

TP d'analyse de données de survie

Correction

Valérie Garès

5 novembre 2013

1 Analyser des sorties STATA

Analyse univariée :

1. Le test du logrank est identique au modèle de Cox dans le cas univarié pour une variable à deux classes.
2. Coef = ln(Haz Ratio). L'intervalle de confiance de coef se calcule de la manière suivante :

$$[I1_m, I1_M] = [\text{Coef} - z_{1-\frac{\alpha}{2}} \times \text{Std. Err}], \text{Coef} + z_{1-\frac{\alpha}{2}} \times \text{Std. Err}]$$

avec $z_{1-\frac{\alpha}{2}}$ le quantile d'ordre $1 - \frac{\alpha}{2}$ de la loi normale. $z_{1-\frac{\alpha}{2}} = 1.96$. L'intervalle de confiance de Haz Ratio est alors :

$$[I2_m, I2_M] = [e^{I1_m}, e^{I1_M}]$$

3. L'âge et l'albumine n'influent pas sur la survie. 1 est dans l'intervalle de confiance.

Analyse multivariée :

1.

$$TauxAFP_{\sim 1} = \begin{cases} 1 & \text{si le taux d'AFP est strictement inférieur à 20} \\ 0 & \text{si le taux d'AFP est supérieur à 20,} \end{cases}$$

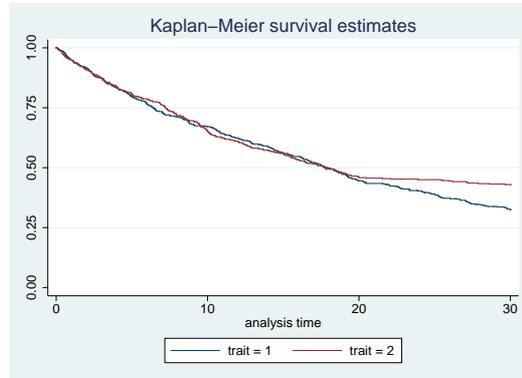
La variable de référence est $TauxAFP_{\sim 1}$.

2. Les variables suivantes influent significativement au seuil de 5% sur la survie : le taux d'AFP, le score de CHILD, la taille de la tumeur, la présence d'un CHC infiltrant et le type de tumeur. Un patient ayant un taux d'AFP supérieur à 500 a un risque de décès significatif environ 2.5 fois plus élevé qu'un patient ayant un taux d'AFP inférieur à 20 ($p = 0.008$). Un patient ayant un score de CHILD de C a un risque de décès significatif environ 4 fois plus élevé qu'un patient ayant un score de A ($p = 0.001$). Une tumeur de taille supérieure à 5 cm entraîne un risque de décès significatif environ 2 fois plus élevé qu'une tumeur de taille inférieure à 5 cm ($p = 0.051$). La présence d'un CHC infiltrant entraîne un risque de décès significatif environ 4.5 fois plus élevé ($p < 0.0001$). Une tumeur multifocale entraîne un risque de décès significatif environ 1.5 fois plus élevé qu'une tumeur unifocale ($p = 0.003$).

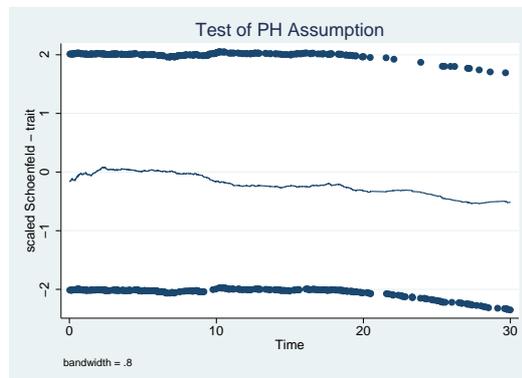
2 Exercices

Exercice 1 En utilisant le test du logrank, on observe un effet significatif du traitement.

Exercice 2 L'hypothèse des risques proportionnels n'est pas vérifiée pour la variable traitement. Ci-dessous, le graphe des courbes de survie de chaque groupe de traitement.



Et celui des résidus de Schoenfeld :



On veut étudier l'effet du traitement sur la survenue de la maladie. Entre 0 et 20, l'hypothèse des risques proportionnels est vérifiée, entre 20 et 30, elle est vérifiée également. On utilise la modélisation par partie : on réalise deux modèles de Cox, un entre 0 et 20 et un entre 20 et 30 (on ne peut pas choisir la stratification car on a besoin d'une p-value et d'un RR).

Exercice 3 On trace $-\ln(-\ln(\text{survie}))$ en fonction de $\ln(t)$ pour la variable traitement et la variable sexe. Pour que l'hypothèse des risques proportionnels soit vérifiée, il faut obtenir deux droites parallèles.

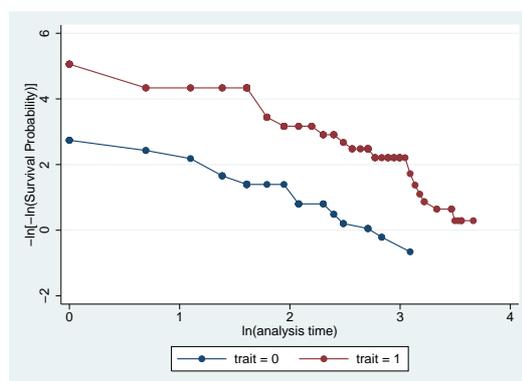


FIGURE 1 – $-\ln(-\ln(\text{survie}))$ en fonction de $\ln(t)$ par groupe de traitement

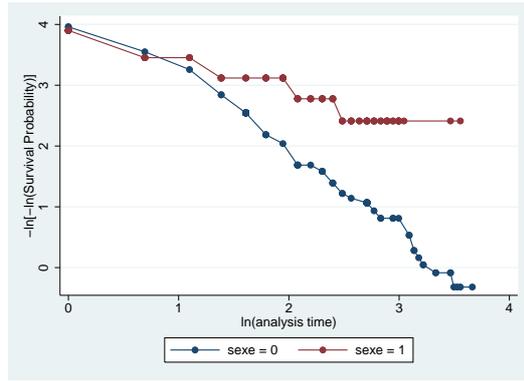


FIGURE 2 – $-\ln(-\ln(\text{survie}))$ en fonction de $\ln(t)$ par sexe

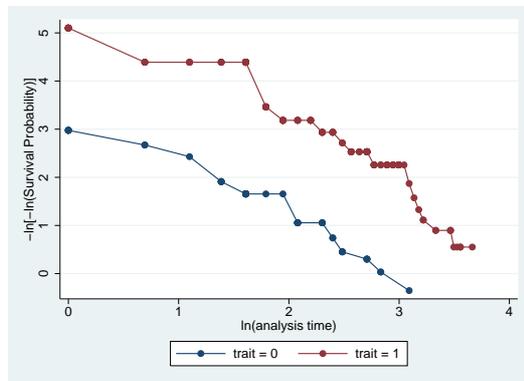


FIGURE 3 – $-\ln(-\ln(\text{survie}))$ en fonction de $\ln(t)$ par groupe de traitement conditionnellement au sexe

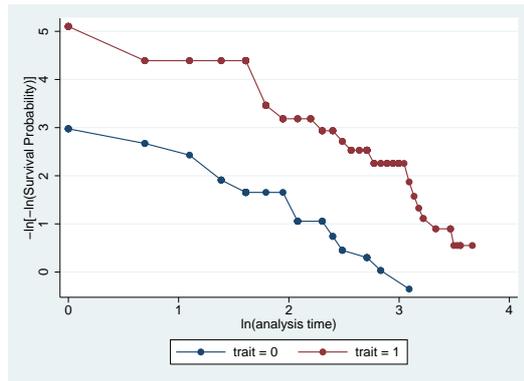


FIGURE 4 – $-\ln(-\ln(\text{survie}))$ en fonction de $\ln(t)$ par sexe conditionnellement au traitement

Les graphes montrent que l'hypothèse des risques proportionnels n'est pas vérifiée pour la variable sexe bien que le test des résidus de Schoenfeld ne soit pas significatif. Il est important de regarder plusieurs méthodes. Ci-dessous, le graphe des résidus de Schoenfeld pour les deux variables.

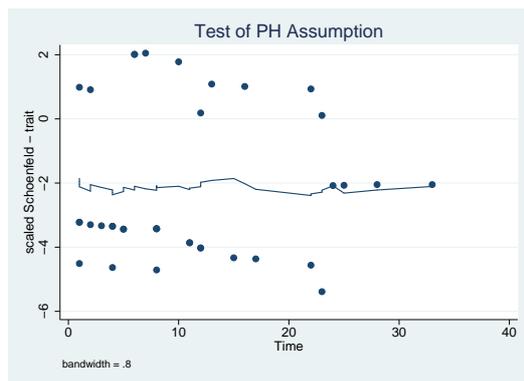


FIGURE 5 – Traitement

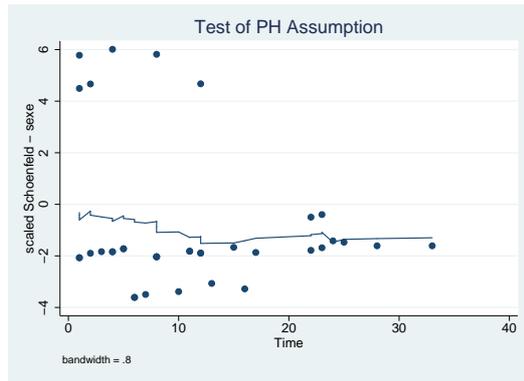
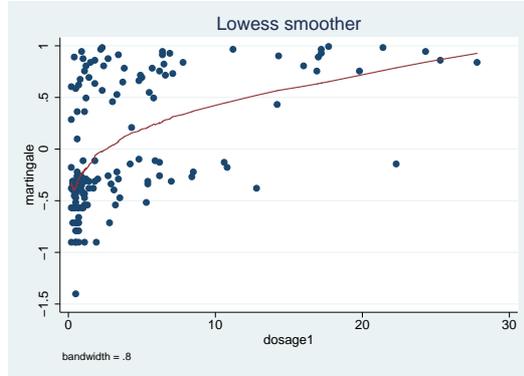


FIGURE 6 – Sexe

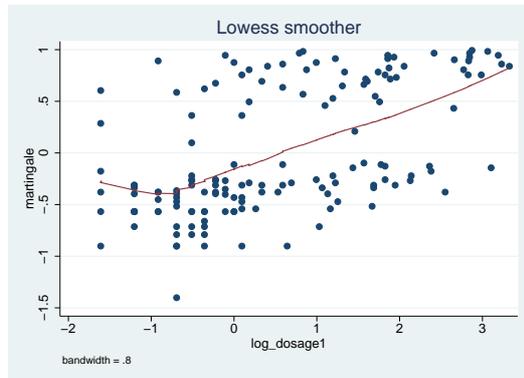
On choisit la stratification par rapport à la variable "sexe" car on s'intéresse seulement à l'effet du traitement et non du sexe.

Exercice 4

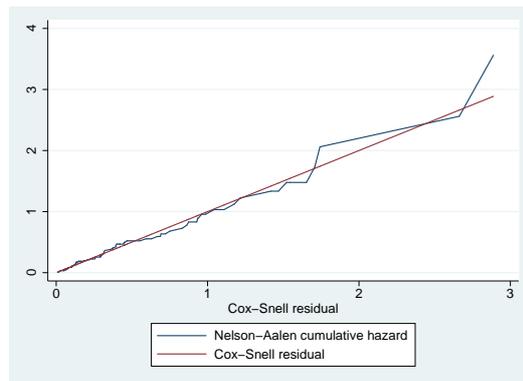
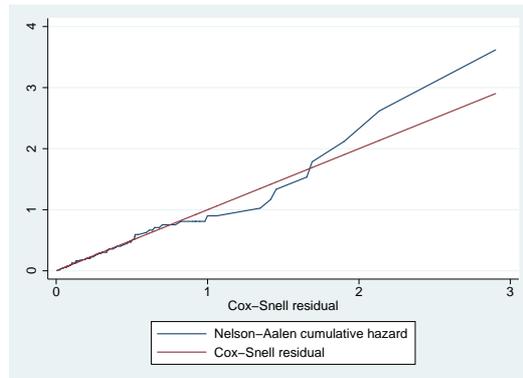
- On trace les résidus de martingales en fonction de la variable "dosage1".



Il est nécessaire de transformer la variable en son logarithme pour avoir des résultats plus linéaires.

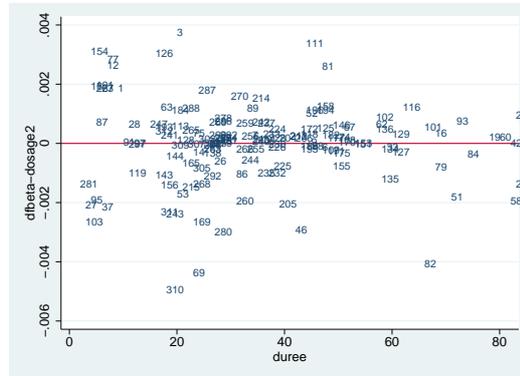


- Ci-dessous le graphe de la fonction de risque cumulé de Nelson-Aalen calculée aux points des résidus de Cox-Snell (pointillé) à comparer avec la fonction de risque cumulé d'une exponentielle de paramètre 1 (bissectrice) lorsque la variable dosage1 est incluse dans le modèle puis la variable log(dosage1).



Les deux courbes doivent se superposer. On valide le modèle avec la variable log(dosage1).

Exercice 5 Voici un exemple du tracé des résidus Dfbeta pour la variable dosage2.



Problème

1. Il faut créer la variable âge.
2. Les patients recevant le traitement et ceux recevant placebo sont comparables (pas de p-value significative). Il faut avoir des échantillons indépendants.
3. Créer la variable "durée de suivi" en considérant si le patient est perdu de vue, si les symptômes sont apparus ou pas avant 36 mois.
4. Le test du logrank pour les variables en classe. On peut aussi utiliser le modèle de Cox.
5. L'hypothèse de log-linéarité est vérifiée pour les variables continues.
6. L'hypothèse des risques proportionnels est vérifiée pour chacune des variables.
7. Il faut penser que la variable stade est en trois classes.
8. Il n'y a aucun sujet influent ou marginal.

3 Commandes

3.1 Avec STATA

```
* Exercice 1
*Variable dc
gen dc=1 if etat=="DCD" & duree<=24
replace dc=0 if dc==.

replace duree=24 if duree>24

*Analyse descriptive des données
tab dc
tab etat
tab sexe
tab sexe dc*

*Déclaration des données de survie
stset duree, failure(dc==1)

*Estimation et représentation de la fonction de survie
sts list
sts graph

*Comparaison des deux groupes
sts list ,by(sexe)
sts test sexe ,logrank
sts graph ,by(sexe)

*Modèle de Cox
*Vérification des risques proportionnels
sts graph ,by(sexe)
stcox sexe, schoenfeld(sch*) scaledsch(sca*)
estat phtest, detail
estat phtest, plot(sexe)

stphplot, by(sexe)

stcox sexe
stcox sexe, nohr

*****

*Exercice 2
```

```

stset duree, failure(maladie==1)

*Vérification des risques proportionnels
sts graph ,by(trait)
stcox trait, schoenfeld(sch*) scaledsch(sca*)
estat phtest, detail
estat phtest, plot(trait)

stphtest, log detail
stphtest, log plot(trait) yline(0)

stphplot, by(trait)
stcoxkm, by(trait)

*Interaction traitement temps
gen dtrait=duree*trait
stcox trait dtrait
stcox trait, tvc(trait) texp(duree)
* on peut utiliser toutes les fonctions de temps C'est la plus commune

*Modélisation du temps
gen traits20=trait*(t>20)
stcox traits20

*****
*Exercice 3
stset duree, failure(maladie==1)

*Vérification des risques proportionnels
stphplot, by(trait)
stphplot, by(sexe)

*Résidus de schenfeld
stcox trait sexe, schoenfeld(sch*) scaledsch(sca*)
estat phtest, detail
estat phtest, plot(trait)
estat phtest, plot(sexe)

*Stratification de la variable sexe
stcox trait, strata(sexe)

*****

```

```

*Exercice 4
stset duree, failure(deces==1)

*Résidus des martingales
stcox trait, mgale(mg) efron estimate
lowess mg dosage1

*****autre façon
stcox trait
predict mgg, mgale
*calcule la même chose que mg=mgg
lowess mgg dosage1

*Transformation de dosage1 en son logarithme
gen log_dosage1=log(dosage1)
lowess mg log_dosage1

*Modèle de Cox
stcox trait log_dosage1

*Validation du modèle
stcox trait log_dosage1, efron mgale(mg2)
*résidus de Cox Snell
predict cs,csnell
*on redéfinit la variable de survie
stset cs, fail(deces)
*fonction cumuléé nelson aalen
sts gen H=na
*bissectrice cs cs
*H au point cs
line H cs cs,sort ytitle("") legend(cols(1))

stset duree, failure(deces==1)
stcox trait dosage1, efron mgale(mg3)
predict cs1,csnell
stset cs1, fail(deces)
sts gen H1=na
list H1
line H1 cs1 cs1,sort ytitle("") legend(cols(1))

*résidus de déviance
stcox trait log_dosage1

```

```

predict mg1, mgale
predict xb, xb
scatter mg1 xb
*non symétrie par rapport à 0

*résidus de déviance symétrie par rapport à 0
predict dev, deviance
scatter dev xb
*On accepte si compris dans [-2,2]

*****

*Exercice 5
stset duree, failure(deces==1)
stcox trait log_dosage1
lowess mg dosage2

stcox trait log_dosage1 dosage2, efron esr(esr*)
predict df*, dfbeta

label var df1 "dfbeta-traitement"
label var df2 "dfbeta-log_dosage1"
label var df3 "dfbeta-dosage2"

scatter df1 duree, yline(0) mlabel(id) msymbol(i)
scatter df2 duree, yline(0) mlabel(id) msymbol(i)
scatter df3 duree, yline(0) mlabel(id) msymbol(i)

*****

*Problème

*Age
list datenaiss
gen datedebb=date(datedeb,"DMY")
gen datenaissb=date(datenaiss,"DMY")
list datenaissb
gen age=round((datedebb-datenaissb)/365.256)
list age

*Analyse descriptive des données
tab sympt
tab trait

```

```

tab sexe
tab stade
sum age
sum dosage

*Comparaison des données
*Variables quantitatives
*Vérification que les variables suivent la loi normale
sktest age
histogram age
sktest dosage
histogram dosage
sdtest age , by(trait)
ttest age, by(trait)
sdtest dosage, by(trait)
ttest dosage, by(trait)
*ranksum dosage, by(trait) *si la variable n'est pas normale
*Variables qualitatives
tab sexe trait , chi2
tab stade trait , chi2

*Création de la variable durée en mois
gen datednd=date(datedn,"DMY")
gen datesymptd=date(datesympt,"DMY")
gen duree=(datednd-datedebb)/30.5
replace duree=(datesymptd-datedebb)/30.5 if sympt==1
replace sympt=0 if duree>36
replace duree=36 if duree>36

list duree

*Estimation et représentation de la fonction de survie
stset duree,failure(sympt==1)
sts list
sts graph

*Analyse bivariée
sts list ,by(trait)
sts test trait ,logrank
sts graph ,by(trait) saving(,replace)
sts list ,by(sexe)
sts test sexe ,logrank
sts graph ,by(sexe) saving(,replace)

```

```

sts list ,by(stade)
sts test stade ,logrank
sts graph ,by(stade) saving(,replace)
stcox trait
stcox sexe
stcox age
stcox dosage
xi:stcox i.stade

*Modèle multivarié
xi:stcox trait sexe age dosage i.stade

*Hypothèse des risques proportionnels
stcox trait sexe age dosage stade , schoenfeld(sch*) scaledsch(sca*)
estat phtest, detail
estat phtest, plot(sexe)
estat phtest, plot(trait)
estat phtest, plot(age)
estat phtest, plot(dosage)
estat phtest, plot(stade)

*Résidus des martingales
xi:stcox trait sexe age dosage i.stade
predict mg,mgale
lowess mg age
lowess mg dosage

*Résidus cox snell
predict cs, csnell
stset cs, fail(sympt)

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

*résidus de déviance
xi:stcox trait sexe age dosage i.stade
predict xb, xb
predict dev, deviance
scatter dev xb

*Indices dfbeta
xi:stcox trait sexe age dosage i.stade , efron esr(esr*)
set matsize 2000

```

```
predict df*, dfbeta
```

```
label var df1 "dfbeta trait"
```

```
label var df2 "dfbeta sexe"
```

```
label var df3 "dfbeta age"
```

```
label var df4 "dfbeta dosage"
```

```
label var df5 "dfbeta stade"
```

```
scatter s1 duree, yline(0) mlabel(id) msymbol(i)
```