à la statistique avec R > Données censurées : introduction



```
> library(survival)
> plot(survfit(Surv(alc$t, alc$SEVRE)~1),
      main="Courbe de maintien dans l'abstinence")
```

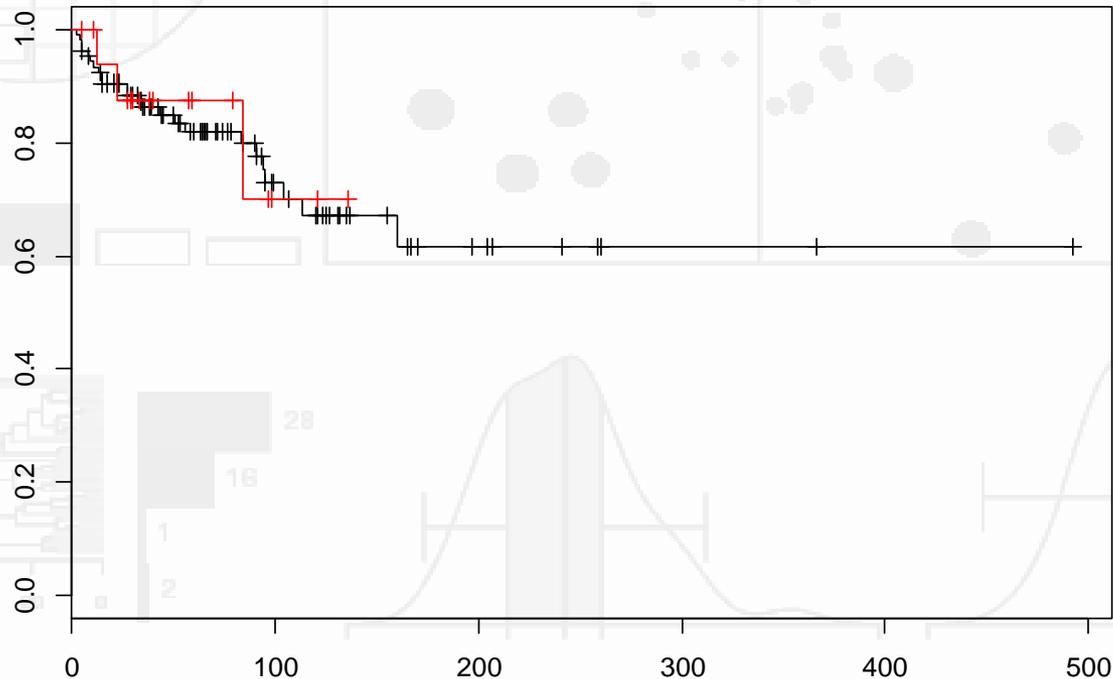
```
> library(survival)
> plot(survfit(Surv(alc$t, alc$SEVRE)~1),
      main="Courbe de maintien dans l'abstinence")
```

Courbe de maintien dans l'abstinence



```
> plot(survfit(Surv(t, SEVRE) ~ SEXE, data=alc),  
      col=c("black", "red"),  
      main="Courbe de maintien dans l'abstinence")
```

Courbe de maintien dans l'abstinence



Médiane de « survie »

Introduction à la statistique avec R > Données censurées : introduction



```
> survfit(Surv(alc$t, alc$SEVRE) ~ 1)
```

```
Call: survfit(formula = Surv(alc$t, alc$SEVRE) ~ 1)
```

records	n.max	n.start	events	median	0.95LCL	0.95UCL
125	125	125	27	NA	160	NA



Médiane de « survie »

Introduction à la statistique avec R > Données censurées : introduction



```
> survfit(Surv(alc$t, alc$SEVRE) ~ 1)
Call: survfit(formula = Surv(alc$t, alc$SEVRE) ~ 1)

records      n.max n.start  events  median 0.95LCL 0.95UCL
      125      125    125      27      NA      160      NA
```



```
alc <- read.csv2("D:/MOOC/Data/alcool.csv")
library(survival)
plot(survfit(Surv(alc$t, alc$SEVRE)~1),
      main="Courbe de maintien dans l'abstinence")
plot(survfit(Surv(t, SEVRE)~SEXE, data=alc),
      col=c("black", "red"), main="Courbe de maintien
      dans l'abstinence")
survfit(Surv(alc$t, alc$SEVRE)~1)
```