

Modèles de Durée

Examen du 14 octobre 2014

Instructions: Durée 1h15 – Les documents sont interdits. La qualité de la présentation sera prise en compte dans la notation.

On s'intéresse à la base **melanoma**. On s'intéresse en particulier à la durée de suivi **days** jusqu'au décès par mélanome, et au statut **dc** (1: mort liée au mélanome, 2 : vivant à la fin de l'étude, 3 : mort d'une autre cause) pour $n = 205$ patients. Pour chaque patient nous disposons des informations suivantes :

- numéro du patient;
- la variable **dc**, 1 : décédé du mélanome, 2 : encore en vie, 3 : décédé d'une autre cause;
- la variable **days** : durée du suivi en jours;
- la variable **level** : niveau d'invasion, 0, 1 ou 2;
- la variable **ici** : niveau d'infiltration des cellules inflammatoires, 0, 1, 2 ou 3;
- la variable **ecel** : présence ou non de cellules épithéloïdes, 1 : absence, 2 : présence;
- la variable **ulc** : présence ou non d'une ulcération, 0 : absence, 1 : présence;
- la variable **thick** : épaisseur de la tumeur (en mm);
- la variable **sex**, 0 : femme, 1 : homme;
- la variable **age** : âge au moment de l'opération (en années).

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

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

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) Quel est le pourcentage de données non censurées? Commenter.
 - (b) Calculer l'estimateur de Kaplan-Meier de la fonction de survie S , pour les valeurs de t dans l'intervalle $t \in [0, 232]$ (tous groupes confondus)
 - (c) On s'intéresse dans un premier temps à l'influence du sexe sur la survie. On s'intéresse donc à l'estimation des fonctions de survie dans les deux groupes définis par le sexe. Commenter le programme et les résultats obtenus.
 - (d) On souhaite tester si le sexe a une influence sur la survie. Quel(s) test(s) proposez-vous ? Donner les hypothèses de(s) test(s), la(les) statistique(s) de test, sa(leur) loi sous l'hypothèse nulle, la zone de rejet et la p-value. Commenter le programme et les résultats obtenus.
 - (e) On s'intéresse maintenant à l'influence de la variable `ulc` sur la survie. Commenter le programme et les résultats obtenus.
 - (f) Quel test proposeriez-vous pour étudier la répartition des individus avec et sans ulcération dans les groupes de sexe? Préciser le test mis en oeuvre dans le programme. Que cherche-t-on à mettre en évidence. Commenter.
 - (g) On sait a priori que l'épaisseur de la tumeur (variable `thick`) est une variable très influente sur la survie. Pour compléter l'étude sur l'influence du sexe sur la survie par mélanome, on étudie la distribution de la variable `thick` dans les groupes de sexe. Commenter le programme. On précisera quels sont les tests utilisés et l'information qu'ils apportent. Commenter tous les résultats obtenus. Que peut-on dire de l'influence de la variable `sex` sur la durée de survie pour le mélanome?
2.
 - (a) On propose maintenant d'expliquer la variable *temps de survie* en fonction de la présence ou non d'une ulcération 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 pour lequel il y a *présence d'une ulcération* au groupe pour lequel il y a *absence d'une ulcération*. Donner son interprétation en fonction du (des) paramètre(s) du modèle de Cox.
 - (c) Rappeler l'expression de la pseudo-vraisemblance de Cox et proposer un estimateur du paramètre de régression du modèle de Cox.
 - (d) Donner l'estimation du rapport de risque instantané comparant le groupe pour lequel il y a *présence d'une ulcération* au groupe pour lequel il y a *absence d'une ulcération*. Commenter.
 - (e) Donner les hypothèses de test de l'influence de la variable `ulc` sur le temps de survie.

Programme et résultats:

```
> library(ISwR)
> library(survival)
> library(MASS)
> melanoma <- read.csv2("http://staff.pubhealth.ku.dk/~linearpredictors/datafiles/Melanoma
+ sep = ";", dec = ".", header=TRUE,
+ colClasses = c("factor", "numeric", "factor", "factor", "factor", "factor", "numeric", "factor")
> melanoma$ecel1[melanoma$ecel==1]=0
> melanoma$ecel1[melanoma$ecel==2]=1
> melanoma$ici1[melanoma$ici==0]=0
> melanoma$ici1[melanoma$ici==1]=1
> melanoma$ici1[melanoma$ici==2]=2
> melanoma$ici1[melanoma$ici==3]=3
> melanoma$age1=(melanoma$age<40)
> melanoma$age2=(melanoma$age>=40 & melanoma$age>60)
> melanoma$age3=(melanoma$age>=60)
> melanoma$thick1=(melanoma$thick<2)
> melanoma$thick2=(melanoma$thick>=2 & melanoma$thick<5)
> melanoma$thick3=(melanoma$thick>=5)
> attach(melanoma)
> head(melanoma)
```

	dc	days	level	ici	ecel	ulc	thick	sex	age	ecel1	ici1	age1	age2	age3	thick1
1	3	10	1	2	2	1	6.76	1	76	1	2	FALSE	TRUE	TRUE	FALSE
2	3	30	0	0	1	0	0.65	1	56	0	0	FALSE	FALSE	FALSE	TRUE
3	2	35	1	2	1	0	1.34	1	41	0	2	FALSE	FALSE	FALSE	TRUE
4	3	99	0	2	1	0	2.90	0	71	0	2	FALSE	TRUE	TRUE	FALSE
5	1	185	2	2	2	1	12.08	1	52	1	2	FALSE	FALSE	FALSE	FALSE
6	1	204	2	2	1	1	4.84	1	28	0	2	TRUE	FALSE	FALSE	FALSE

	thick2	thick3
1	FALSE	TRUE
2	FALSE	FALSE
3	FALSE	FALSE
4	TRUE	FALSE
5	FALSE	TRUE
6	TRUE	FALSE

```
> melanoma$days<-as.numeric(melanoma$days)
> base<-Surv(melanoma$days,melanoma$dc==1)
> base
```

[1]	10+	30+	35+	99+	185	204	210	232	232+	279	295	355+
[13]	386	426	469	493+	529	621	629	659	667	718	752	779
[25]	793	817	826+	833	858	869	872	967	977	982	1041	1055
[37]	1062	1075	1156	1228	1252	1271	1312	1427+	1435	1499+	1506	1508+
[49]	1510+	1512+	1516	1525+	1542+	1548	1557+	1560	1563+	1584	1605+	1621
[61]	1627+	1634+	1641+	1641+	1648+	1652+	1654+	1654+	1667	1678+	1685+	1690
[73]	1710+	1710+	1726	1745+	1762+	1779+	1787+	1787+	1793+	1804+	1812+	1836+
[85]	1839+	1839+	1854+	1856+	1860+	1864+	1899+	1914+	1919+	1920+	1927+	1933
[97]	1942+	1955+	1956+	1958+	1963+	1970+	2005+	2007+	2011+	2024+	2028+	2038+

```
[109] 2056+ 2059+ 2061 2062 2075+ 2085+ 2102+ 2103 2104+ 2108 2112+ 2150+
[121] 2156+ 2165+ 2209+ 2227+ 2227+ 2256 2264+ 2339+ 2361+ 2387+ 2388 2403+
[133] 2426+ 2426+ 2431+ 2460+ 2467 2492+ 2493+ 2521+ 2542+ 2559+ 2565 2570+
[145] 2660+ 2666+ 2676+ 2738+ 2782 2787+ 2984+ 3032+ 3040+ 3042 3067+ 3079+
[157] 3101+ 3144+ 3152+ 3154+ 3180+ 3182+ 3185+ 3199+ 3228+ 3229+ 3278+ 3297+
[169] 3328+ 3330+ 3338 3383+ 3384+ 3385+ 3388+ 3402+ 3441+ 3458+ 3459+ 3459+
[181] 3476+ 3523+ 3667+ 3695+ 3695+ 3776+ 3776+ 3830+ 3856+ 3872+ 3909+ 3968+
[193] 4001+ 4103+ 4119+ 4124+ 4207+ 4310+ 4390+ 4479+ 4492+ 4668+ 4688+ 4926+
[205] 5565+
```

```
> n=205
> ind<-(dc==1)
> sum(ind)/n
```

```
[1] 0.2780488
```

```
> fit<-survfit(base~1,data=melanoma)
> fit
```

```
Call: survfit(formula = base ~ 1, data = melanoma)
```

records	n.max	n.start	events	median	0.95LCL	0.95UCL
205	205	205	57	NA	NA	NA

```
> summary(fit)
```

```
Call: survfit(formula = base ~ 1, data = melanoma)
```

time	n.risk	n.event	survival	std.err	lower 95% CI	upper 95% CI
185	201	1	0.995	0.00496	0.985	1.000
204	200	1	0.990	0.00700	0.976	1.000
210	199	1	0.985	0.00855	0.968	1.000
232	198	1	0.980	0.00985	0.961	1.000
279	196	1	0.975	0.01100	0.954	0.997
295	195	1	0.970	0.01202	0.947	0.994
386	193	1	0.965	0.01297	0.940	0.991
426	192	1	0.960	0.01384	0.933	0.988
469	191	1	0.955	0.01465	0.927	0.984
529	189	1	0.950	0.01542	0.920	0.981
621	188	1	0.945	0.01615	0.914	0.977
629	187	1	0.940	0.01683	0.907	0.973
659	186	1	0.935	0.01748	0.901	0.970
667	185	1	0.930	0.01811	0.895	0.966
718	184	1	0.925	0.01870	0.889	0.962
752	183	1	0.920	0.01927	0.883	0.958
779	182	1	0.915	0.01981	0.877	0.954
793	181	1	0.910	0.02034	0.871	0.950
817	180	1	0.904	0.02084	0.865	0.946
833	178	1	0.899	0.02134	0.859	0.942
858	177	1	0.894	0.02181	0.853	0.938
869	176	1	0.889	0.02227	0.847	0.934

872	175	1	0.884	0.02272	0.841	0.930
967	174	1	0.879	0.02315	0.835	0.926
977	173	1	0.874	0.02357	0.829	0.921
982	172	1	0.869	0.02397	0.823	0.917
1041	171	1	0.864	0.02436	0.817	0.913
1055	170	1	0.859	0.02474	0.812	0.909
1062	169	1	0.854	0.02511	0.806	0.904
1075	168	1	0.849	0.02547	0.800	0.900
1156	167	1	0.844	0.02582	0.794	0.896
1228	166	1	0.838	0.02616	0.789	0.891
1252	165	1	0.833	0.02649	0.783	0.887
1271	164	1	0.828	0.02681	0.777	0.883
1312	163	1	0.823	0.02713	0.772	0.878
1435	161	1	0.818	0.02744	0.766	0.874
1506	159	1	0.813	0.02774	0.760	0.869
1516	155	1	0.808	0.02805	0.755	0.865
1548	152	1	0.802	0.02837	0.749	0.860
1560	150	1	0.797	0.02868	0.743	0.855
1584	148	1	0.792	0.02899	0.737	0.851
1621	146	1	0.786	0.02929	0.731	0.846
1667	137	1	0.780	0.02963	0.725	0.841
1690	134	1	0.775	0.02998	0.718	0.836
1726	131	1	0.769	0.03033	0.712	0.831
1933	110	1	0.762	0.03085	0.704	0.825
2061	95	1	0.754	0.03155	0.694	0.818
2062	94	1	0.746	0.03221	0.685	0.812
2103	90	1	0.737	0.03290	0.676	0.805
2108	88	1	0.729	0.03358	0.666	0.798
2256	80	1	0.720	0.03438	0.656	0.791
2388	75	1	0.710	0.03523	0.645	0.783
2467	69	1	0.700	0.03619	0.633	0.775
2565	63	1	0.689	0.03729	0.620	0.766
2782	57	1	0.677	0.03854	0.605	0.757
3042	52	1	0.664	0.03994	0.590	0.747
3338	35	1	0.645	0.04307	0.566	0.735

```

> fitsex<-survfit(base~sex,data=melanoma)
> plot(fitsex,lty=c(1,2),col=c(1,2))
> legend(3000,0.4,c("F","H"),lty=c(1,2),col=c(1,2))
> title(main="Kaplan-Meier par sexe")
> survdiff(base~sex,data=melanoma)

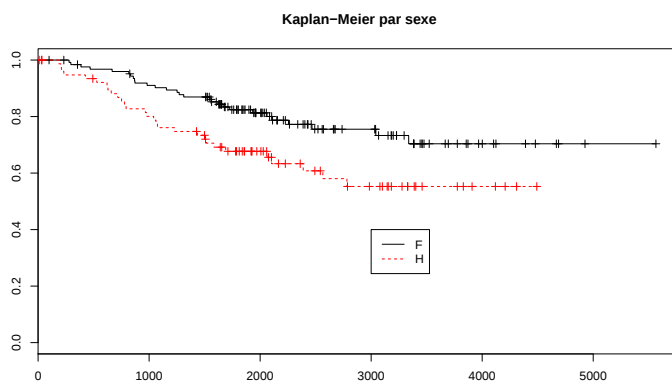
```

Call:

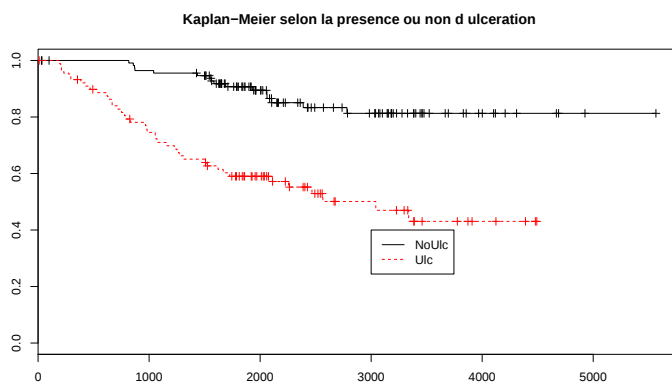
```
survdiff(formula = base ~ sex, data = melanoma)
```

	N	Observed	Expected	(O-E) ² /E	(O-E) ² /V
sex=0	126	28	37.1	2.25	6.47
sex=1	79	29	19.9	4.21	6.47

Chisq= 6.5 on 1 degrees of freedom, p= 0.011



```
> fitulc<-survfit(base~ulc,data=melanoma)
> plot(fitulc,lty=c(1,2),col=c(1,2))
> legend(3000,0.4,c("NoUlc","Ulc"),lty=c(1,2),col=c(1,2))
> title(main="Kaplan-Meier selon la presence ou non d ulceration")
```



```
> survdiff(base~ulc,data=melanoma)
```

Call:

```
survdiff(formula = base ~ ulc, data = melanoma)
```

	N	Observed	Expected	(O-E) ² /E	(O-E) ² /V
ulc=0	115	16	35.8	10.9	29.6
ulc=1	90	41	21.2	18.5	29.6

Chisq= 29.6 on 1 degrees of freedom, p= 5.41e-08

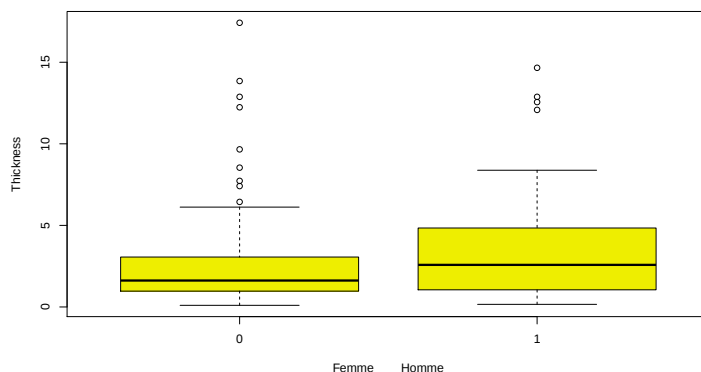
```
> chisq.test(melanoma$sex,melanoma$ulc,correct=FALSE)
```

Pearson's Chi-squared test

data: melanoma\$sex and melanoma\$ulc

X-squared = 5.7845, df = 1, p-value = 0.01617

```
> boxplot(thick~sex,col="yellow2",xlab='Femme Homme', ylab='Thickness',data=melanoma)
```



```
> wilcox.test(melanoma$thick[melanoma$sex==0],melanoma$thick[melanoma$sex==1])
```

Wilcoxon rank sum test with continuity correction

```
data: melanoma$thick[melanoma$sex == 0] and melanoma$thick[melanoma$sex == 1]
W = 3794.5, p-value = 0.004213
alternative hypothesis: true location shift is not equal to 0
```

```
> t.test(melanoma$thick[melanoma$sex==1],melanoma$thick[melanoma$sex==0])
```

Welch Two Sample t-test

```
data: melanoma$thick[melanoma$sex == 1] and melanoma$thick[melanoma$sex == 0]
t = 2.6059, df = 149.094, p-value = 0.01009
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.2718653 1.9775560
sample estimates:
mean of x mean of y
 3.611139  2.486429
```

```
> coxulc<-coxph(base~ulc,data=melanoma,method="exact")
> summary(coxulc)
```

Call:

```
coxph(formula = base ~ ulc, data = melanoma, method = "exact")
```

n= 205, number of events= 57

```
      coef exp(coef) se(coef)      z Pr(>|z|)
ulc1 1.4717   4.3567   0.2954  4.982 6.29e-07 ***
---
Signif. codes:  0
```