

TP d'analyse de données de survie

3 heures

Valérie Garès

5 novembre 2013

Introduction

L'objectif de ce TP est de mettre en pratique sous R, STATA ou SAS les méthodes d'analyse de données de survie, vues en cours, à savoir :

- estimation paramétrique ou non-paramétrique de la fonction de survie,
- représentation des courbes de survie,
- le test du logrank,
- le modèle de Cox et ses hypothèses.

1 Quelques rappels de cours

1.1 Définitions

- **T** : délai d'apparition de l'évènement d'intérêt,
- **S(t)** : probabilité pour qu'un patient soit encore vivant après un délai t,
- **F(t)** : fonction de répartition. On a $F(t) = 1 - S(t)$,
- **f(t)** : densité de probabilité. On a $f(t) = F'(t)$,
- $\lambda(t)$: fonction de risque. On a $\lambda(t) = \frac{f(t)}{S(t)}$,
- $\Lambda(t)$: fonction de risque cumulé. On a $\Lambda(t) = -\ln(S(t))$,
- **Censure à droite** : l'évènement T a eu lieu après une certaine date C ($T > C$),
- **Date d'origine** : date à laquelle le sujet peut être considéré comme entrant dans l'étude et à partir de laquelle on comptera les délais le concernant,
- **Date des dernières nouvelles** : date la plus tardive pour laquelle on est assuré que le sujet n'a pas eu l'évènement (DDN),
- **Perdu de vue** : si le patient est non décédé aux dernières nouvelles,
- **Date de surveillance** : délai écoulé entre la date d'origine et la date des dernières nouvelles,
- **Date de point** : date à laquelle l'étude se termine, *commune à tous les individus*, (fin du recueil des informations) (DP),
- **Exclu-vivant** : si $DDN > DP$,
- **Temps de participation** : temps qui sera reporté dans l'analyse comme "durée de suivi",
 - si $T \leq DP$, temps de participation=T (non censuré),

- sinon le sujet est censuré :
 - si $DDN \geq DP$ ou $T \geq DP$, temps de participation=DP,
 - si $DDN < DP$ temps de participation=DDN.

1.2 Estimation des fonctions de survie

1.2.1 Estimation non paramétrique de Kaplan-Meier

- $t_1 < t_2 < \dots < t_k$, les k temps d'apparition de l'évènement,
- N_i : nombre de sujets connus "sans l'évènement" après t_{i-1} ,
- C_i : nombre de patients censurés entre t_{i-1} et t_i ,
- D_i : nombre de sujets ayant eu l'évènement au temps t_i ,
- $N_i - C_i$: nombre de sujets à "risque".

$$\hat{S}_{KM}(t_j) = \prod_{0 \leq t_i \leq t_j} 1 - \frac{D_i}{N_i - C_i}.$$

Les bornes de l'intervalle de confiance de Greenwood de la probabilité de survie au temps t_i sont :

$$\hat{S}_{KM}(t_j) \pm 1.96 \times \sqrt{\sum_{t_i \leq t_j} \frac{D_i}{(N_i - C_i)(N_i - C_i - D_i)}}.$$

1.2.2 Méthode actuarielle

- Soient les intervalles : $[a_0, a_1[$, $[a_1, a_2[$, ...
- N_i : nombre de sujets connus "sans l'évènement" en a_i ,
- C_i : nombre de patients censurés dans l'intervalle $[a_{i-1}, a_i[$,
- D_{i-1} : nombre de sujets ayant eu l'évènement dans l'intervalle $[a_{i-1}, a_i[$.

$$\hat{S}_{AC}(a_j) = \prod_{a_i \leq a_j} 1 - \frac{D_{i-1}}{N_{i-1} - \frac{C_{i-1}}{2}}.$$

Elle est semblable à la méthode Kaplan-Meier mais les intervalles de temps ne sont plus déterminés par la survenue des évènements. La taille des intervalles de temps est fixée a priori : 1 semaine, 1 mois, 1 an... On calcule la probabilité de survie dans chaque intervalle (moins exacte que Kaplan-Meier). Le nombre de sujets à risque dans l'intervalle est le nombre de sujets à risque en début d'intervalle moins la moitié des perdus de vue dans l'intervalle. Puis les calculs sont identiques.

1.2.3 Estimation dans un modèle paramétrique

- Modèle exponentiel : $S(t) = \exp(-\lambda \times t)$,
- Modèle de Weibull : $S(t) = \exp[-(\lambda \times t)^\gamma]$,
- Modèle log normal, log logistique ...

1.3 Estimation de la fonction de risque cumulé de Nelson-Aalen

$$\hat{\Lambda}_{NA}(t) = \sum_{t_i \leq t} \frac{D_i}{N_i - C_i}.$$

1.4 Comparaison de groupe : test du logrank (non paramétrique)

On veut tester l'égalité des fonctions de survie de deux groupes (A et B). On ajoute l'indice A ou B pour distinguer les groupes.

- D_A et D_B : nombre total d'évènements observés dans chaque groupe,
- E_A et E_B : nombre total d'évènements attendus dans chaque groupe,

$$E_A = \sum_{t_i \leq t_k} D_i \frac{N_i^A - C_i^A}{N_i - C_i} \quad \text{et} \quad E_B = \sum_{t_i \leq t_k} D_i \frac{N_i^B - C_i^B}{N_i - C_i}.$$

$$LR = \frac{(D_A - E_A)^2}{E_A} + \frac{(D_B - E_B)^2}{E_B}.$$

suit un χ^2 à 1 ddl.

Si on veut comparer L courbes de survie, on calcule de même les effectifs totaux attendus E_l et observés D_l dans chaque groupe l ($1 \leq l \leq L$). Alors

$$\sum_{l=1}^L \frac{(D_l - E_l)^2}{E_l}$$

suit un χ^2 à $L - 1$ ddl.

1.5 Approche semi-paramétrique : Le modèle de Cox

1.5.1 Définition

Le modèle de Cox permet d'établir une relation paramétrique entre la fonction de risque et les facteurs de risque (covariables)

$$\begin{aligned} \lambda(t, (Z_1, Z_2, \dots, Z_p)) &= \lambda_0(t) e^{\beta_1 Z_1 + \beta_2 Z_2 + \dots + \beta_p Z_p}, \\ &= \lambda_0(t) e^{\beta^T Z} \end{aligned}$$

où λ_0 est la fonction de risque de base, et $Z = (Z_1, Z_2, \dots, Z_p)$ sont les covariables. Le but est d'estimer les β_j .

Ce modèle est à risque proportionnel :

- **Si Z est une variable à deux classes.** Prenons l'exemple où Z est la covariable prenant 0 ou 1.
Exemple : $Z = 1$ pour le groupe A et $Z = 0$ pour le groupe B.

$$RR = \frac{\lambda(t, Z = 1)}{\lambda(t, Z = 0)} = e^\beta.$$

Si on prend comme référence les individus ayant la covariable $Z = 0$, e^β est le **risque relatif** (RR) du groupe A par rapport au groupe B. Il est **constant**. Autrement dit, les individus ayant la covariable $Z = 1$ ont un risque instantané d'avoir l'évènement e^β fois celui des individus ayant $Z = 0$.

- Si $\beta > 0$ alors $RR > 1$, risque plus élevé d'avoir l'évènement pour le groupe A,
- Si $\beta = 0$ alors $RR = 1$, risques égaux pour les deux groupes,
- Si $\beta < 0$ alors $RR < 1$, risque plus faible pour le groupe A.

- **Si Z est une variable qualitative à n classes** ($n \geq 3$), $Z = 0, 1, 2, \dots, n - 1$. Le modèle de Cox considère $n - 1$ variables à 2 classes :

$$- Z_{1,1} = \begin{cases} 1 & \text{si } Z = 1 \\ 0 & \text{si } Z \neq 1, \end{cases}$$

$$- Z_{1,2} = \begin{cases} 1 & \text{si } Z = 2 \\ 0 & \text{si } Z \neq 2, \end{cases}$$

- ...,

$$- Z_{1,n-1} = \begin{cases} 1 & \text{si } Z = n - 1 \\ 0 & \text{si } Z \neq n - 1. \end{cases}$$

Alors

$$\lambda(t, Z) = \lambda_0(t) e^{\beta_1 Z_{1,1} + \beta_2 Z_{1,2} + \dots + \beta_n Z_{1,n-1}}.$$

$RR_i = e^{\beta_i}$ mesure le rapport des risques d'un sujet de classe i par rapport à un individu de classe de référence i_r ($i_r \neq i$). Ici $i_r = 0$.

- **Si Z est une variable continue,**

$$RR = \frac{\lambda(t, Z + 1)}{\lambda(t, Z)} = \frac{\lambda_0(t) e^{\beta(Z+1)}}{\lambda_0(t) e^{\beta Z}} = e^{\beta}.$$

$RR = e^{\beta}$ mesure le rapport des risques de deux sujets dont les valeurs diffèrent d'une unité.

1.5.2 Estimation des paramètres et tests de significativité

Estimation des paramètres. Soient D durées à l'échec distinctes, triées dans un ordre tel que $t_{(1)} < \dots < t_{(D)}$ et soient R_i les individus toujours soumis à risque juste avant le temps t_i . La probabilité conditionnelle que cet i -ième individu meure en t_i sachant que l'on avait R_i sujets à risque en t_i s'exprime comme suit :

$$\frac{e^{\beta^T Z^i}}{\sum_{j \in R_i} e^{\beta^T Z^j}}$$

La vraisemblance que Cox a proposée repose sur le produit de ces quantités.

$$L(\beta) = \prod_{i=1}^D \frac{e^{\beta^T Z^i}}{\sum_{j \in R_i} e^{\beta^T Z^j}}$$

L'estimation du maximum de vraisemblance $\hat{\beta}$ du vecteur β vérifie :

$$\frac{\delta \ln(L(\beta))}{\delta \beta} \Big|_{\beta=\hat{\beta}} = 0$$

On note $I(\beta)$, la matrice des dérivées secondes par rapport à β .

Tests de significativité des paramètres Le maximum de vraisemblance $\hat{\beta}$ est asymptotiquement normal de moyenne β et de variance $I(\beta)^{-1}$. L'intervalle de confiance de β est donné par

$$[\hat{\beta} - z_{1-\frac{\alpha}{2}} I(\hat{\beta})^{-1}, \hat{\beta} + z_{1-\frac{\alpha}{2}} I(\hat{\beta})^{-1}],$$

$z_{1-\frac{\alpha}{2}}$ est le quantile de la loi normale d'ordre $1 - \frac{\alpha}{2}$.

Rappelons

$$RR(t, Z = a, Z = b) = \frac{\lambda(t, Z = a)}{\lambda(t, Z = b)} = e^{(a-b)\beta}$$

- Si Z est une variable à classe, l'estimateur du rapport de risque est donné par

$$\widehat{RR}(t, a, b) = e^{(a-b)\widehat{\beta}}.$$

L'intervalle de confiance est

$$[\exp\left((a-b)\widehat{\beta} - |a-b|z_{1-\frac{\alpha}{2}}I(\widehat{\beta})^{-1}\right), \exp\left((a-b)\widehat{\beta} + |a-b|z_{1-\frac{\alpha}{2}}I(\widehat{\beta})^{-1}\right)].$$

- Si Z est une variable continue, l'estimateur du rapport de risque est donné par

$$\widehat{RR}(t, a+c, a) = e^{c\widehat{\beta}}.$$

$RR = e^{c\widehat{\beta}}$ mesure le rapport des risques de deux sujets dont les valeurs diffèrent de c unités. L'intervalle de confiance est

$$[\exp[c\widehat{\beta} - cz_{1-\frac{\alpha}{2}}I(\widehat{\beta})^{-1}], \exp[c\widehat{\beta} + cz_{1-\frac{\alpha}{2}}I(\widehat{\beta})^{-1}].$$

Pour tester l'hypothèse nulle $\mathcal{H}_0 : \beta = 0$, soit le vecteur des effets $(\beta_1, \dots, \beta_p)$ est nul, les trois statistiques de test classiquement utilisées sont :

1. Test de Wald

$$(\widehat{\beta})^T I(\widehat{\beta}) \widehat{\beta}$$

2. Test de vraisemblance

$$2(\ln L(\widehat{\beta}) - \ln L(0))$$

3. Test du score

$$U^T(0)I(0)^{-1}U(0)$$

Ces trois statistiques suivent approximativement sous l'hypothèse nulle, des distributions de Chi-deux à p degrés de liberté (p est le nombre de variables explicatives prises en compte dans le modèle). Si les valeurs de ces statistiques sont supérieures à la valeur limite du χ^2 à p degrés de liberté au risque $\alpha = 5\%$, alors on rejette l'hypothèse nulle et on conclut à l'effet significatif des covariables sur la survie.

1.5.3 Diagnostic du modèle de Cox

Il y a deux hypothèses à vérifier :

1. la **log-linéarité** pour les variables continues,
2. les **risques proportionnels**.

Les résidus. De nombreuses procédures de vérification des hypothèses du modèle de Cox sont basées sur des quantités que l'on appelle résidus. Il existe plusieurs résidus :

1. Les résidus de **Schoenfeld** servent à tester l'**hypothèse des risques proportionnels**.
 - Si l'hypothèse des risques proportionnels est vérifiée, ces résidus ont en théorie un aspect totalement aléatoire et l'évolution temporelle moyenne est une droite horizontale.

- Si l’hypothèse des risques proportionnels n’est pas vérifiée, par exemple si le facteur de risque est important au début du suivi du patient mais pas à la fin, alors les résidus seront, sur le schéma, négatifs puis positifs, et l’évolution temporelle moyenne sera une courbe croissante (ou le contraire selon le codage du facteur risque).

2. Les résidus des **Martingales** permettent de détecter la **non linéarité**, c’est-à-dire une forme fonctionnelle mal spécifiée dans la partie paramétrique du modèle.

Le tracé des résidus des martingales en fonction des variables explicatives incluses dans le modèle peut être utilisé pour indiquer si certaines variables ont besoin d’être transformées avant d’être incorporées dans le modèle. Pour cela, on ajoute une courbe lisse sur les points obtenus. La forme fonctionnelle est alors suggérée par la forme de la courbe lisse :

- une croissance lente suggèrera une transformation logarithmique ou racine.
- A l’inverse, une croissance rapide suggèrera une transformation puissance avec une puissance supérieure à 1.

Par ailleurs, le tracé des résidus des martingales en fonction des variables explicatives non incluses dans le modèle peut être utilisé pour indiquer si certaines variables devraient être incluses dans le modèle, ce qui est le cas si une dépendance apparaît.

3. Les résidus de **Cox-Snell** servent à valider l’ensemble du modèle. L’idée est que si T suit une loi de fonction de risque cumulé $\Lambda(t)$, alors $Y = \Lambda(T)$ suit une loi exponentielle de paramètre 1. On procède de la façon suivante : on estime Λ du modèle supposé par $\hat{\Lambda}$. On calcule les résidus de Cox-Snell $Y_i = \hat{\Lambda}(X_i)$. On estime la fonction de risque cumulé des Y_i de façon non paramétrique (Nelson-Aalen) par $\hat{\Lambda}_{NA}$. Si le modèle est correct, $\hat{\Lambda}_{NA}$ est à peu près égal au Λ d’une exponentielle de paramètre 1 (=1a bissectrice).

4. Les résidus de **déviance** (resymétrisation des résidus des martingales, pour corriger leur asymétrie) sont compris entre -1 et 1. Ils valent 0 sous l’hypothèse des risques proportionnels.

5. Les résidus du **score** permettent d’identifier les observations qui contribuent fortement à la détermination des paramètres du modèle.

6. Il y a aussi les indices **dfbetas** qui permettent de rechercher les sujets affluents (ou marginaux). Cette recherche se limite à la représentation des indices dfbetas de chaque sujet en fonction du temps. Des points extrêmes ou isolés doivent faire examiner en détail les sujets correspondants à la recherche d’une erreur de cotisation ou d’une configuration exceptionnelle des variables explicatives. Il pourra parfois être intéressant d’extraire provisoirement ces sujets de l’échantillon et de recalculer le modèle : une variabilité importante fera craindre une instabilité importante et jettera un doute sur les résultats.

Hypothèse des risques proportionnels. On peut tester l’hypothèse de proportionnalité des risques de la façon suivante :

1. Utiliser l’interaction avec le temps,
2. Méthodes graphiques pour les variables qualitatives. Le tracé de $\ln(-\ln(S(t)))$ en fonction de $\ln t$ dans chaque groupe, dans le cas de risques proportionnels, devrait se traduire par des courbes translatées.
3. Test basé sur les résidus de Schoenfeld (cf 1).

Dans le cas de risques non proportionnels, on peut utiliser une des méthodes suivantes :

1. Transformer les variables quantitatives en variables qualitatives,

2. Pour les variables qualitatives, essayer de changer de classes,
3. Si hypothèse de proportionalité vérifiée sur des intervalles de temps court, faire une modélisation par partie (faire différents modèles sur ces intervalles),
4. Stratification.

Si on connaît la valeur de la variable à différents instants, il existe le modèle de Cox avec variables dépendantes du temps pour des variables longitudinales. La vraisemblance est alors la suivante :

$$L(\beta, t) = \prod_{i=1}^D \frac{e^{\beta^T Z^i(t)}}{\sum_{j \in R_i} e^{\beta^T Z^j(t)}}$$

Modélisation par partie ou Stratification : Lorsque l'hypothèse des risques proportionnels n'est pas vérifiée, s'il s'agit de l'effet d'un traitement sur la survenue d'une maladie, on utilise la modélisation par partie car on a besoin d'un "RR" et d'un "p" pour la variable traitement puisque c'est la variable d'intérêt. S'il s'agit par exemple de l'effet "centre", on utilisera la stratification car c'est simple à mettre en oeuvre et la plupart du temps, on veut seulement ajuster sur le centre et non expliquer l'effet centre.

Hypothèse de Log-linéarité. Si l'on peut raisonnablement accepter l'hypothèse de log-linéarité, la variable doit rester continue. Cette hypothèse a pour conséquences qu'une augmentation de la variable par x multiplie le risque de survie par $\exp(x\beta)$ que l'on passe de x à $x + a$ ou de $2x$ à $2x + a$.

Par exemple si l'évènement est le décès et la covariable est l'âge, alors cette hypothèse entraîne qu'une augmentation de l'âge de 5 ans multiplie le risque de décéder par $\exp(5\beta)$ que l'on passe de 20 à 25 ans ou de 70 à 75 ans.

C'est une hypothèse forte qui n'est généralement pas recommandée. Si l'on choisit de transformer une variable quantitative continue en une variable à n classes ordonnées, se pose le problème du nombre de classes et du choix des seuils pour découper cette variable. La transformation peut être simple en classes en utilisant la moyenne ou la médiane de la variable continue ou plus complexe en utilisant les tertiles pour un découpage en 3 classes ou les quartiles pour un découpage en 4 classes. L'information la plus réduite est associée au plus petit nombre de classes. Attention à ce que l'hypothèse de proportionnalité des risques soit vérifiée. On peut également essayer de transformer la variable (logarithme, racine, puissance...) afin qu'elle vérifie l'hypothèse de log-linéarité (cf 3).

Voici maintenant une présentation des logiciels R, STATA et SAS en analyse de données de survie. On utilisera les variables :

- *duree* : durée de suivi du patient,
- *evt* : variable qui indique si l'évènement d'"échec" a eu lieu (survenue de la maladie ou le décès). C'est une indicatrice qui vaut 1 resp 0 (oui resp non). A noter que la variable décrivant la survenue de l'évènement est aussi la variable décrivant la censure : si l'évènement ne s'est pas réalisé pendant le suivi, alors l'information est censurée.
- *groupe* : le groupe auquel appartient le sujet,
- *cov*, *cov1*, *cov2* : des covariables,
- *id* : identifiant du sujet.

2 Sous STATA

Il faut avant tout ouvrir la base de données que l'on souhaite analyser.

2.1 Déclaration des données de survie

Les données de survie doivent toujours être déclarées comme telles :

- **stset** : permet de déclarer que les données sont des données de survie,
- **streset** : permet de changer la déclaration des données de survie,
- **st** : permet de voir de quelle manière les données sont actuellement déclarées.

Lorsque on tape **stset** ou **streset**, STATA fait toute une série de vérifications sur les données pour s'assurer que la manière dont elles ont été déclarées est correcte et a un sens.

Les syntaxes de ces trois commandes sont les suivantes :

```
stset duree, failure(evt[==numlist]) [id(varid) enter(tpsdbt) exit(tpsfin) origin(tpsrisk)]
streset duree, failure(evt[==numlist]) [id(varid) enter(tpsdbt) exit(tpsfin) origin(tpsrisk)]
st
```

- **failure(evt)** :

Si on ne spécifie rien après *evt*, STATA considère qu'un évènement a eu lieu lorsque *evt* est différent de 0 et non manquant.

Si on spécifie une liste de valeurs après *evt* (*evt*=0,2,5... par exemple), STATA considère qu'un évènement a eu lieu lorsque *evt* prend n'importe laquelle des valeurs de la liste spécifiée.

- **id(varid)** : ici *varid* correspond à une variable identifiant chaque sujet, à indiquer dans le cas où on a plus d'une ligne d'observation par sujet,
- **enter** : permet de spécifier le moment du début du suivi du sujet,
- **exit** : permet de spécifier la fin du suivi du sujet,
- **origin** : permet de spécifier le moment à partir duquel un sujet devient "à risque".

Lorsque on tape **stset**, **clear**, STATA oublie que ces données ont été déclarées comme étant des données de survie.

ATTENTION : STATA ne prendra pas en compte dans l'analyse des sujets ayant un suivi nul (exemple : la personne décède ou est perdue de vue le jour du diagnostic). Or il est important que les vécus de ces sujets

soient comptabilisés dans l'analyse. Pour qu'ils le soient, il suffit de leur attribuer un suivi minimum (exemple 1 jour=0.0333 mois=0.00274 an).

Deux commandes permettent d'obtenir un résumé des données de survie ayant été déclarées au préalable :

- **stdes** : donne une description des données de survie ayant été déclarées au préalable
- **stsum** : donne un résumé des données de survie ayant été déclarées au préalable.

2.2 Estimation de la fonction de survie et de fonctions cumulatives de risque

Le groupe de commandes commençant par **sts** dans STATA permet de générer, lister, représenter graphiquement et tester les fonctions de survie et les fonctions cumulatives de risque. Plus précisément, la commande **sts** produit des estimations de survie de Kaplan-Meier et les tests associés. Dans le cadre de la régression de Cox, elle permet d'obtenir des estimations ajustées.

- **sts list** : permet de lister les fonctions de survie estimées et autres fonctions liées,
- **sts test** : permet de tester l'égalité de la fonction de survie entre différents groupes de patients (différents tests sont disponibles, qu'il faut indiquer en option, par défaut si aucun test n'est précisé, la commande donne le test du logrank).

Voici la syntaxe :

```
sts list [if exp] [in range] [,by(groupe) at(numlist)]
sts test varlist [if exp] [in groupe] [,logrank |wilcoxon|cox ...]
```

2.3 Graphique des fonctions de survie et de risque

- **sts graph** : permet d'obtenir la représentation graphique de la fonction de survie estimée par la méthode de Kaplan-Meier.

La syntaxe est :

```
sts graph [,by(varlist) ... graph_options]
```

- **stcurve** : permet d'obtenir la représentation graphique de fonctions de survie estimées par d'autres méthodes.

2.4 Modélisation par le modèle de Cox

2.4.1 Estimation

- **stcox** : permet d'estimer un modèle de Cox à risques proportionnels,
- **streg** : permet d'estimer d'autres modèles de survie paramétrique (exponentiel, Weibull, Gompertz, lognormal, loglogistic, gamma).

Il y a de nombreuses autres commandes **st ...** permettant d'effectuer différentes analyses sur des données de survie. (cf aide de STATA : **help st**). *stcox* donne l'estimation du risque relatif et son intervalle de confiance. Pour obtenir l'estimation des paramètres, il faut ajouter la commande suivante :

```
stcox cov1, nohr
```

Pour une variable qualitative *var* à plusieurs classes, il suffit d'utiliser la commande :

```
xi: stcox i.var
```

Pour transformer la variable *cov*, en son logarithme par exemple, il faut la définir de la façon suivante :

```
gen lncov1=ln(cov)
```

Pour inclure l'interaction des covariables *cov1* et *cov2* dans le modèle, il faut également la définir :

```
gen intcov=cov1*cov2
```

2.4.2 Calculs des résidus.

L'hypothèse de log-linéarité. Le calcul des résidus des martingales se fait avec la commande suivante :

```
stcox, mgale(mg)
```

où *mg* est une nouvelle variable. Pour déterminer sous quelle forme fonctionnelle inclure les covariables, on trace *mg* en fonction de chaque covariable par la commande **lowess**.

```
lowess mg cov1.
```

On doit obtenir une droite linéaire, sinon une transformation est nécessaire. Une autre façon de calculer les résidus de martingale est :

```
predict mg, mgale
```

Pour calculer et tracer les résidus de déviance, on utilise :

```
predict mg, mgale  
predict xb, xb  
scatter mg xb
```

L'hypothèse des risques proportionnels. Pour tester l'hypothèse des risques proportionnels, les commandes sont les suivantes :

- **stphplot** ou **stcoxkm** : permettent d'évaluer graphiquement la validité de l'hypothèse des risques proportionnels (tracé de $\ln(-\ln(S(t)))$ en fonction de $\ln(t)$),
- **estat phtest** : permet de tester l'hypothèse de proportionnalité des risques. Pour l'utiliser il faut calculer les résidus de Schoenfeld et Schoenfeld standardisé avec les options **schoenfeld()** et **scaledsch()**, `stcox cov1 cov2, schoenfeld(sch*) scaledsch(sca*)`
sch1 et *sch2* sont créés par **schoenfeld(sch*)** et *sca1* et *sca2* par **scaledsch(sca*)**. Ensuite, on utilise les commandes suivantes :

```
estat phtest, detail  
estat phtest, plot(cov1)  
estat phtest, plot(cov2)
```

La première instruction donne le test des risques proportionnels. La deuxième trace les résidus en fonction de chaque covariable.

Stratification et modélisation par partie. La variable *cov2* ne respecte pas l'hypothèse des risques proportionnels. Si on veut stratifier sur la variable *cov2*, la syntaxe est la suivante :

```
stcox cov1, strata(cov2)
```

Si la variable varie avec le temps et est continue, on peut également inclure l'interaction avec la fonction *f* du temps dans le modèle de la manière suivante :

```
stcox cov1, tvc(cov1) texp(f(_t))
```

Le modèle correspondant est le suivant :

$$\lambda(t, cov1) = \lambda_0(t)e^{\beta_1 cov1 + \beta_2 cov1 * f(t)}.$$

Si on veut faire une modélisation par partie : $t \leq 5$ et $t > 5$, on crée une nouvelle variable *new* de la façon suivante :

```
generate id=_n
streset, id(id)
stsplit new, at(5) id(id)
```

new = 0 lorsque $t \leq 5$, et *new* = 5 lorsque $t > 5$. On fait alors un modèle de Cox sur chaque partie. Par exemple, pour la partie $t > 5$, on utilise les commandes suivantes :

```
gen cov2a=cov2*(new=5)
stcox cov2a
```

On peut utiliser directement :

```
gen cov25=cov2*(t>5)
stcox cov25
```

stsplit peut être utilisé pour une discrétisation en différents instants :

```
stsplit new, at(5 10 15 20)
```

Validation du modèle. Pour obtenir les résidus de Cox-Snell, on utilise la fonction **predict**. Pour cela, il faut avant tout sauver les résidus de martingales, sinon **predict** refuse de calculer les résidus de Cox-Snell.

```
stcox cov1, efron mgale(mg)
predict cs, csnell
stset cs, fail(evt)
```

On peut générer la fonction de risque cumulé de Nelson-Aalen et la tracer aux points correspondant aux résidus de Cox-Snell de la façon suivante :

```
sts gen H=na
line H cs cs, sort ytilte("") legend(cols(1))
```

Dans la commande **line**, on spécifie deux fois **cs** pour obtenir une ligne à 45°, la ligne de référence. Les deux courbes doivent se superposer.

Sujets affluents ou marginaux. La commande pour obtenir les variables `dfbeta` est :

```
stcox cov1 cov2  
predict df*, dfbeta
```

Si on veut calculer les résidus de toutes les variables, il faut spécifier **df***. Cette commande correspond donc au vecteur indice `dfbeta` sauvé dans les variables *df1* et *df2*.

```
label var df1 "dfeta cov1"  
label var df2 "dfeta cov2"
```

Pour identifier les observations avec une influence disproportionnée, on trace par exemple :

```
scatter df1 duree, yline(0) mlabel(id) msymbol(i)
```

3 Sous R

3.1 Introduction

Pour lire les fichiers de données :

```
donnee<-read.csv("donnee.csv")
```

La commande **attach()** permet de manipuler les variables séparément en les appelant par leur nom.

Pour l'analyse des données de survie sous R, il est nécessaire de charger ce package au début de chaque nouvelle session R par l'instruction suivante :

```
library(splines)
library(survival)
```

Pour voir quels sont les packages disponibles dans R :

```
library()
```

Pour voir ce que contient un package donné

```
help(package="survival")
```

Pour plus de détails sur les fonctions décrites ci-après, vous pouvez utiliser l'aide en ligne de R :

```
?survfit
```

3.2 Estimation non paramétrique de Kaplan-Meier

Les objets de survie sont créés au moyen de fonctions **Surv(duree,evt)**.

L'instruction de base pour créer un objet de survie à partir des colonnes *duree* et *evt* de la base de données *donnee* est :

```
donnee.surv=Surv(duree,evt,data=donnee)
```

Lorsqu'il y a plus de deux modalités pour *evt* ou que les modalités ne sont pas parmi celles citées plus haut, il faut préciser quelles modalités correspondent aux observations non censurées, par exemple,

```
donnee.surv=surv(duree,evt!=0,data=donnee)
```

L'estimation d'une fonction de survie à partir d'un objet de survie se fait au moyen de la fonction **survfit**. L'instruction de base est :

```
survfit(donnee.surv)
```

ou les estimations de la fonction de survie en fonction du temps :

```
summary(survfit(donnee.surv))
```

Pour tracer la courbe de survie avec les intervalles de confiance à 95% :

```
plot(survfit(donnee.surv),xlab="Durée de suivi",ylab="Taux de survie")
```

et sans les intervalles de confiance (option utile quand on trace plusieurs courbes sur un même graphique) :

```
plot(survfit(donnee.surv),conf.type="none",xlab="Durée de suivi",ylab="Taux de survie")
```

3.3 Comparaison de deux courbes par le test du logrank

Pour estimer la fonction de survie par groupe, l'instruction est la suivante :

```
donnee.survfit<-survfit(donnee.surv~groupe,data=donnee,conf.int=.95,
                      type=c("kaplan-meier","fleming-harrington"),
                      error=c("greenwood","tsiatis"),
                      conf.type=c("log","log-log","plain","none"))
```

- **conf.type** : le niveau de confiance des intervalles de confiance,
- **type** : l'estimateur de la fonction de survie employé,
- **error** : l'estimateur de la variance employé,
- **conf.type** : la formule des intervalles de confiance.

Si on dispose de deux covariables *cov1* et *cov2* de deux classes chacune alors,

```
survfit(donnee.surv~cov1+cov2,data=donnee)
```

estime la fonction de survie de 4 sous-groupes selon *cov1* et *cov2*. La fonction **survdif** permet de comparer deux ou plusieurs courbes de survie par le test du logrank.

```
survdif(donnee.surv~groupe,data=donnee)
```

Représentons maintenant, sur le même graphique, les deux courbes de survie correspondant à chaque groupe :

```
plot(survfit(donnee.surv~groupe,conf.type="none",lty=1:2,xlab="Durée de suivi",
            ylab="Taux de survie"))
```

3.4 Modèle de Cox

3.4.1 Estimation

La fonction **coxph** de R permet de modéliser le risque instantané de décès en fonction d'une liste de variables explicatives. La syntaxe est la suivante :

```
coxph(donnee.surv~groupe,data=donnee,method=c("efron","breslow","exact"))
```

L'option **method** spécifie la méthode employée pour traiter les ex-aequo. Voyons maintenant quelles instructions permettent d'ajuster un modèle de Cox sur des données. Dans l'exemple suivant, un modèle de Cox est ajusté sur l'objet de survie *donnee.surv* avec en prédicteurs les covariables *cov1* et *cov2* : les effets simples dus respectivement à *cov1* et *cov2* sont estimés.

```
cox.donnee=coxph(donnee.surv~cov1+cov2)
```

Dans l'exemple suivant, un modèle de Cox est ajusté sur l'objet de survie *donnee.surv* avec en prédicteurs les covariables *cov1* et *cov2* : les effets simples dus respectivement à *cov1* et *cov2* sont estimés ainsi que l'interaction entre *cov1* et *cov2*.

```
cox.donnee=coxph(donnee.surv~cov1*cov2)
```

Dans l'exemple suivant, un modèle de Cox est ajusté sur l'objet de survie `donnee.surv` avec en prédicteur la covariable `cov1` : l'effet simple dû à `cov1` est estimé mais l'effet dû à `cov2` n'est pas estimé : le modèle est seulement ajusté dans chaque strate.

```
cox.donnee=coxph(donnee.surv~cov1+strata(cov2))
```

3.4.2 Calcul des résidus

L'instruction R qui permet d'effectuer le calcul des différents résidus et indices est la suivante

```
residuals(cox.donnee,type=c("martingale","deviance","score","'schoenfeld","dfbetas"))
```

Les résidus de Schoenfeld sont également déterminés par la fonction suivante :

```
donnee.cox<-coxph(donnee.surv~groupe)
donnee.test<-cox.zph(donnee.cox,transform=c("km","rank","identity")
print(donnee.test)
plot(donnee.test)
```

L'option **transform** indique comment les temps de survie doivent être transformés avant d'effectuer le test. Pour des censures à droite, la valeur par défaut est **transform="km"**. Cette fonction fournit en parallèle :

1. les résultats d'un test statistique valant la pertinence de l'hypothèse pour chaque variable explicative introduite dans le modèle. Une p-value supérieure à 5% indique que la variable explicative associée n'est pas en contradiction avec l'hypothèse des risques proportionnels.
2. des graphiques donnant l'allure de l'évolution temporelle des résidus.

3.5 Sorties graphiques

Pour sauvegarder les graphiques élaborés sous R, après une instruction `plot`, taper :

```
dev.print(postscript,file="figure.ps")
```

On peut également sauvegarder les graphiques en format jpeg en tapant, avant l'instruction `plot` :

```
jpeg(filename=figure.jpeg)
plot(...)
dev.off()
```

Il est possible de partager l'écran graphique pour y afficher plusieurs figures :

```
par(mfrow=c(li,co))
```

où `li` et `co` désignent respectivement le nombre de lignes et le nombre de colonnes structurant la fenêtre graphique. Pour annuler le partage :

```
layout()
```

4 Sous SAS

Pour accéder à une table de données SAS, on doit passer par une librairie, qui elle-même est associée à un répertoire physique de windows. Ainsi, pour accéder à la table *donnee* stockée dans le répertoire "repert", on devra créer une librairie associée au répertoire "repert". On utilise pour cela l'instruction **libname**.

```
libname survie 'repert';
```

4.1 Estimation non-paramétrique

La procédure **lifetest** de SAS permet l'estimation non-paramétrique de fonction de survie, c'est-à-dire le calcul des probabilités de survie par la méthode de Kaplan-Meier ou par la méthode actuarielle à partir des données pouvant être censurées à droite. Elle permet aussi des tests de liaisons avec des covariables (test du logrank).

Sa syntaxe est :

```
proc lifetest data=surveie.donnee [options];  
time duree [*evt(liste)];  
[by variable];  
[freq variables];  
[strata variable];  
[test variable];  
run;
```

Les termes ou instructions notés dans [...] sont optionnels.

Parmi les options de cette procédure :

- **method** : permet de spécifier la méthode d'estimation de la fonction de survie : Kaplan-Meier (**KM** ou **PL** ou actuarielle (**ACT** ou **LIFE** ou **LT**). Par défaut, la procédure utilise la méthode de Kaplan-Meier.
- Pour déterminer les intervalles utilisés dans la méthode actuarielle :
 - **intervals** : donne une liste de valeurs correspondant aux bornes des intervalles, sous la forme : 1 3 5 ou 1,3,5 ou 1 to 5 by 2; le premier intervalle est supposé commencer à 0, même si 0 n'est pas spécifié.
 - **ninterval** : donne le nombre d'intervalles utilisés pour calculer la fonction de survie actuarielle (calculé par SAS).
 - **width** : donne la largeur des intervalles utilisés pour calculer la fonction de survie actuarielle.
- **plots=(S,LS,LLS,H,P)** : permet de demander le graphique de la fonction de survie estimée (**S**), de son logarithme (**LS**), de son log-log (**LLS**) et pour l'estimation actuarielle seulement, des graphiques de la fonction de risque (**H**) ou de la densité (**P**), en fonction du temps (en abscisse).

Les instructions (non obligatoires) données dans la syntaxe ci-dessus peuvent être utilisées pour :

- **freq** : indiquer l'effectif associé à chaque valeur de durée, en cas de données groupées (une ligne=plusieurs observations caractérisées par la même durée).
- **by** : réaliser des analyses séparées par groupe (sur un fichier trié par la procédure SORT).
- **test** : indiquer des covariables (en format numérique) dont on veut étudier la liaison avec la durée (entre autre par le test du logrank).
- **strata** : indiquer des covariables qualitatives définissant un découpage des analyses.

4.2 Modélisation par le modèle de Cox

4.2.1 Estimation

La procédure **phreg** permet l'estimation du modèle de Cox, supposé à risques proportionnels. Sa syntaxe est la suivante :

```
proc phreg data=survie.donnee;
model duree*evt(liste)=cov1 cov2 ...[/options];
[strata variables];
[Assess var=() / resample];
run;
```

Les covariables doivent être au format numérique qu'elles soient quantitatives ou qualitatives.

L'option **strata** permet de stratifier sur une variable. Si on veut transformer par exemple la covariable *cov1* en son logarithme :

```
proc phreg survie.donnee;
model duree*censure(liste)=logcov1[/options];
logcov1=log(cov1);
run;
```

Pour inclure une interaction entre la covariable *cov1* et *cov2*

```
proc phreg survie.donnee;
model duree*censure(liste)=intcov[/options];
intcov=cov1*cov2;
;
run;
```

4.2.2 Calcul des résidus

L'instruction **assess** crée un graphe des résidus des martingales cumulés permettant de valider la proportionnalité des risques. On obtient deux courbes, une courbe solide correspondant au score observé et des courbes en pointillés correspondant à des scores simulés lorsque l'hypothèse de proportionnalité est vérifiée. Pour valider l'hypothèse, il ne faut pas qu'il y ait d'écarts entre la courbe solide et celles en pointillés. L'option **resample** calcule la p-value des tests du supremum de type Kolmogorov, testant la validation des risques proportionnels.

On peut également vérifier l'hypothèse des risques proportionnels en utilisant les résidus de Schoenfeld de la façon suivante :

```
proc phreg data=survie.donnee;
model duree*evt(liste)=cov1 cov2 ...[/options];
output out=survie.resultat ressch=rcov 1 rcov2 lmax=infl;
run;

proc corr data=survie.resultat;
var duree rcov1 rcov2;
run;
```

La procédure **phreg** dispose d'options pour le calcul de différents résidus :

```
proc phreg data=survie.donnee;  
model duree*evt(liste)=cov1 cov2 ...[/options];  
output out=survie.resultat dfbeta=dfbeta logsurv=cs resmart=mart resdev=dev ressch=ressch lmax=infl;  
run;
```

- **logsurv** : résidus de Cox-Snell,
- **resmart** : résidus des martingales,
- **resdev** : résidus de déviance
- **lmax**, l'influence relative des observations,
- **ressch** : résidus de Schoenfeld,
- **dfbeta** : indice dfbeta.

5 Applications

5.1 Analyser des sorties STATA

Age

```
Log-rank test for equality of survivor functions
-----+-----
Age_c | Events | Events
-----+-----
Age_c | observed | expected
-----+-----
moins de 60 ans | 33 | 25.29
60 ans et plus | 53 | 60.71
-----+-----
Total | 86 | 86.00
-----+-----
chi2(1) = 3.37
Pr > chi2 = 0.0662

Cox regression -- Breslow method for ties
-----+-----
No. of subjects = 178 | Number of obs = 178
No. of failures = 86
Time at risk = 3839.470219
-----+-----
Log likelihood = -385.20408 | LR chi2(1) = 3.21
Prob > chi2 = 0.0732
-----+-----
             _t_ | Hag. Ratio | Std. Err. | z | P>|z| | [95% Conf. Interval]
-----+-----
             Age_c | .6651158 | .1487698 | -1.82 | 0.068 | -.4290468 | 1.031074

Cox regression -- Breslow method for ties
-----+-----
No. of subjects = 178 | Number of obs = 178
No. of failures = 86
Time at risk = 3667.318269
-----+-----
Log likelihood = -385.20408 | LR chi2(1) = 3.21
Prob > chi2 = 0.0732
-----+-----
             _t_ | Coef. | Std. Err. | z | P>|z| | [95% Conf. Interval]
-----+-----
             Age_c | -.4077941 | .2236751 | -1.82 | 0.068 | -.8461893 | .030601
```

FIGURE 1 – Analyse univariée, âge.

Albumine

```

Log-rank test for equality of survivor functions
-----+-----
Events      Events
albumine_c | observed expected
-----+-----
<35        | 34      30.82
normale de 35 à 50 | 36      39.18
-----+-----
Total      | 70      70.00
-----+-----
chi2(1) = 0.60
Pr>chi2 = 0.4403

Cox regression -- Breslow method for ties
-----+-----
No. of subjects = 149      Number of obs = 149
No. of failures = 70
Time at risk = 3127.464064
-----+-----
LR chi2(1) = 0.59
Log likelihood = -305.05141      Prob > chi2 = 0.4418
-----+-----
      _t_ | Haz. Ratio   Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
albumine_c | .8304406   .2003822   -0.77   0.441   .5175077   1.332602
-----+-----

Cox regression -- Breslow method for ties
-----+-----
No. of subjects = 149      Number of obs = 149
No. of failures = 70
Time at risk = 2983.359337
-----+-----
LR chi2(1) = 0.59
Log likelihood = -305.05141      Prob > chi2 = 0.4418
-----+-----
      _t_ | Coef.   Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
albumine_c | -.1857989   .2412963   -0.77   0.441   -.6587309   .2871331
-----+-----

```

FIGURE 2 – Analyse univariée, Albumine.

Analyse univariée :

1. Comparer le modèle de Cox avec le test du logrank. Dans quel cas donnent-ils approximativement le même résultat ?
2. Retrouver la valeur du coefficient du risque relatif et le calcul des différents intervalles de confiance.
3. Interpréter les résultats.

MULTIVARIE

```

Cox regression -- Breslow method for ties
No. of subjects = 140      Number of obs = 1719
No. of failures = 689
Time at risk = 2741.420942
LR chi2(8) = 76.439
Log likelihood = -258.64289      Prob > chi2 = 0.00009

-----+-----
      _t_ | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-----+-----
  ITauxAFP ~2 |   .91615   .330024   -0.24   0.808   .4522087   1.856069
  ITauxAFP ~3 |   1.053553   .50595    0.11   0.913   .4110383   2.700416
  ITauxAFP ~4 |   2.535122   .8826863   2.67   0.008   1.281224   5.016176
  TailleTum c |   1.985842   .6985157   1.95   0.051   .9966356   3.956882
  Ichild_c 2 |   1.136196   .3454737   0.42   0.675   .6260881   2.061916
  Ichild_c 3 |   4.15282    1.753523   3.37   0.001   1.8152    9.500835
  CHCinfilt |   4.49782    1.595263   4.24   0.000   2.244411   9.013673
  UniMultifo c |   1.472104   .1918652   2.97   0.003   1.140245   1.900549
-----+-----

avec
  ITauxAFP ~1 : le taux d'AFP est strictement inférieur à 20
  ITauxAFP ~2 : le taux d'AFP est compris entre 20 et 200
  ITauxAFP ~3 : le taux d'AFP est compris entre 200 et 500
  ITauxAFP ~4 : le taux d'AFP est supérieur ou égal à 500

et
  Ichild_c 1 : le score de CHILD est A ou D
  Ichild_c 2 : le score de CHILD est B
  Ichild_c 3 : le score de CHILD est C

```

FIGURE 3 – Analyse multivariée.

Analyse multivariée :

1. Expliquer la signification des variables $TauxAFP_{\sim i}$, $i = 1, \dots, 4$. Quel est la variable de référence?
2. Interpréter les résultats.

5.2 Exercices

On dispose de cinq jeux de données `survie1`, `survie2`, `survie3`, `survie4` et `survie5`. Les données sont censurées à 2 ans.

5.2.1 Estimation de la fonction de survie / test du logrank

Exercice 1 (base de donnée 1) Le jeu de données **survie1** contient les variables :

- *duree* : durée de suivi (en jours),
- *état* : état à la fin du suivi (DCD, VIV ou PDV),
- *sexe* : sexe (1=homme, 2=femme).

1. Avant toute analyse de survie, réaliser une analyse descriptive rapide des données.
2. On veut étudier la survenue du décès. Créer la variable "deces", correspondant à la survenue du décès déduite de la variable état. Estimer la fonction de survie par la méthode de Kaplan-Meier, tous groupes confondus. Représenter la courbe de survie correspondante.
3. Réaliser une étude par groupe. Puis comparer la survie des deux groupes.

5.2.2 Modèle de Cox : Hypothèse des risques proportionnels

Exercice 2 (base de données 2) Le jeu de données **survie2** contient

- *durée* : durée de suivi (en jours),
- *maladie* : survenue de la maladie (0=non, 1=oui),
- *trait* : prise du traitement (1=traitement n°1, 2=traitement n°2) ;,

1. On veut étudier l'effet du traitement sur la survenue d'une maladie. Vérifier l'hypothèse des risques proportionnels pour la variable traitement. Si elle n'est pas vérifiée, quelle méthode utilisez-vous ?
2. Evaluer l'effet du traitement sur la survenue de la maladie, en utilisant le modèle de Cox.

Exercice 3 (base de données 3) Le jeu de données **survie3** contient

- *duree* : durée de suivi (en mois),
- *maladie* : survenue de la maladie (0=non, 1=oui),
- *trait* : prise du traitement (1=traitement n°1, 2=traitement n°1) ;,
- *sexe* : sexe (1=homme, 2=femme).

1. On veut étudier l'effet du traitement sur la survenue d'une maladie. Vérifier l'hypothèse des risques proportionnels pour la variable *traitement* et la variable *sexe*.
2. Construisez un modèle de Cox stratifié selon la variable *sexe*.

Le jeu de données **survie4** contient

- *id* : identifiant,
- *duree* : durée de suivi (en mois),
- *deces* : décès (0=non, 1=oui),
- *trait* : prise du traitement (0=non, 1=oui) ;,
- *dos1* : valeur d'un dosage biologique 1,
- *dos2* : valeur d'un dosage biologique 2,
- *sexe* : sexe du patient.

5.2.3 Hypothèse de log-linéarité

Exercice 4 (base de données 4)

1. Déterminer la forme fonctionnelle à utiliser pour la variable dosage 1.
2. Construisez un modèle de Cox avec la variable traitement et la variable dosage 1 ou sa transformation.
3. Vérifier que le modèle est valide en utilisant les résidus de Cox-Snell.
4. Utiliser les résidus de déviance et interpréter.

5.2.4 Sujets influents ou marginaux

Exercice 5 (base de données 4)

1. Construisez un modèle de Cox avec les variables traitement, dosage 1 et dosage 2.
2. Tracer les résidus *dfbeta* pour chacune des covariables.

6 Problème

Après avoir étudié les différentes fonctions et procédures de ces logiciels sur des exemples simples, vous allez maintenant travailler sur une étude plus conséquente.

Le fichier de données **survie5** contient 1000 observations et 6 variables. On étudie ici la survenue des symptômes pour le processus d'une maladie.

Les variables sont les suivantes :

- *id* : identifiant du patient,
- *datenaiss* : date de naissance du patient,
- *sympt* : la survenue de symptômes,
- *datesymp* : date de la survenue des symptômes,
- *datedn* : date des dernières nouvelles,
- *sexe* : le sexe du patient (1=homme, 2=femme),
- *trait* : la prise de traitement (0=non, 1 =oui),
- *dosage* : la valeur d'un dosage biologique,
- *stade* : stade de la maladie au début de l'étude (3 états).

Les données sont censurées à 36 mois.

1. Faites une analyse descriptive des données (âge, sexe...).
2. Les deux groupes (traitement versus placebo) sont-ils comparables ? Pourquoi cela est-il important ?
3. Créez la variable "durée de suivi".
4. Estimez la fonction de survie par la méthode de Kaplan-Meier, donnez la survie cumulative à 1 an, 2 ans, 3 ans, 5 ans. Faire une représentation graphique de la survie
5. Reliez la survie avec chacune des covariables (approche bivariée). Quelles méthodes pouvez-vous utiliser ?
6. Vérifier la log-linéarité des variables continues. Effectuez les changements si nécessaire.
7. Vérifier l'hypothèse de proportionnalité des risques pour chaque variable par différentes méthodes.
8. Construisez un modèle multiple de la survie en fonction des covariables.
9. Utilisez les divers résidus et les interpréter.