

Modèles de Durée

Examen du 18 novembre 2014

Instructions : *Durée 1h45 – Les documents sont interdits. La qualité de la présentation sera prise en compte dans la notation.*

On s'intéresse à un essai clinique sur le traitement du cancer du poumon. Cette base de données contient les informations suivantes sur $n = 137$ individus :

- *trait* : (traitement) avec 0=standard 1=test
- *celltype* : squamous, smallcell, adeno, large
- *cellule* : codage des valeurs de *celltype* par 1, 2, 3, et 4
- *time* : survival time
- *status* : censoring status (0=censurée)
- *karno* : Karnofsky performance score (100=good)
- *diagtime* : months from diagnosis to randomisation
- *age* : in years
- *prior* : prior therapy 0=no, 1=yes

Dans la suite on note T_i le temps de suivi (temps de survie) du patient i , C_i la censure du patient i et δ_i l'indicateur de censure. Pour chaque patient i , $i = 1, \dots, n$, on observe

$$X_i = \min(T_i, C_i), \text{ et } \delta_i = \mathbb{1}_{T_i \leq C_i}, \text{ où } \mathbb{1}_{T_i \leq C_i} = 1 \text{ si } T_i \leq C_i \text{ et } \mathbb{1}_{T_i \leq C_i} = 0 \text{ sinon.}$$

On note également $Z^{(i)} = (Z_1^{(i)}, \dots, Z_p^{(i)})^T$ le vecteur des p covariables de l'individu i , pour $i = 1, \dots, n$. Le programme *R* se trouve à la fin de l'énoncé, suivi des résultats.

1. (a) Donner la définition de la fonction de survie S , du risque instantané, h , et du risque cumulé H ainsi que les relations qui les lient.
- (b) Considérons $\hat{S}(t) = \frac{1}{\sum_{i=1}^n \delta_i} \sum_{i=1}^n \mathbb{1}_{X_i > t, \delta_i = 1}$. Expliquer ce que représente cette quantité.
- (c) Donner la définition de l'estimateur de Kaplan-Meier, noté $\hat{S}_{KM}$.
- (d) On s'intéresse dans un premier temps à l'influence du traitement sur la survie. Quel(s) test(s) proposez-vous ? Donner les hypothèses de(s) test(s) et préciser simplement la différence entre ces tests (les statistiques de test ne sont pas exigées).
- (e) On s'intéresse maintenant à l'estimation des fonctions de survie dans les quatre groupes définis par la variable *celltype*. Commenter le programme et les résultats obtenus. Que concluez-vous ?
- (f) Quel test proposeriez-vous pour étudier la répartition des traitements par type de cellule (variable *celltype*) ? Préciser le test mis en oeuvre dans le programme. Que cherche-t-on à mettre en évidence. Commenter.
2. (a) On propose maintenant de modéliser la relation entre la variable *temps de survie* et le type de traitement (variable *trait*) par un modèle de Cox. Ecrire le modèle de Cox correspondant.
- (b) Donner la définition du rapport des risques (ou hazard ratio) comparant le groupe de traitement *test* au groupe de traitement *standard*. Donner son interprétation en fonction du (des) paramètre(s) du modèle de Cox.

- (c) Donner l'estimation du rapport de risque instantané comparant le groupe de traitement *test* au groupe de traitement *standard*. Commenter.
 - (d) Donner les hypothèses de test de l'influence de la variable *trait* sur le temps de survie et proposer trois tests de l'influence de la variable *trait* sur le temps de survie.
 - (e) Commenter le programme correspondant et les résultats obtenus. Que concluez-vous ?
 - (f) Donner l'intervalle de confiance à 95% pour le rapport des risques comparant le groupe de traitement *test* au groupe de traitement *standard*. Commenter.
 - (g) Comparer l'un des trois tests précédents avec le test effectué lors de la comparaison des fonctions de survie dans les deux groupes de traitements.
3. (a) On propose maintenant de modéliser la relation entre la variable *temps de survie* et le type de cellule (*cellule*), par un modèle de Cox. Ecrire le modèle de Cox correspondant en rappelant la définition des quantités qui interviennent dans la modélisation, ainsi que l'interprétation des coefficients.
 - (b) Commenter la pertinence du codage de la variable *cellule*. Quel codage proposeriez-vous ? Justifier.
 - (c) Commenter les résultats obtenus pour le modèle de Cox avec la variable *celltype*.
 4. (a) Ecrire le modèle de Cox complet (avec toutes les variables explicatives). Discuter les effets des différentes variables.
 - (b) Donner l'estimation du rapport de risque instantané entre deux patients qui ne diffèrent que par la variable *prior*. Commenter.
 - (c) Donner l'interprétation du rapport de risque instantané entre deux patients qui ne diffèrent que par la variable *age*. Commenter.
 - (d) Quel codage proposeriez-vous pour la variable *age* ? Justifier.
 - (e) Préciser (en le justifiant) quelles sont les variables qui influent sur la survie de ces patients.
 5. Proposer (en la justifiant) une méthode graphique permettant d'étudier la pertinence de l'hypothèse de proportionalité portant sur la variable *trait* ?

Programme et résultats :

```
> library(ISwR)
> library(survival)
> library(MASS)
> veteran$cellule[veteran$celltype=="squamous"]=1
> veteran$cellule[veteran$celltype=="smallcell"]=2
> veteran$cellule[veteran$celltype=="adeno"]=3
> veteran$cellule[veteran$celltype=="large"]=4
> veteran$prior[veteran$prior==0]=0
> veteran$prior[veteran$prior==10]=1
> veteran$trait[veteran$trt==1]=0
> veteran$trait[veteran$trt==2]=1
> veteran$age1=(veteran$age<40)
> veteran$age2=(veteran$age>=40 & veteran$age<60)
> veteran$age3=(veteran$age>=60)
> head(veteran)
```

	trt	celltype	time	status	karno	diagtime	age	prior	cellule	prior	trait	age1
1	1	squamous	72	1	60	7	69	0	1	0	0	FALSE
2	1	squamous	411	1	70	5	64	10	1	1	0	FALSE
3	1	squamous	228	1	60	3	38	0	1	0	0	TRUE
4	1	squamous	126	1	60	9	63	10	1	1	0	FALSE
5	1	squamous	118	1	70	11	65	10	1	1	0	FALSE
6	1	squamous	10	1	20	5	49	0	1	0	0	FALSE

```

    age2 age3
1 FALSE TRUE
2 FALSE TRUE
3 FALSE FALSE
4 FALSE TRUE
5 FALSE TRUE
6 TRUE FALSE

> names(veteran)

[1] "trt"      "celltype" "time"      "status"    "karno"     "diagtime"
[7] "age"      "prior"     "cellule"   "priort"    "trait"     "age1"
[13] "age2"     "age3"

> attach(veteran)
> base=Surv(time,status==1)
> base

  [1] 72 411 228 126 118 10 82 110 314 100+ 42 8 144 25+ 11
 [16] 30 384 4 54 13 123+ 97+ 153 59 117 16 151 22 56 21
 [31] 18 139 20 31 52 287 18 51 122 27 54 7 63 392 10
 [46] 8 92 35 117 132 12 162 3 95 177 162 216 553 278 12
 [61] 260 200 156 182+ 143 105 103 250 100 999 112 87+ 231+ 242 991
 [76] 111 1 587 389 33 25 357 467 201 1 30 44 283 15 25
 [91] 103+ 21 13 87 2 20 7 24 99 8 99 61 25 95 80
[106] 51 29 24 18 83+ 31 51 90 52 73 8 36 48 7 140
[121] 186 84 19 45 80 52 164 19 53 15 43 340 133 111 231
[136] 378 49

> n=dim(veteran)[1]
> n

[1] 137

> ind=status==1
> sum(ind)/n

[1] 0.9343066

> fittrait=survfit(base~trait,data=veteran)
> plot(fittrait,lty=c(1,2),col=c(1,2))
> legend(200,0.4,c("Standard","Test"),lty=c(1,2),col=c(1,2))
> title(main="Courbes de survie par groupe de traitement")
> survdiff(base~trait,data=veteran,rho=1)

Call:
survdiff(formula = base ~ trait, data = veteran, rho = 1)

      N Observed Expected (O-E)^2/E (O-E)^2/V
trait=0 69 32.2 35.4 0.279 0.871
trait=1 68 35.2 32.1 0.308 0.871

Chisq= 0.9 on 1 degrees of freedom, p= 0.351

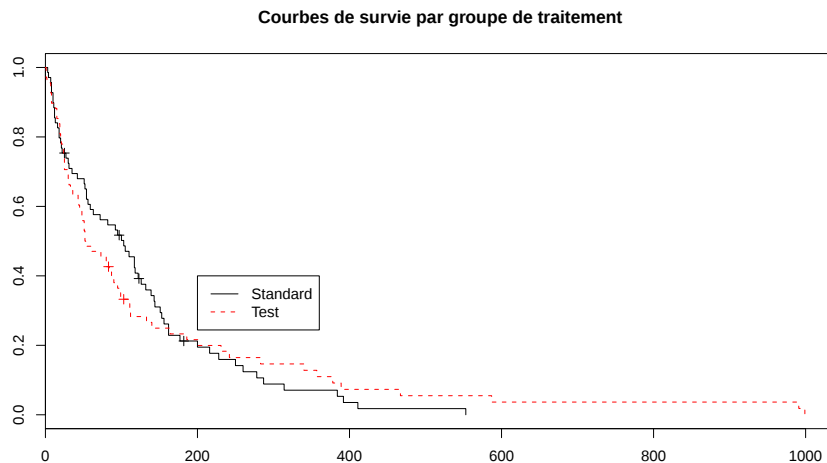
> survdiff(base~trait,data=veteran,rho=0)

Call:
survdiff(formula = base ~ trait, data = veteran, rho = 0)

```

	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
trait=0	69	64	64.5	0.00388	0.00823
trait=1	68	64	63.5	0.00394	0.00823

Chisq= 0 on 1 degrees of freedom, p= 0.928



```
> fitcell=survfit(base~celltype,data=veteran)
> plot(fitcell,lty=c(1,2,3,4),col=c(1,2,3,4))
> legend(400,0.7,c("squamous","smallcell","adeno","large"),lty=c(1,2,3,4),col=c(1,2,3,4))
> title(main="Courbes de survie par type de cellules")
> survdiff(base~celltype,data=veteran,rho=1)
```

Call:

```
survdiff(formula = base ~ celltype, data = veteran, rho = 1)
```

	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
celltype=squamous	35	13.39	20.2	2.27	5.05
celltype=smallcell	48	28.43	19.0	4.67	9.33
celltype=adeno	27	16.07	10.8	2.56	4.23
celltype=large	27	9.56	17.5	3.59	7.57

Chisq= 19.7 on 3 degrees of freedom, p= 0.000195

```
> survdiff(base~celltype,data=veteran,rho=0)
```

Call:

```
survdiff(formula = base ~ celltype, data = veteran, rho = 0)
```

	N	Observed	Expected	$(O-E)^2/E$	$(O-E)^2/V$
celltype=squamous	35	31	47.7	5.82	10.53
celltype=smallcell	48	45	30.1	7.37	10.20
celltype=adeno	27	26	15.7	6.77	8.19
celltype=large	27	26	34.5	2.12	3.02

Chisq= 25.4 on 3 degrees of freedom, p= 1.27e-05

```
> chisq.test(trait,celltype,correct=FALSE)
```

Pearson's Chi-squared test

data: trait and celltype

X-squared = 7.0407, df = 3, p-value = 0.07061

```
> chisq.test(trait, celltype, correct=FALSE)$expected
```

```
      celltype
trait squamous smallcell      adeno      large
      0 17.62774 24.17518 13.59854 13.59854
      1 17.37226 23.82482 13.40146 13.40146
```

```
> chisq.test(trait, celltype, correct=FALSE)$observed
```

```
      celltype
trait squamous smallcell adeno large
      0      15      30      9      15
      1      20      18     18     12
```

```
> coxtrait=coxph(base~trait, data=veteran, method="exact")
```

```
> summary(coxtrait)
```

Call:

```
coxph(formula = base ~ trait, data = veteran, method = "exact")
```

n= 137

```
      coef exp(coef) se(coef)      z Pr(>|z|)
trait 0.01644  1.01658  0.18129 0.091  0.928
```

```
      exp(coef) exp(-coef) lower .95 upper .95
trait      1.017      0.9837  0.7126  1.450
```

Rsquare= 0 (max possible= 0.999)

Likelihood ratio test= 0.01 on 1 df, p=0.9277

Wald test = 0.01 on 1 df, p=0.9277

Score (logrank) test = 0.01 on 1 df, p=0.9277

```
> coxcellule=coxph(base~cellule, data=veteran)
```

```
> summary(coxcellule)
```

Call:

```
coxph(formula = base ~ cellule, data = veteran)
```

n= 137

```
      coef exp(coef) se(coef)      z Pr(>|z|)
cellule 0.05447  1.05598  0.07403 0.736  0.462
```

```
      exp(coef) exp(-coef) lower .95 upper .95
cellule      1.056      0.947  0.9134  1.221
```

Rsquare= 0.004 (max possible= 0.999)

Likelihood ratio test= 0.54 on 1 df, p=0.4629

Wald test = 0.54 on 1 df, p=0.4618

Score (logrank) test = 0.54 on 1 df, p=0.4616

```
> coxcelltype=coxph(base~celltype, data=veteran)
```

```
> summary(coxcelltype)
```

Call:

```
coxph(formula = base ~ celltype, data = veteran)
```

n= 137

	coef	exp(coef)	se(coef)	z	Pr(> z)
celltypesmallcell	1.0013	2.7217	0.2535	3.950	7.83e-05 ***
celltypeadeno	1.1477	3.1510	0.2929	3.919	8.90e-05 ***
celltypelarge	0.2301	1.2588	0.2773	0.830	0.407

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

	exp(coef)	exp(-coef)	lower .95	upper .95
celltypesmallcell	2.722	0.3674	1.656	4.473
celltypeadeno	3.151	0.3174	1.775	5.594
celltypelarge	1.259	0.7944	0.731	2.168

Rsquare= 0.166 (max possible= 0.999)

Likelihood ratio test= 24.85 on 3 df, p=1.661e-05

Wald test = 24.09 on 3 df, p=2.387e-05

Score (logrank) test = 25.51 on 3 df, p=1.208e-05

```
> coxcomplet=coxph(base~trait+celltype+karno+age2+age3+priort,data=veteran)
```

```
> summary(coxcomplet)
```

Call:

```
coxph(formula = base ~ trait + celltype + karno + age2 + age3 +  
      priort, data = veteran)
```

n= 137

	coef	exp(coef)	se(coef)	z	Pr(> z)
trait	0.266665	1.305603	0.199148	1.339	0.180561
celltypesmallcell	0.812054	2.252529	0.268389	3.026	0.002481 **
celltypeadeno	1.106691	3.024334	0.302340	3.660	0.000252 ***
celltypelarge	0.330336	1.391436	0.282727	1.168	0.242648
karno	-0.031383	0.969105	0.005182	-6.056	1.4e-09 ***
age2TRUE	-0.697151	0.498002	0.374494	-1.862	0.062662 .
age3TRUE	-0.430824	0.649973	0.345568	-1.247	0.212502
priort	0.096724	1.101556	0.207807	0.465	0.641608

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

	exp(coef)	exp(-coef)	lower .95	upper .95
trait	1.3056	0.7659	0.8837	1.929
celltypesmallcell	2.2525	0.4439	1.3311	3.812
celltypeadeno	3.0243	0.3307	1.6722	5.470
celltypelarge	1.3914	0.7187	0.7995	2.422
karno	0.9691	1.0319	0.9593	0.979
age2TRUE	0.4980	2.0080	0.2390	1.038
age3TRUE	0.6500	1.5385	0.3302	1.280
priort	1.1016	0.9078	0.7330	1.655

Rsquare= 0.377 (max possible= 0.999)

Likelihood ratio test= 64.79 on 8 df, p=5.309e-11

Wald test = 66.29 on 8 df, p=2.682e-11

Score (logrank) test = 69.87 on 8 df, p=5.215e-12

```
> stepAIC(coxcomplet,scope = list(upper = ~trait+celltype+karno+age2+age3+priort), direction="backwards")
```

Start: AIC=962.11

```
base ~ trait + celltype + karno + age2 + age3 + priort
```

	Df	AIC
- priort	1	960.32
- age3	1	961.52
- trait	1	961.90
<none>		962.11
- age2	1	963.20
- celltype	3	972.92
- karno	1	995.43

Step: AIC=960.32

```
base ~ trait + celltype + karno + age2 + age3
```

	Df	AIC
- age3	1	959.78
- trait	1	960.21
<none>		960.32
- age2	1	961.41
- celltype	3	970.93
- karno	1	993.47

Step: AIC=959.78

```
base ~ trait + celltype + karno + age2
```

	Df	AIC
- trait	1	959.36
<none>		959.78
- age2	1	959.83
- celltype	3	969.76
- karno	1	992.04

Step: AIC=959.36

```
base ~ celltype + karno + age2
```

	Df	AIC
<none>		959.36
- age2	1	959.53
- celltype	3	968.81
- karno	1	990.81

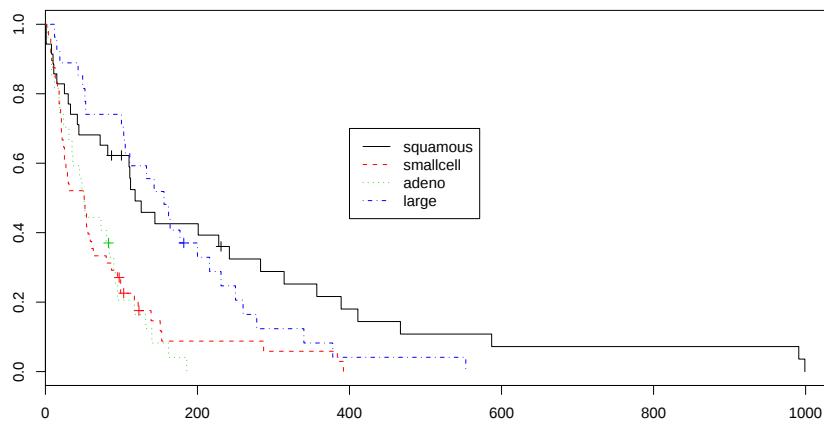
Call:

```
coxph(formula = base ~ celltype + karno + age2, data = veteran)
```

	coef	exp(coef)	se(coef)	z	p
celltypesmallcell	0.6798	1.974	0.25318	2.69	7.3e-03
celltypeadeno	1.0899	2.974	0.29528	3.69	2.2e-04
celltypelarge	0.2917	1.339	0.27790	1.05	2.9e-01
karno	-0.0304	0.970	0.00516	-5.90	3.6e-09
age2TRUE	-0.3044	0.738	0.21110	-1.44	1.5e-01

Likelihood ratio test=61.5 on 5 df, p=5.85e-12 n= 137

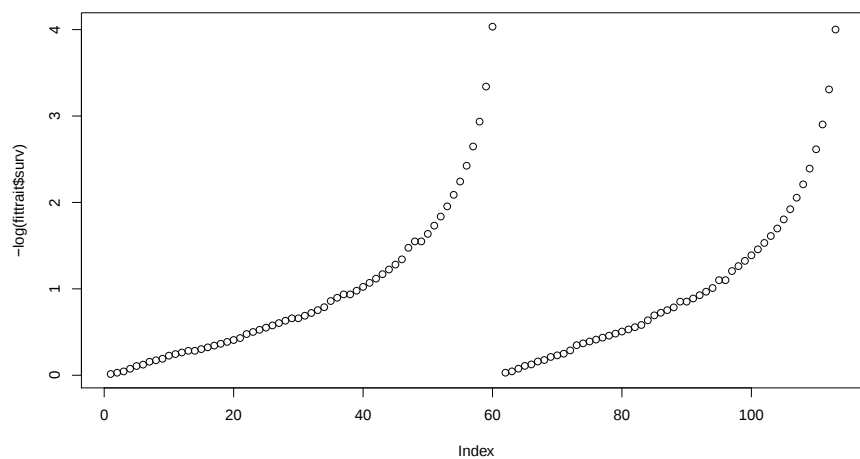
Courbes de survie par type de cellules



GRAPHES (P)

```
> plot(-log(fittrait$urv), lty = 2:3)
> title(main="Estimateur de Breslow du Risque cumulé par groupe de traitement")
```

Estimateur de Breslow du Risque cumulé par groupe de traitement



```
> plot(log(-log(fittrait$urv)))
> title(main="log(-log(S)) par groupe de traitement")
```

