

MATH 60604

Modélisation statistique

§ 7e - Test du log rang

Léo Belzile

HEC Montréal
Département de sciences de la décision

Comparaison de courbes de survie

Les données cancersein contiennent les résultats d'une étude sur la survie de femmes atteintes du cancer du sein.

- temps: temps avant la mort, ou la fin de l'étude, en mois.
- mort: variable indicatrice pour la mort, 0 pour les survivantes et 1 pour les décédées
- repimmuno: réponse immunohistochimique, soit négative (0) ou positive (1)

On s'intéresse à la question suivante:

- Est-ce que les femmes qui répondent positivement à l'examen immunohistochimique ont tendance à survivre moins longtemps que celles qui répondent négativement?

Comparaison de courbes de survie

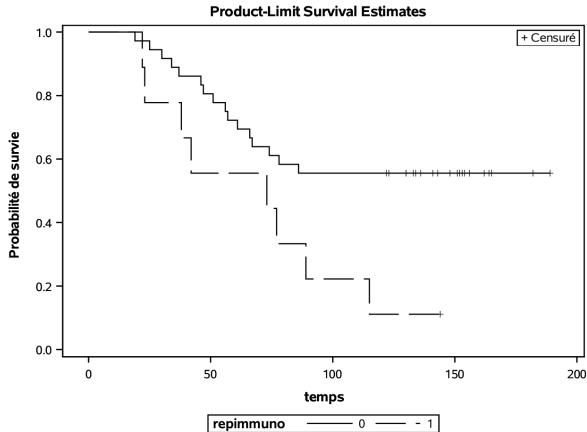
On peut ajuster des courbes de survie différentes par groupe avec l'option `strata`.

Code SAS pour le modèle de Kaplan–Meier

```
proc lifetest data=modstat.cancersein method=km;  
time temps*mort(0);  
strata repimmuno;  
run;
```

SAS va estimer la courbe de survie pour les individus avec une réaction négative (groupe `repimmuno=0`) séparément de ceux qui ont une réaction positive (groupe `repimmuno=1`).

Courbes de survie (Kaplan–Meier)



Comparaison de courbes de survie

Il semble que les femmes ayant une réaction négative à l'examen (repimmuno=0) ont un meilleur taux de survie que celles qui ont une réaction positive (repimmuno=1).

- Pour la majorité des temps t , $\hat{S}_1(t) > \hat{S}_2(t)$ et donc ceux avec repimmuno=0 ont une probabilité de survie supérieure à ceux avec repimmuno=1

Est-ce que la fonction de survie est significativement différente dans les deux groupes repimmuno=0 et repimmuno=1?

$$\mathcal{H}_0 : S_0(t) = S_1(t) \text{ pour tout } t,$$

$$\mathcal{H}_1 : S_0(t) \neq S_1(t) \text{ pour au moins une valeur de } t.$$

Test du log rang

Considérons un modèle à risques proportionnels de Cox avec fonction de risque

$$h(t) = h_0(t) \exp(\beta \text{repimmuno}). \quad (\star)$$

- L'hypothèse nulle pour l'égalité des fonctions de survie est équivalente à $\mathcal{H}_0 : \beta = 0$.
- La statistique du score permet de tester cette hypothèse sans ajuster le modèle.
 - On recouvre l'estimateur de Kaplan–Meier de la fonction de survie si $\beta = 0$.
- Il suffit de calculer le gradient et la hessienne du modèle décrit par (\star) et l'évaluer en $\beta = 0$.
 - Ce sont des fonctions simples du nombre de personnes à risque dans chaque groupe aux temps t_i .

Ajustement du modèle à risques proportionnels

Code SAS pour le modèle de risques proportionnels

```
proc phreg data=modstat.cancersein;  
model temps*mort(0) = repimmuno;  
run;
```

Test d'égalité sur les niveaux de discrétisation				Test de l'hypothèse nulle globale : BETA=0			
Test	khi-2	DDL	Pr > khi-2	Test	khi-2	DDL	Pr > khi-2
Log-rang	5.4943	1	0.0191	Rapport de vrais	4.4463	1	0.0350
Wilcoxon	4.3512	1	0.0370	Score	5.4943	1	0.0191
-2Log(LR)	5.6708	1	0.0172	Wald	5.0804	1	0.0242

Le test du log rang est aussi présenté par défaut dans la sortie SAS de la procédure `lifetest` (gauche).

Test du log rang

- Sous $\mathcal{H}_0 : \beta = 0$, la loi nulle de la statistique de score est approximativement χ_1^2 .
- La valeur- p est 0.0191: on rejette \mathcal{H}_0 à niveau 5% et on conclut que les fonctions de survie sont significativement différentes pour les femmes avec des réactions négatives / positives à l'examen immunohistochimique.
- On peut généraliser le test du log rang en utilisant un modèle de Cox qui n'inclut qu'une variable catégorielle à k niveaux
 - la loi nulle de la statistique du test de score sera χ_{k-1}^2 .