

1 Le modèle de Cox

- Généralités sur le modèle de Cox
- Interprétation du modèle de Cox
 - Modèle de Cox avec une covariable binaire
 - Modèle de Cox avec une ou plusieurs variables catégorielles
 - Modèle de Cox avec une variable quantitative
 - Modèle de Cox avec plusieurs covariables
- Estimation des paramètres
- Sélection de variables
- Validation du Modèle de Cox

Modèle de Cox avec une covariable binaire I

Avec le logiciel R : données de Freireich

```
> library(survival)
> freireich<-read.table("Datafreireich.csv", sep=";", header=T)
> freireich.surv=Surv(freireich$temps,freireich$statut)
> fit.x=coxph(Surv(temps,statut)~groupe,data=freireich,method="breslow")
> summary(fit.x)
```

Call:

```
coxph(formula = Surv(temps, statut) ~ groupe, data = freireich,
      method = "breslow")
```

n= 42

	coef	exp(coef)	se(coef)	z	Pr(> z)	
groupePlacebo	1.5092	4.5231	0.4096	3.685	0.000229	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Modèle de Cox avec une covariable binaire II

Avec le logiciel *R* : données de Freireich

	<code>exp(coef)</code>	<code>exp(-coef)</code>	lower .95	upper .95
<code>groupePlacebo</code>	4.523	0.2211	2.027	10.09

`Rsquare= 0.304` (max possible= 0.989)

Likelihood ratio test= 15.21 on 1 df, p=9.615e-05

Wald test = 13.58 on 1 df, p=0.0002288

Score (logrank) test = 15.93 on 1 df, p=6.571e-05

Modèle de Cox avec une variable catégorielle

Données kidney

Données kidney : 38 patients, avec les variables "id", "time", "status", "disease" où "disease" (type d'atteinte rénale) a 4 modalités (0=GN, 1=AN, 2=PKD, 3=Other).

2 possibilités :

- première possibilité :

$$Z = \text{disease}, Z \in \{0, 1, 2, 3\} \implies h(t|Z) = \alpha_0(t)e^{\beta Z}.$$

- deuxième possibilité

$$GN = 1 \text{ si } \text{disease} = GN, 0 \text{ sinon}$$

$$AN = 1 \text{ si } \text{disease} = AN, 0 \text{ sinon}$$

$$PKD = 1 \text{ si } \text{disease} = PKD, 0 \text{ sinon}$$

$$Other = 1 \text{ si } \text{disease} = Other, 0 \text{ sinon}$$

$$\text{et } h(t|Z) = \alpha_0(t) \exp(\beta_1 AN + \beta_2 PKD + \beta_3 Other).$$

Le paramètre $\beta = (\beta_1, \beta_2, \beta_3)^T$ est inconnu.

Modèle de Cox avec une variable catégorielle I

Avec le logiciel *R* : données kidney

Première possibilité

```
> kidney$X[kidney$disease=="GN"]=0  
> kidney$X[kidney$disease=="AN"]=1  
> kidney$X[kidney$disease=="PKD"]=2  
> kidney$X[kidney$disease=="Other"]=3
```

Deuxième possibilité

```
> kidney$AN=(kidney$disease=="AN")  
> kidney$GN=(kidney$disease=="GN")  
> kidney$PKD=(kidney$disease=="PKD")  
> kidney$Other=(kidney$disease=="Other")
```

Comparaison des deux possibilités

```
> kidney.surv=Surv(kidney$time,kidney$status)  
> coxph(Surv(time,status)~X,data=kidney,method="breslow")
```

Modèle de Cox avec une variable catégorielle II

Avec le logiciel *R* : données kidney

Call:

```
coxph(formula = Surv(time, status) ~ X, data = kidney, method = "breslow")
```

	coef	exp(coef)	se(coef)	z	p
X	-0.144	0.866	0.113	-1.27	0.2

Likelihood ratio test=1.62 on 1 df, p=0.203 n= 76

```
> coxph(Surv(time,status)~AN+PKD+Other,data=kidney,method="breslow")
```

Call:

```
coxph(formula = Surv(time, status) ~ AN + PKD + Other, data = kidney,  
      method = "breslow")
```

	coef	exp(coef)	se(coef)	z	p
ANTRUE	0.031	1.032	0.364	0.0853	0.93

Modèle de Cox avec une variable catégorielle III

Avec le logiciel R : données kidney

PKDTRUE	-0.619	0.538	0.531	-1.1674	0.24
OtherTRUE	-0.351	0.704	0.354	-0.9912	0.32

Likelihood ratio test=2.71 on 3 df, p=0.439 n= 76

```
> coxph(Surv(time,status)~disease,data=kidney,method="breslow")
```

Call:

```
coxph(formula = Surv(time, status) ~ disease, data = kidney,  
      method = "breslow")
```

	coef	exp(coef)	se(coef)	z	p
diseaseGN	0.351	1.420	0.354	0.991	0.32
diseaseAN	0.382	1.465	0.336	1.135	0.26
diseasePKD	-0.269	0.764	0.507	-0.531	0.60

Likelihood ratio test=2.71 on 3 df, p=0.439 n= 76

Modèle de Cox avec deux variables catégorielles

Données de Peto

Données de Peto (1979) : durées de survie de deux groupes de patients sous traitement.

Variable supplémentaire : la fonction rénale.

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.

Modèle de Cox avec deux variables catégorielles

Données de Peto : résultats sans interaction

Analyse des estimations de la vraisemblance maximum

Variable	DF	Estimation param.	Erreur std	Khi 2	Pr > Khi 2
groupe	1	1.24308	0.59932	4.3021	0.0381
rein	1	4.10540	1.16451	12.4286	0.0004
	Rapport	95% Limites de confiance		Libellé	
Variable	de risque	du rapport de risque		variable	
groupe	3.466	1.071 - 11.220		traitement	
rein	60.667	6.190 - 594.548		fonction rénale	

Modèle de Cox avec deux variables catégorielles

Données de Peto : résultats avec interaction

Analyse des estimations de la vraisemblance maximum

Variable	DF	Estimation param.	Erreur std	Khi 2	Pr > Khi 2
groupe	1	1.78145	1.84974	0.9275	0.3355
rein	1	4.67334	2.17557	4.6143	0.0317
grein	1	-0.38564	1.23621	0.0973	0.7551
variable	Rapport de risque	95% Limites de confiance du rapport de risque			
groupe	5.938	0.158 - 222.936			
rein	107.055	1.506 - 7611.224			
grein	0.680	0.060 - 7.670			

Validation du Modèle de Cox avec le logiciel R I

```
> base<-Surv(aml$time,aml$status)
```

On recode la variable traitement :

```
> aml$z[aml$x=="Nonmaintained"]=0
```

```
> aml$z[aml$x=="Maintained"]=1
```

Modèle de Cox :

```
> res<-coxph(base~aml$z,data=aml)
```

```
> summary(res)
```

Call:

```
coxph(formula = base ~ aml$z, data = aml)
```

n= 23

	coef	exp(coef)	se(coef)	z	Pr(> z)
aml\$z	-0.9155	0.4003	0.5119	-1.788	0.0737 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Validation du Modèle de Cox avec le logiciel R II

```
      exp(coef) exp(-coef) lower .95 upper .95  
aml$z      0.4003      2.498    0.1468    1.092
```

```
Rsquare= 0.137    (max possible= 0.976 )
```

```
Likelihood ratio test= 3.38  on 1 df,    p=0.06581
```

```
Wald test          = 3.2  on 1 df,    p=0.07371
```

```
Score (logrank) test = 3.42  on 1 df,    p=0.06454
```

Critère AIC :

```
> AIC=-2*res$loglik[2]+2*1
```

```
> AIC
```

```
[1] 84.06523
```

Validation du modèle de Cox :

Validation du Modèle de Cox avec le logiciel R III

```
> Main<-aml$x=="Maintained"  
> amlM<-aml[Main, ]  
> NonMain<-aml$x=="Nonmaintained"  
> amlNM<-aml[NonMain, ]  
> fitMain<-survfit(Surv(time,status)~1,data=amlM)  
> str(fitMain)
```

List of 13

```
$ n      : int 11  
$ time    : num [1:10] 9 13 18 23 28 31 34 45 48 161  
$ n.risk   : num [1:10] 11 10 8 7 6 5 4 3 2 1  
$ n.event  : num [1:10] 1 1 1 1 0 1 1 0 1 0  
$ n.censor : num [1:10] 0 1 0 0 1 0 0 1 0 1  
$ surv     : num [1:10] 0.909 0.818 0.716 0.614 0.614 ...  
$ type     : chr "right"  
$ std.err  : num [1:10] 0.0953 0.1421 0.1951 0.2487 0.2487 ...  
$ upper    : num [1:10] 1 1 1 1 1 ...  
$ lower    : num [1:10] 0.754 0.619 0.488 0.377 0.377 ...  
$ conf.type: chr "log"
```

Validation du Modèle de Cox avec le logiciel R IV

```
$ conf.int : num 0.95  
$ call      : language survfit(formula = Surv(time, status) ~ 1, data=  
- attr(*, "class")= chr "survfit"
```

Estimateur de Nelson-Aalen du risque cumulé dans le groupe "Maintained" :

```
> HMNA<-cumsum(fitMain$n.event/fitMain$n.risk)
```

Estimateur de Breslow du risque cumulé dans le groupe "Maintained" :

```
> HMB<-log(fitMain$surv)  
> fitNonMain<-survfit(Surv(time,status)~1,data=amlNM)
```

Estimateur de Nelson-Aalen du risque cumulé dans le groupe "Nonmaintained" :

```
> HMNA<-cumsum(fitNonMain$n.event/fitNonMain$n.risk)
```

Estimateur de Breslow du risque cumulé dans le groupe "Nonmaintained" :

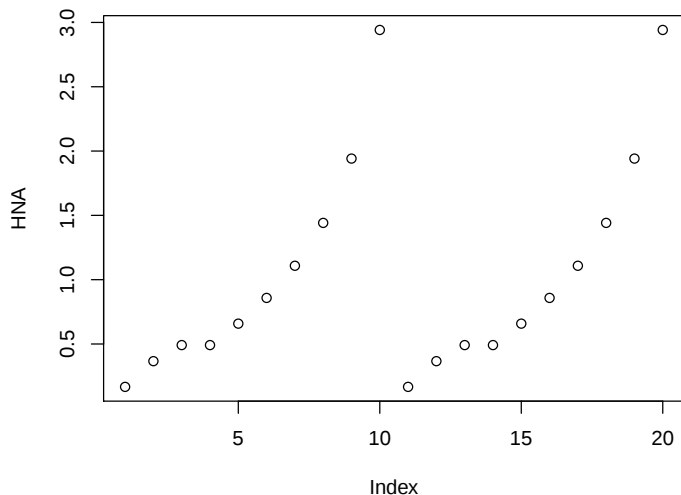
```
> HNMB<-log(fitNonMain$surv)
```

On trace les deux risques cumulés obtenus avec l'estimateur de Nelson-Aalen dans les deux groupes :

Validation du Modèle de Cox avec le logiciel R V

```
> HNA=c(HMNA, HMNA)
> plot(HNA, lty=2:3)
> title(main="Risques cumules de Nelson-Aalen par traitement")
```

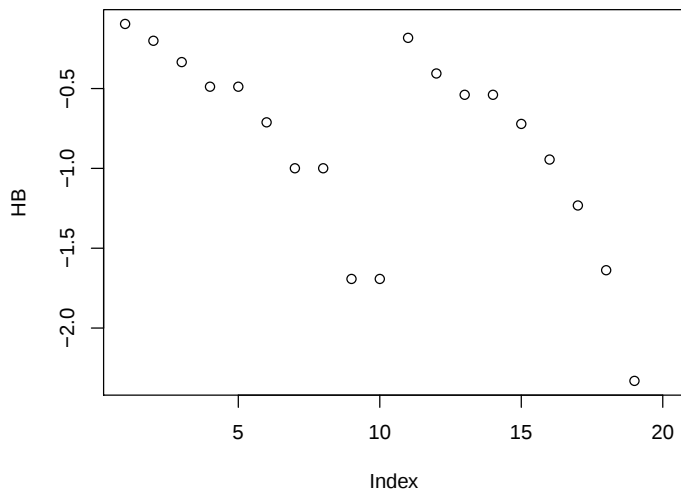
Risques cumules de Nelson–Aalen par traitement



On trace les deux risques cumulés obtenus avec l'estimateur de Breslow dans les deux groupes :

```
> HB=c(HMB,HNMB)
> plot(HB,lty=2:3)
> title(main="Risques cumules de Breslow par traitement")
```

Risques cumules de Breslow par traitement

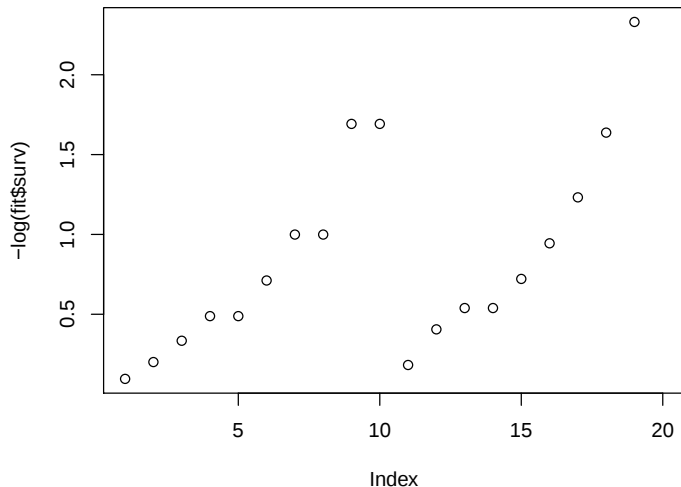


Estimateur de Breslow du risque cumulé par traitement

```
> fit<-survfit(Surv(time,status)~x,data=aml)
> plot(-log(fit$surv),lty=2:3)
> title(main="Estimateur de Breslow du risque cumule par traitement")
```

Validation du Modèle de Cox avec le logiciel R II

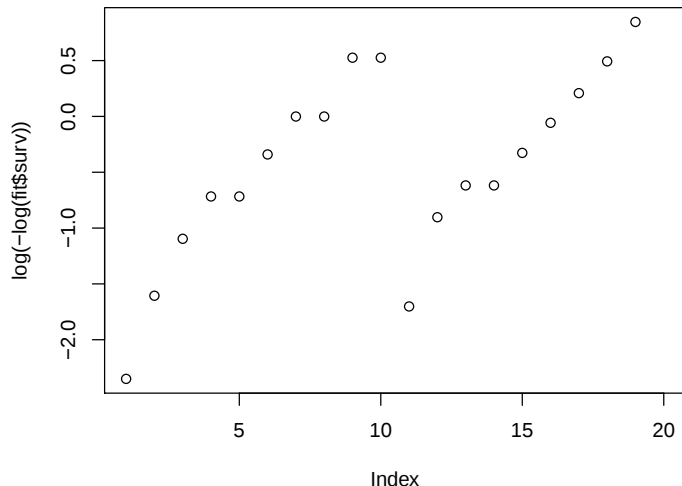
Estimateur de Breslow du risque cumule par traitement



```
> plot(log(-log(fit$surv)),lty=2:3)
> title(main="log(-log(S)) par groupe de traitement")
```

Validation du Modèle de Cox avec le logiciel R II

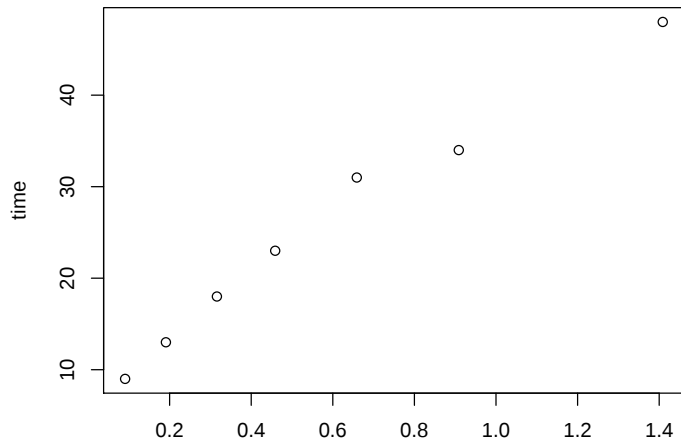
$\log(-\log(S))$ par groupe de traitement



Egalité des risques de base :

```
> rbaseM<-basehaz(coxph(Surv(time,status)~1,data=amlM))  
> rbaseNM<-basehaz(coxph(Surv(time,status)~1,data=amlNM))  
> plot(rbaseM,rbaseNM)
```

Validation du Modèle de Cox avec le logiciel R II



i3535353535, 0.677020202020202, 0.877020202020202, 1.12702020202020, 1.4
c(5, 8, 12, 23, 27, 30, 33, 43, 45)

Validation du Modèle de Cox avec le logiciel R I

Test de proportionnalité :

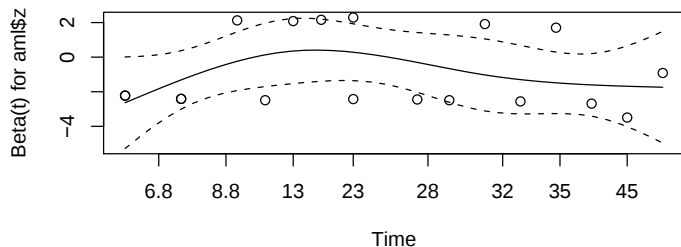
```
> cox.zph(res)
```

```
          rho    chisq      p  
aml$z -0.0198 0.00691 0.934
```

```
> par(mfrow=c(2,1))
```

```
> plot(cox.zph(res))
```

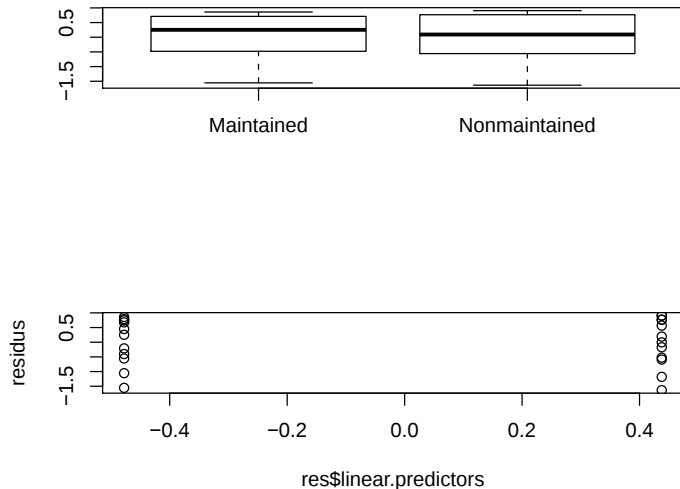
Validation du Modèle de Cox avec le logiciel R II



Vérification de l'hypothèse de linéarité :

```
> residus<-residuals(res,type="martingale")  
> par(mfrow=c(2,1))  
> plot(aml$x,residus)  
> plot(res$linear.predictors,residus)
```

Validation du Modèle de Cox avec le logiciel R II



Validation du Modèle de Cox avec le logiciel R I

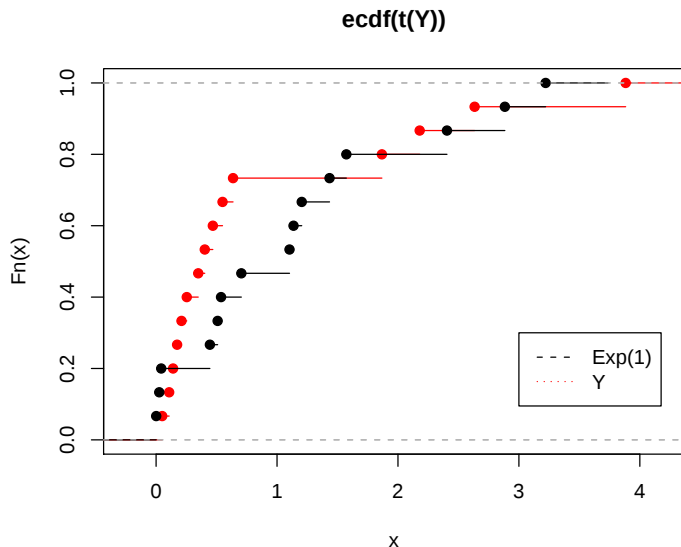
Résidus de Cox-Snell

```
> rbasec<-basehaz(coxph(Surv(time,status)~x,data=aml))
> Y<-rbasec[1]*exp(res$linear.predictors)
> plot(ecdf(t(Y)),col='red')
> E<-rexp(15, rate = 1)
> lines(ecdf(t(E)))
> legend(3,0.3,c("Exp(1)", "Y"),lty=2:3, col=c(1,2))
> ks.test(t(Y),rexp(15,rate=1))
```

Two-sample Kolmogorov-Smirnov test

```
data:  t(Y) and rexp(15, rate = 1)
D = 0.1333, p-value = 0.9998
alternative hypothesis: two-sided
```

Validation du Modèle de Cox avec le logiciel R II



Validation du Modèle de Cox avec le logiciel R I

Algorithmes de comparaison de modèles :

```
> library(MASS)
> AIC<-stepAIC(res,direction="both", trace=T,k=2)
```

```
Start:  AIC=84.07
base ~ aml$z
```

	Df	AIC
<none>		84.065
- aml\$z	1	85.450

```
> BIC<-stepAIC(res,direction="backward",trace=T,k=log(23))
```

```
Start:  AIC=85.2
base ~ aml$z
```

	Df	AIC
<none>		85.201
- aml\$z	1	85.450