

## Devoir no. 2- MAT 898A-2015

### Partie I: une dernière vérification de formules...

**Calcul avec la vraisemblance partielle de Cox et l'estimateur Kaplan-Meier:** révision de nos formules sur un mini exemple.

Soit les données (extraites de la banque Stanford data):

Mismatch, $x$	2.09	0.36	0.6	1.44	0.92
Survie, $y$	55	120+	280	389+	1516+

Ces données portent sur des greffes d'organes, et par le score de *mismatch* on indique la ressemblance entre le donneur et le patient qui reçoit le transplant.

(a) Calculer l'estimateur de Kaplan-Meier pour ces quelques données. Comparer cet estimateur avec une estimation qui ne tient pas compte de la censure,  $1 - F_n(y)$ , où  $F_n(y)$  est la fonction de répartition empirique.

(b) Écrire le modèle de Cox qui incorpore une variable explicative, le score de *mismatch*.

(c) Écrire la vraisemblance partielle de Cox,  $L_c(\beta)$ .

(d) Tester l'hypothèse  $H_0 : \beta = 1.3$  (why not !?) en utilisant le test de Rao (score), basé sur la statistique  $(u_c(\beta_0))^2 / I(\beta_0)$  où

$$u_c(\beta) = \frac{\partial}{\partial \beta} \log L_c(\beta), \quad I(\beta) = -\frac{\partial^2}{\partial \beta^2} \log L_c(\beta).$$

On suppose que sous  $H_0$  la statistique suit une loi du  $\chi^2$  avec 1 degré de liberté (c'est une approximation assez grossière avec si peu de données). Effectuer le test "à la main" et comparer avec la sortie R (qui contient un résultat à partir duquel on peut faire ce test).

(e) Estimer la fonction de survie avec un modèle de Cox (à  $x = 0.92$ , valeur médiane) et comparer avec le Kaplan-Meier obtenu en (a).

### Partie II: petite analyse de données

Considérer la banque de données sur la cirrhose et prendre seulement quatre variables explicatives, Z1 (IV, code du traitement), Z8 (XI, *bilirubine*), Z16 (XIX, temps de *prothrombine*) et Z17 (XX, stade de la maladie), et ajuster un modèle de Cox avec ces 4 variables "explicatives". (L'idée de départ a été de comparer la survie entre les traités-non traités et prendre plein de variables "confondantes".) Selon la description des données, seulement les patients correspondant aux premières 312 lignes ont participé à l'étude clinique et donc on utilise seulement ceux-là. (Attention aux codes des données manquantes.) La raison de considérer la variable Z16 est qu'elle correspond à un test sanguin de l'état de santé du foie; Z8 est un bon indicateur de la condition de la vésicule biliaire, donc cette variable est reliée au foie et elle-aussi peut influencer sérieusement la survie.

(i) Interpréter les résultats des tests (si on garde ou non les variables dans le modèle). Nous savons de fait que le traitement ne fait pas de différence pour la survie. Par contre, on peut se poser des questions sur l'apport des autres variables explicatives. Entre autres, il est à noter que la variable Z17 est une autre variable qualitative, avec laquelle on peut former des groupes et reprendre le même type d'analyse que celle vue au cours. (Noter par contre qu'il y a peu de données, 16, au stade 1, donc regrouper 1 et 2 peut-être?)

(ii) Une fois le modèle de Cox ajusté, considérer les deux divisions par groupe : (a) traité ou non et (b) stade de la maladie, et représenter graphiquement la survie selon les conclusions en (i) (il y a deux graphiques avec plusieurs courbes chacun).

**Note :** Par défaut, la fonction "survfit" appliquée à un objet de type "coxph" calcule la survie à une valeur "moyenne" des variables explicatives. Tel que montré dans les sorties, pour avoir une courbe par groupe, selon une variable nommée gr par exemple, on va mettre dans l'équation de "coxph" `strata(gr)` plutôt que seulement gr. Après, il reste à faire `plot(survfit(...))`.

(iii) On pourrait se poser des questions s'il y a une proportionnalité de Cox quand on sépare selon les stades de maladie (mais en tenant compte des autres variables dans l'équation). On peut répondre à cette question de plusieurs manières différentes (il y a aussi une fonction dans R qui fait un test pour chaque variable, `cox.zph()`). Avec la théorie vue au cours on peut (choisir une des méthodes) :

(a) comparer les fonctions de survie appropriées ;

(b) faire le graphique (QQ-plot) qui permet de voir si le modèle de Cox (avec des groupes) est adéquat ou non.

(iv) Finalement, le comportement des résidus (Cox-Snell modifiés) donne une bonne idée sur l'ajustement avec toutes les variables explicatives considérées.

### **Partie III: lecture de tableaux dans un article**

Dans l'article ci-joint, *Movement disorders (2011) vol. 26, p. 2522-2529*, on veut comparer la mortalité de patients atteints de PD (Parkinson) et non atteints (catégorie de référence). On calcule des "hazard ratio" ajustés et non ajustés et on tire des conclusions.

On vous demande de présenter sur maximum (!!) deux pages :

(i) une courte explication des procédures utilisées dans la section *Data Analyses* ;

(ii) les équations qui correspondent au tableau 3 (vous pouvez déduire le type des variables à partir du tableau 2, par exemple) ;

(iii) une explication (de bon sens) du fait que la valeur du "hazard ratio" descend quand on passe du modèle non ajusté au modèle 1 (ajusté).

**À remettre au plus tard le 22 décembre 2015, 10h.**