

Durées de survie

Sarah Lemler

Laboratoire "Statistique et Génome"

ENSIIE

1 Généralités sur l'analyse de survie, analyse de données de survie

- Contexte
- Les données censurées
- Exemples
- Les outils spécifiques aux données censurées

2 Méthodes non-paramétriques

- Estimateur de Kaplan-Meier
- Application aux données de Freireich
- D'autres estimateurs obtenus à l'aide de l'estimateur de Kaplan-Meier
- Tests d'égalités de fonctions de survie
- Avec le logiciel R : Kaplan-Meier et tests

3 Le modèle de Cox

- 1 Généralités sur l'analyse de survie, analyse de données de survie
 - Contexte
 - Les données censurées
 - Exemples
 - Les outils spécifiques aux données censurées
- 2 Méthodes non-paramétriques
- 3 Le modèle de Cox

En médecine ou en biologie, on s'intéresse souvent à des durées :

- durée de survie de patients ayant eu un infarctus
- durée de rémission d'une leucémie aiguë
- durée de séropositivité sans symptômes de patients infectés par le VIH
- durée de fièvre chez un patient atteint de pneumonie

Exemple 1 : données de Freireich (1963)

Durée de rémission (en semaines) obtenue par des stéroïdes chez des patients atteints de leucémie aiguë, traités soit par placebo soit par 6-mercaptopurine (6-MP).

6-MP	6	6	6	6 ⁺	7	9 ⁺	10	10 ⁺	11 ⁺	13		
	16	17 ⁺	19 ⁺	20 ⁺	22	23	25 ⁺	32 ⁺				
	32 ⁺	34 ⁺	35 ⁺									
Placebo	1	1	2	2	3	4	4	5	5	8	8	8
	8	11	11	12	12	15	17	22	23			

Le signe ⁺ correspond à des patients qui ont quitté l'étude à la date considérée. Pour ces individus les durées donc **censurées**. Par exemple le 4^{ème} patient est perdu de vue au bout de 6 semaines de traitement avec le 6-MP : il a donc une durée de rémission supérieure à 6 semaines.

Exemple 2 : données de Peto (1979)

Durées de survie de deux groupes de patients à qui l'on a administré deux types de traitement. Information supplémentaire : la fonction rénale connue pour influencer sur la survie.

durée de survie	groupe de traitement	fonction rénale	durée de survie	groupe de traitement	fonction rénale
8	1	A	220	1	N
8	1	N	365*	1	N
13	2	A	632	2	N
18	2	A	700	2	N
23	2	A	852*	1	N
52	1	A	1296	2	N
63	1	A	1296*	1	N
63	1	A	1328*	1	N
70	2	N	1460*	1	N
76	2	N	1976*	1	N
180	2	N	1990*	2	N
195	2	N	2240*	2	N
210	2	N			

Note : * : durée non observée. Les données avec * sont **censurées**.

N : fonction rénale normale ; A : fonction rénale anormale.

1 Généralités sur l'analyse de survie, analyse de données de survie

2 Méthodes non-paramétriques

- Estimateur de Kaplan-Meier
- Application aux données de Freireich
- D'autres estimateurs obtenus à l'aide de l'estimateur de Kaplan-Meier
- Tests d'égalités de fonctions de survie
- Avec le logiciel R : Kaplan-Meier et tests

3 Le modèle de Cox

Méthodes non-paramétriques

Calcul pour le groupe Placebo (sans censure)

semaine i	nombre de rémissions à la semaine i	proportion de rémissions à la semaine i
0	21	1
1	19	$19/21 = 0,90$
2	17	$17/21 = 0,81$
3	16	0,76
4	14	0,67
5	12	0,57
8	8	0,38
11	6	0,29
12	4	0,19
15	3	0,14
17	2	0,10
22	1	0,05
23	0	0,00

Méthodes non-paramétriques

Calcul pour le groupe 6-MP (avec données censurées)

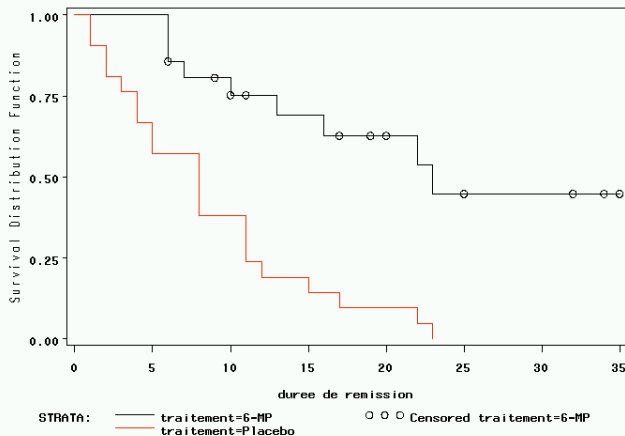
durée de rémission observées	rechutes observées en la semaine i	sujets en rémission au début de la semaine i	prob. de ne pas rechuter à la semaine i sachant qu'on est en rémission à la semaine $(i - 1)$	prob. d'être en rémission à la semaine i
0	0	21	$21/21 = 1$	1
6	3	21	$18/21 = 0,857$	$1 * 18 / 21 = 0,857$
7	1	17	$16/17 = 0,941$	$0,857 * 16/17 = 0,807$
10	1	15	$14/15 = 0,933$	$0,807 * 14/15 = 0,753$
13	1	12	$11/12 = 0,917$	$0,753 * 11/12 = 0,690$
16	1	11	$10/11 = 0,909$	$0,690 * 10/11 = 0,627$
22	1	7	$6/7 = 0,857$	$0,627 * 6/7 = 0,538$
23	1	6	$5/6 = 0,833$	$0,538 * 5/6 = 0,448$

Méthodes non-paramétriques

Courbes de survie

La courbe de survie est la représentation graphique de l'estimateur de la fonction de survie

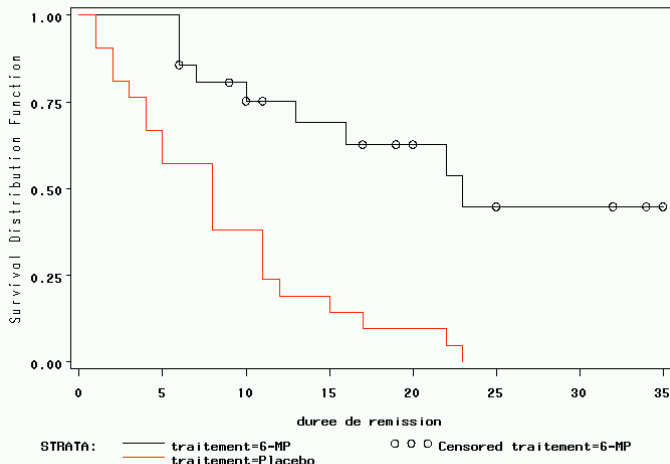
Durees de remission avant rechute de patients atteints de leucemie



Tests de comparaison

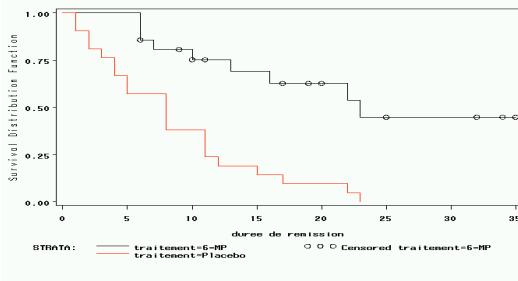
On souhaite comparer la survie de **deux groupes** ayant reçu des traitements différents.

Durees de remission avant rechute de patients atteints de leucemie



Application aux données de Freireich : comparaison suivant traitement

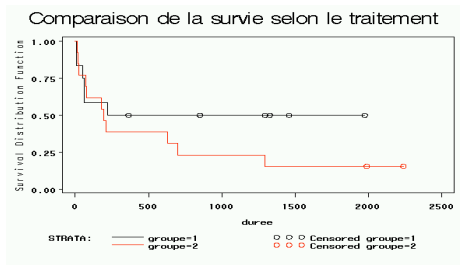
Durees de remission avant rechute de patients atteints de leucemie



test	Khi2	DF	Pr > Khi2
Log-Rank	16.8246	1	<0.001
Wilcoxon	13.5383	1	0.002

Rejet de H_0 : les 2 courbes de survie sont différentes $\Rightarrow$ effet traitement
On conclut que le 6-MP a un effet positif sur la fonction de survie.

Application aux données de Peto : comparaison suivant traitement



Test	Khi 2	DF	Pr >Khi 2
Log-Rank	1.3126	1	0.2519
Wilcoxon	0.2490	1	0.6178

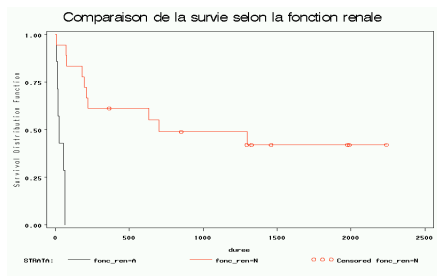
Non rejet de H_0 : les 2 courbes de survie ne semblent pas différentes
⇒ pas d'effet traitement.

Tests de comparaison

Analyse des résultats

- Dans l'exemple 1 (données de Freireich) nous concluons au rejet de H_0 : les courbes de survie dans les 2 groupes de traitement sont différentes. On conclut donc que le 6-MP a un effet (positif) sur la fonction de survie.
- Dans l'exemple 2, nous concluons au non rejet de H_0 : nous n'avons pas de raison de rejeter l'égalité des courbes de survie. Pas d'effet traitement a priori. L'effet traitement est peut-être caché par la fonction rénale. Pour cela, les deux groupes, fonction rénale anormale (A) et normale (N) vont être comparés (tous traitements confondus).

Données de Peto : tests de comparaison suivant la fonction rénale quelque soit le traitement



Test	Khi 2	DF	$Pr > Khi2$
Log-Rank	24.0188	1	< .0001
Wilcoxon	20.8739	1	< .0001

Rejet de H_0 : les 2 courbes de survie sont différentes
⇒ influence de la fonction rénale.

Tests de comparaison

Données de Peto : conclusions sur la fonction rénale

→ Ces résultats peuvent faire penser que la fonction rénale masque l'effet du traitement. **C'est ce qu'il faut tenter de vérifier.**

→ Problème : comment relier la durée de survie d'un patient à **plusieurs facteurs pronostiques** : traitement et fonction rénale.

Avec le logiciel R : méthode de Kaplan-Meier I

```
> library(survival)
> freireich<-read.table("Datafreireich.csv", sep=";", header=T)
> head(freireich)
```

	id	groupe	temps	statut
1	1	6MP	6	1
2	2	6MP	6	1
3	3	6MP	6	1
4	4	6MP	6	0
5	5	6MP	7	1
6	6	6MP	9	0

```
> freireich.surv=Surv(freireich$temps,freireich$statut)
> fit<-survfit(freireich.surv~ groupe,data=freireich)
> summary(fit)
```

Avec le logiciel R : méthode de Kaplan-Meier II

```
Call: survfit(formula = freireich.surv ~ groupe, data = freireich)
```

groupe=6MP

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
6	21	3	0.857	0.0764	0.720	1.000
7	17	1	0.807	0.0869	0.653	0.996
10	15	1	0.753	0.0963	0.586	0.968
13	12	1	0.690	0.1068	0.510	0.935
16	11	1	0.627	0.1141	0.439	0.896
22	7	1	0.538	0.1282	0.337	0.858
23	6	1	0.448	0.1346	0.249	0.807

groupe=Placebo

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
1	21	2	0.9048	0.0641	0.78754	1.000
2	19	2	0.8095	0.0857	0.65785	0.996
3	17	1	0.7619	0.0929	0.59988	0.968
4	16	2	0.6667	0.1029	0.49268	0.902

Avec le logiciel *R* : méthode de Kaplan-Meier III

5	14	2	0.5714	0.1080	0.39455	0.828
8	12	4	0.3810	0.1060	0.22085	0.657
11	8	2	0.2857	0.0986	0.14529	0.562
12	6	2	0.1905	0.0857	0.07887	0.460
15	4	1	0.1429	0.0764	0.05011	0.407
17	3	1	0.0952	0.0641	0.02549	0.356
22	2	1	0.0476	0.0465	0.00703	0.322
23	1	1	0.0000	NaN	NA	NA

Avec le logiciel R : courbes de survie et tests de comparaison I

```
> plot(fit, lty=2:3)
> legend(6,.7,c("6-MP","Placebo"),lty=2:3)
> title(main="Kaplan Meier par traitement")
> survdiff(freireich.surv~ groupe,data=freireich,rho=0) #log-rank
```

Call:

```
survdiff(formula = freireich.surv ~ groupe, data = freireich,
  rho = 0)
```

	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
groupe=6MP	21	9	19.3	5.46	16.8
groupe=Placebo	21	21	10.7	9.77	16.8

Chisq= 16.8 on 1 degrees of freedom, p= 4.17e-05

```
> survdiff(freireich.surv~ groupe,data=freireich,rho=1) #Peto
```

Avec le logiciel *R* : courbes de survie et tests de comparaison II

Call:

```
survdifff(formula = freireich.surv ~ groupe, data = freireich,  
           rho = 1)
```

	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
groupe=6MP	21	5.12	12.00	3.94	14.5
groupe=Placebo	21	14.55	7.68	6.16	14.5

Chisq= 14.5 on 1 degrees of freedom, p= 0.000143

Plan du cours

- 1 Généralités sur l'analyse de survie, analyse de données de survie
- 2 Méthodes non-paramétriques
- 3 Le modèle de Cox