

Modèles d'Analyse de la Survie

Le Modèle de Cox

Pr Roch Giorgi

 roch.giorgi@univ-amu.fr

Modèles de régression

- Différentes méthodes existes pour estimer la distribution de durée de survie
- Possibilité de comparer (tester) des distributions
- Besoin de modèles de régression pour gérer simultanément plusieurs co-variables

Modèle de Cox (1)

$$\lambda(t, \mathbf{z}) = \lambda_0(t) \exp(\boldsymbol{\beta} \mathbf{z})$$

- $\lambda_0(t)$: taux de mortalité de base
- \mathbf{z} : vecteur de covariables
- $\boldsymbol{\beta}$: vecteur de paramètres associés à \mathbf{z}

Modèle de Cox (2)

- C'est un **modèle linéaire généralisé** pour la probabilité de survie

Taux cumulé : $\Lambda(t, \mathbf{z}) = \Lambda_0(t) \exp(\boldsymbol{\beta}\mathbf{z})$

$$\Lambda(t) = \int_0^t \lambda(u) du$$

Survie : $S(t, \mathbf{z}) = \exp(-\Lambda(t, \mathbf{z})) = S_0(t)^{\exp(\boldsymbol{\beta}\mathbf{z})}$

$$\text{Log}[S(t, \mathbf{z})] = -\Lambda_0(t) \exp(\boldsymbol{\beta}\mathbf{z})$$

$$\text{Log}[-\text{Log}[S(t, \mathbf{z})]] = \text{Log}[\Lambda_0(t)] + \boldsymbol{\beta}\mathbf{z}$$

Hypothèse des Taux Proportionnels (1)

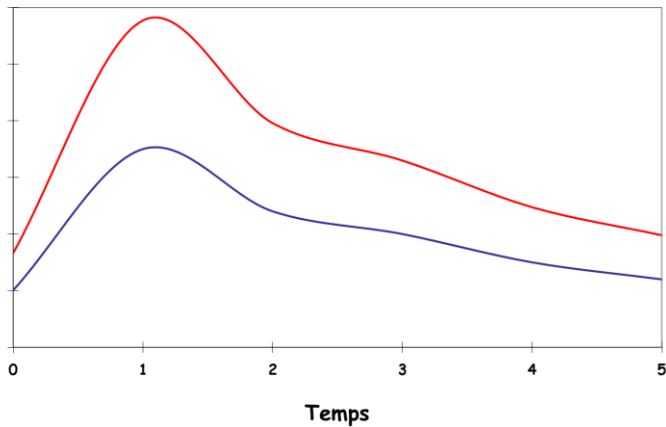
- Exemple : $z = \begin{cases} 1, \text{ femme} \\ 0, \text{ homme} \end{cases}$ $\lambda(t, \mathbf{z}) = \lambda_0(t) \exp(\beta \mathbf{z})$
- Taux de décès chez les femmes : $\lambda(t, z = 1) = \lambda_0(t) \exp(\beta)$
- Taux de décès chez les hommes : $\lambda(t, z = 0) = \lambda_0(t)$
- Taux relatif de décès des femmes r/r hommes : $\exp(\beta)$

Taux Relatif de décès indépendant du temps

$$\frac{\lambda(t, z_1, \dots, z_i, \dots, z_m)}{\lambda(t, z_1, \dots, 0, \dots, z_m)} = \exp(\beta_i z_i)$$

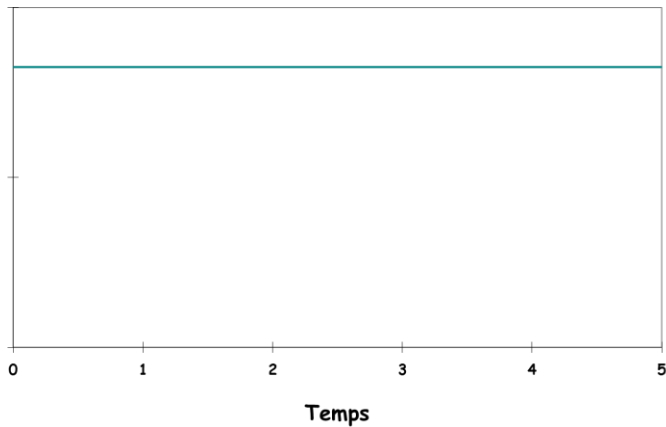
Hypothèse des Taux Proportionnels (2)

Modèle de Cox : $\lambda_{Exp}(t) = \lambda_{Contr}(t) \exp(\beta_{Exp})$



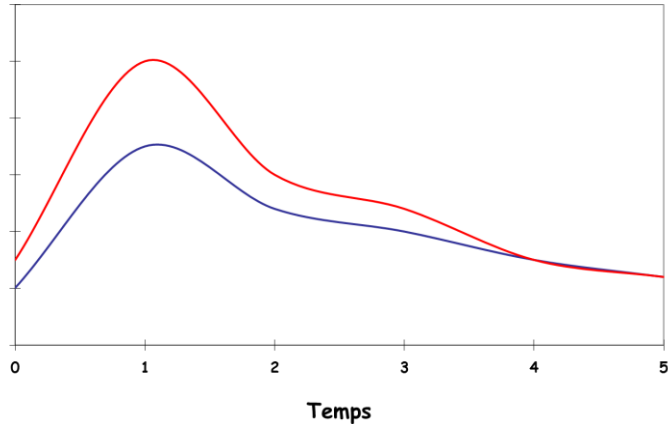
$\lambda_{Exp}(t)$

$\lambda_{Contr}(t)$



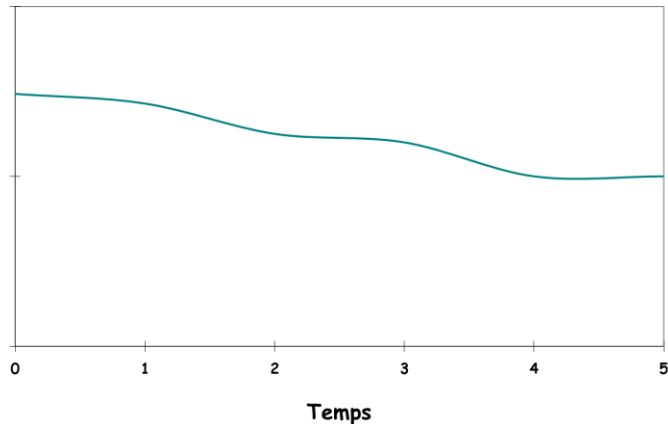
$$HR(t) = \frac{\lambda_{Exp}(t)}{\lambda_{Contr}(t)} = \exp(\beta_{Exp})$$

Taux Non Proportionnels



$$\lambda_{Exp}(t)$$

$$\lambda_{Contr}(t)$$

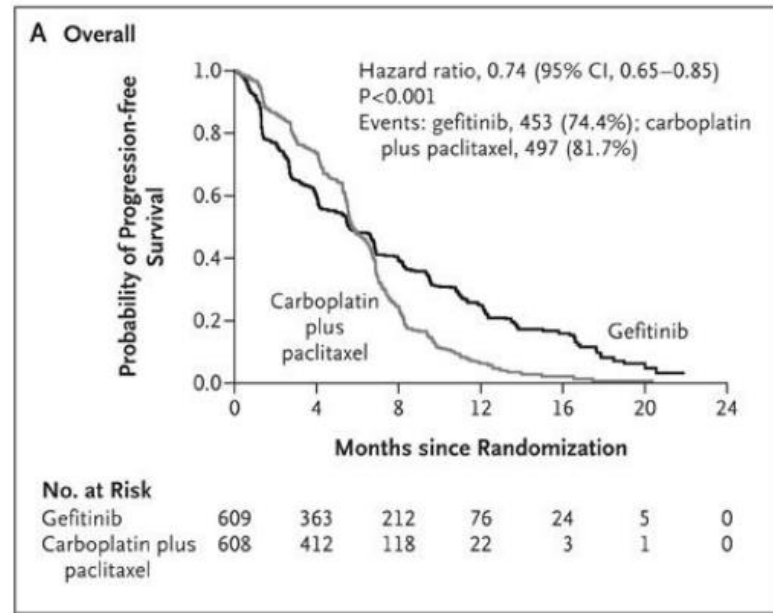


$$HR(t) = \frac{\lambda_{Exp}(t)}{\lambda_{Contr}(t)} = \exp(\beta_{Exp}(t))$$

$$\lambda_{Exp}(t) = \exp(\beta_{Exp}(t)) \lambda_{Contr}(t)$$

Hypothèse des Taux Proportionnels (3)

- L'effet d'un traitement reste le même à 1 an, 2 ans, 5 ans, 10 ans...
 - Pas forcément



Kaplan-Meier curves for progression-free survival

Source : Mok TS, Wu YL, Thongprasert S, Yang CH, Chu DT, Saijo N, Sunpaweravong P, Han B, Margono B, Ichinose Y, Nishiwaki Y, Ohe Y, Yang JJ, Chewaskulyong B, Jiang H, Duffield EL, Watkins CL, Armour AA, Fukuoka M. Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N Engl J Med* 2009;361(10):947-57.

Hypothèse des Taux Proportionnels (4)

- Modélisation à tort de l'effet d'un facteur par un modèle à taux proportionnel
 - Estimation d'un effet moyen
 - Possible conclusion à tort à l'absence d'effet

Hypothèse de Linéarité (1)

- Exemple : $z = \text{age}$

$$\lambda(t, \mathbf{z}) = \lambda_0(t) \exp(\boldsymbol{\beta} \mathbf{z})$$

$$z_1 = 26 \text{ et } z_0 = 25 \text{ ou } z_1 = 83 \text{ et } z_0 = 82$$

$$\text{On a : } \frac{\lambda(t, z = z_1)}{\lambda(t, z = z_0)} = \exp(\beta(z_1 - z_0))$$

$$\text{et donc : } \text{Log} \left(\frac{\lambda(t, z = z_1)}{\lambda(t, z = z_0)} \right) = \beta(z_1 - z_0)$$

Le Log du taux relatif de décès est une fonction linéaire des covariables

$$\log \left(\frac{\lambda(t, z_1, \dots, z_{i1}, \dots, z_m)}{\lambda(t, z_1, \dots, z_{i0}, \dots, z_m)} \right) = \beta_i (z_{i1} - z_{i0})$$

Hypothèse de Linéarité (2)

- Modélisation à tort de l'effet d'un facteur par une fonction linéaire
 - Estimations biaisées
 - Possible conclusion à tort à l'absence d'effet

Modèle de Cox

$$\lambda(t, \mathbf{z}) = \lambda_0(t) \exp(\boldsymbol{\beta} \mathbf{z})$$

- $\exp(\beta_j)$: taux relatif des sujets pour lesquels $z_j=1$ par rapport à ceux pour lesquels $z_j=0$, toutes choses égales par ailleurs
 - $\exp(\beta_j) > 1$: effet néfaste
 - $\exp(\beta_j) = 1$: pas d'effet
 - $\exp(\beta_j) < 1$: effet protecteur
- $\lambda_0(t)$: taux de mortalité de base ($\mathbf{z}=0$)
- Estimation des paramètres : méthode du maximum de vraisemblance

Maximum de Vraisemblance

- Vraisemblance d'une valeur donnée
 - Probabilité d'obtenir une valeur telle que celle observée
- Estimateur du maximum de vraisemblance
 - Estimateur qui associe aux observations la valeur pour laquelle la probabilité de l'observation est la plus forte dans le modèle
- Vraisemblance d'un modèle
 - Valeur de la vraisemblance des estimations du maximum de vraisemblance de ses paramètres

Vraisemblance Partielle de Cox (1)

- A l'instant t_i ($i=1, \dots, n$ décès), il y a \mathcal{R}_i individus encore à risque
- La probabilité de décès en t_i de chaque sujets j est

$$\lambda_0(t_i) \exp(\beta z_j) \Delta t$$

- La probabilité que ce soit le sujet i qui décède est

$$v_i(\beta) = \frac{\lambda_0(t_i) e^{\beta z_i} \Delta t}{\sum_{j \in \mathcal{R}_i} \lambda_0(t_i) e^{\beta z_j} \Delta t} = \frac{e^{\beta z_i}}{\sum_{j \in \mathcal{R}_i} e^{\beta z_j}}$$

$$V(\beta) = \prod_{i=1}^n v_i(\beta) = \prod_{i=1}^n \frac{e^{\beta z_i}}{\sum_{j \in \mathcal{R}_i} e^{\beta z_j}}$$

Vraisemblance Partielle de Cox (2)

- Les censures ne participent pas au calcul de la vraisemblance :
 - Hypothèse de censures non-informatives
 - Hypothèse d'indépendance entre les temps de censures et les temps de décès
- La vraisemblance partielle ne dépend pas de $\lambda_0(t)$

$$V(\beta) = \prod_{i=1}^n v_i(\beta) = \prod_{i=1}^n \frac{e^{\beta z_i}}{\sum_{j \in \mathcal{R}_i} e^{\beta z_j}}$$

Estimation des Paramètres (β)

- Maximisation de la log-vraisemblance

$$L(\beta) = \sum_{i=1}^n \left[\beta z_i - \log \left(\sum_{j \in \mathcal{R}_i} e^{\beta z_j} \right) \right]$$

- L'estimation du maximum de vraisemblance est la valeur $\hat{\beta}$ de β qui rend maximum $L(\beta)$ (ou $V(\beta)$)

- Au maximum : $U(\hat{\beta}) = \frac{\partial L(\hat{\beta})}{\partial \beta} = 0$

- On a, pour la k^{ème} composante du vecteur score :

$$U_k(\beta) = \sum_{i=1}^n \left[z_{ki} - \frac{\sum_{j \in \mathcal{R}_i} e^{\beta z_j} z_{kj}}{\sum_{j \in \mathcal{R}_i} e^{\beta z_j}} \right] = \sum_{i=1}^n [z_{ki} - \bar{z}_{ki}(\beta)]$$

Estimation des Paramètres

- Cox : *a priori* pour les β
- Possibilité d'estimation du taux cumulé de mortalité de base : généralisation de l'estimateur de Nelson (d/n)

$$\hat{\Lambda}_0(t) = \sum_{t_i < t} \frac{d_i}{\sum_{j \in \mathcal{R}_i} e^{\hat{\beta} z_j}}$$

- Estimation de la fonction de survie

$$\hat{S}(t_i, \mathbf{z}) = \exp \left\{ -\hat{\Lambda}_0(t_i) \exp(\hat{\beta} \mathbf{z}) \right\}$$

Tests de l'Hypothèse Nulle des β (1)

$$H_0 : \beta = \beta^{(0)}$$

Dans le cas d'une seule variable

- Wald (maximum de vraisemblance)

$$\hat{\beta}^2 / \text{Var}(\hat{\beta}) \sim \chi^2 \text{ à 1 ddl}$$

- Rapport de vraisemblance

$$2(L(\hat{\beta}) - L(\beta^{(0)})) \sim \chi^2 \text{ à 1 ddl}$$

- Score

$$U(\beta^{(0)})^2 / \text{Var}[U(\beta^{(0)})] \sim \chi^2 \text{ à 1 ddl}$$

Tests de l'Hypothèse Nulle des β

$$H_0 : \beta = \beta^{(0)} \quad \beta = 1, \dots, p$$

Généralisation

- Wald (maximum de vraisemblance)

$$\left(\hat{\beta} - \beta^{(0)} \right)^T I \left(\hat{\beta} \right)^{-1} \left(\hat{\beta} - \beta^{(0)} \right) \sim \chi^2 \text{ à } p \text{ ddl}$$

- Rapport de vraisemblance

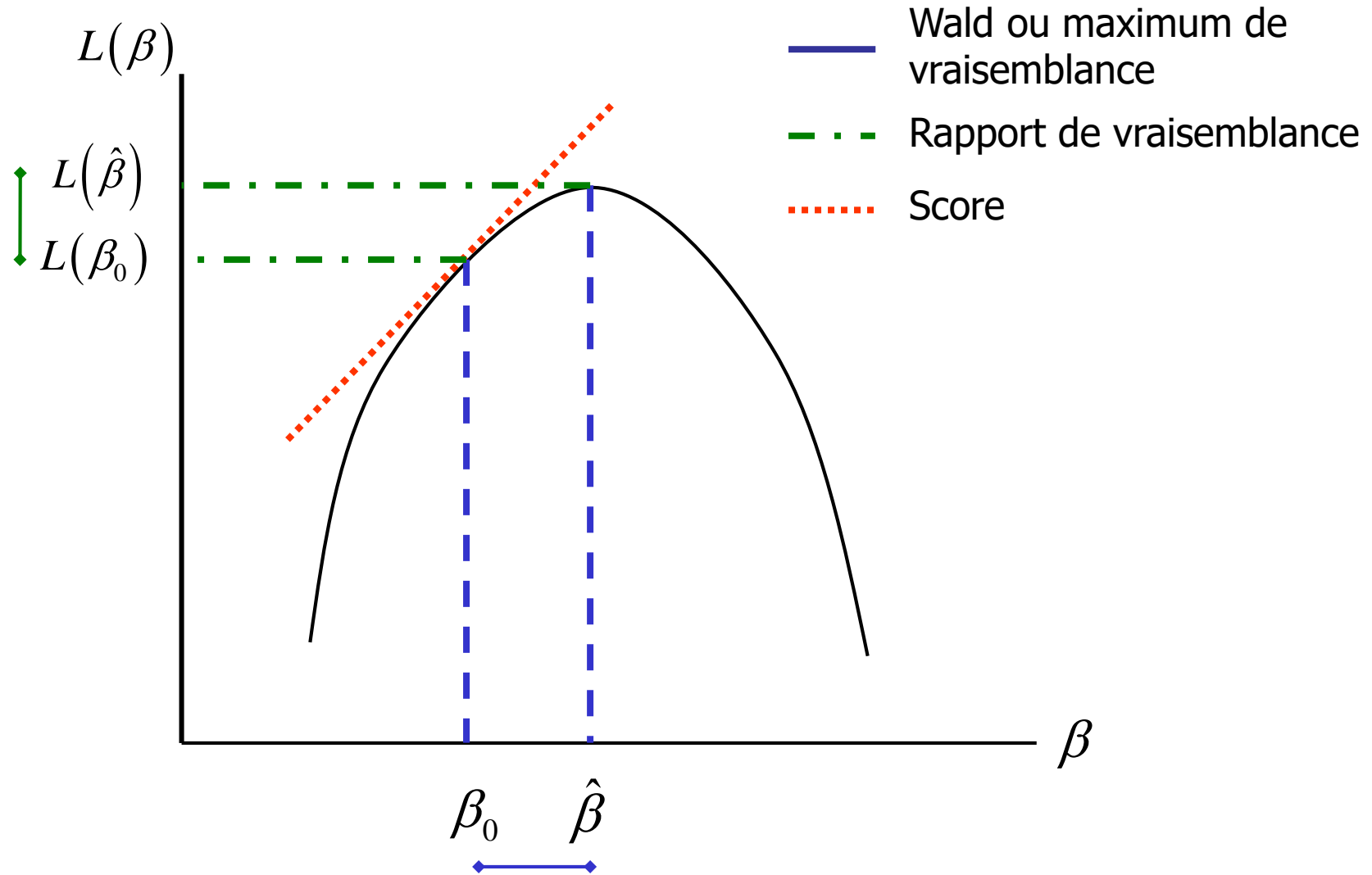
$$2 \left(L \left(\hat{\beta} \right) - L \left(\beta^{(0)} \right) \right) \sim \chi^2 \text{ à } p \text{ ddl}$$

- Score

$$U^T \left(\beta^{(0)} \right) I \left(\beta^{(0)} \right)^{-1} U \left(\beta^{(0)} \right) \sim \chi^2 \text{ à } p \text{ ddl}$$

où I est la matrice d'information de Fisher

Représentation Graphique des Tests



Exemple (Peto, 1979) : Données

Temps	Traitement	Fonction rénale	Temps participation	Traitement	Fonction rénale
8	A	AN	220	A	N
8	A	N	365*	A	N
13	B	AN	632	B	N
18	B	AN	700	B	N
23	B	AN	852*	A	N
52	A	AN	1296	B	N
63	A	AN	1296*	A	N
63	A	AN	1328*	A	N
70	B	N	1460*	A	N
76	B	N	1976*	A	N
180	B	N	1990*	B	N
195	B	N	2240*	B	N
210	B	N			

* Censures ; N : normale ; AN : anormale

Exemple (Peto, 1979) : Traitement

Modèle : $\lambda(t, \mathbf{z}) = \lambda_0(t) \exp(\beta_1 z_1)$ avec $z_1 = 1$ si traitement B

	coef	exp(coef)	se(coef)	z	p
TRT	0.561	1.75	0.51	1.1	0.27

	exp(coef)	exp(-coef)	lower .95	upper .95
TRT	1.75	0.571	0.645	4.76

Log Vraisemblance initiale = -46.9

Log Vraisemblance finale = -46.3

Likelihood ratio test = 1.26 on 1 df, p=0.261

Wald test = 1.21 on 1 df, p=0.271

Score (logrank) test = 1.24 on 1 df, p=0.265

→ $2(-46.3 + 46.9)$

→ $0.561^2 / 0.51^2$

Exemple (Peto, 1979) : Fonction Rénale

Modèle : $\lambda(t, \mathbf{z}) = \lambda_0(t) \exp(\beta_2 z_2)$ avec $z_2=1$ si fonction rénale AN

	coef	exp(coef)	se(coef)	z	p
FR	3.64	38.1	1.1	3.31	0.00094

	exp(coef)	exp(-coef)	lower .95	upper .95
FR	38.1	0.0262	4.41	330

Log Vraisemblance initiale = -46.9
Log Vraisemblance finale = -37.6

Likelihood ratio test= 18.7 on 1 df, p=0.0000152
Wald test = 10.9 on 1 df, p=0.000939
Score (logrank) test = 24.9 on 1 df, p=5.96e-007

Exemple (Peto, 1979) : Ajustement

Modèle : $\lambda(t, \mathbf{z}) = \lambda_0(t) \exp(\beta_1 z_1 + \beta_2 z_2)$

$z_1=1$ si traitement B et $z_2=1$ si fonction rénale AN

	coef	exp(coef)	se(coef)	z	p
TRT	1.22	3.4	0.598	2.05	0.04100
FR	4.28	72.1	1.191	3.59	0.00033

	exp(coef)	exp(-coef)	lower .95	upper .95
TRT	3.4	0.2945	1.05	11
FR	72.1	0.0139	6.98	745

Log Vraisemblance initiale = -46.9

Log Vraisemblance finale = -35.2

Likelihood ratio test = 23.5 on 2 df, p=7.86e-006

Wald test = 13.9 on 2 df, p=0.000982

Score (logrank) test = 28.3 on 2 df, p=7.32e-007

Codage des Variables Catégorielles (1)

$$\lambda(t, \mathbf{z}) = \lambda_0(t) \exp(\beta_1 z_1 + \beta_2 z_2 + \beta_3 z_3)$$

	Z1	Z2	Z3
Stade 1	0	0	0
Stade 2	1	0	0
Stade 3	0	1	0
Stade 4	0	0	1

$\lambda_0(t)$: taux de décès pour les patients en stade 1

$\lambda_1(t, \mathbf{z}) = \lambda_0(t) \exp(\beta_1 z_1)$ pour les patients en stade 2

$\lambda_2(t, \mathbf{z}) = \lambda_0(t) \exp(\beta_2 z_2)$ pour les patients en stade 3

$\lambda_3(t, \mathbf{z}) = \lambda_0(t) \exp(\beta_3 z_3)$ pour les patients en stade 4

$\exp(\beta_i z_i)$: taux relatif de décès pour les patients en stade i
par rapport aux patients stade 1

Codage des Variables Catégorielles (2)

$$\lambda(t, \mathbf{z}) = \lambda_0(t) \exp(\beta_1 z_1 + \beta_2 z_2 + \beta_3 z_3)$$

	Z1	Z2	Z3
Stade 1	0	0	0
Stade 2	1	0	0
Stade 3	1	1	0
Stade 4	1	1	1

$\lambda_0(t)$: taux de décès pour les patients en stade 1

$\lambda_1(t, \mathbf{z}) = \lambda_0(t) \exp(\beta_1 z_1)$ en stade 2

$\lambda_2(t, \mathbf{z}) = \lambda_0(t) \exp(\beta_1 z_1 + \beta_2 z_2)$ en stade 3

$\lambda_3(t, \mathbf{z}) = \lambda_0(t) \exp(\beta_1 z_1 + \beta_2 z_2 + \beta_3 z_3)$ en stade 4

$\exp(\beta_i z_i)$: taux relatif de décès pour les patients
en stade $i+1$ par rapport aux patients stade i

Codage des Variables Quantitatives

- Exprimée dans son unité
 - Hypothèse de linéarité
 - Transformation de la variable telle que le taux de mortalité augmente linéairement avec $f(\text{variable})$
- Regroupement par classes
 - Perte d'information
 - Hypothèse des taux proportionnels
- L'étude des « résidus » permet d'aider dans ce choix

Modèle de Cox Stratifié (1)

$$\lambda(t, \mathbf{z}) = \lambda_{0k}(t) \exp(\boldsymbol{\beta} \mathbf{z})$$

- Stratification sur une variable catégorielle (le centre dans un essai thérapeutique, ...) à K classes
- Les strates répartissent les sujets dans des groupes disjoints
- Le taux de base est différent dans chaque strates
- Le taux relatif des autres covariables est identique dans chacune des strates

Modèle de Cox Stratifié (2)

- Vraisemblance du modèle stratifié

$$l(\beta) = \sum_{k=1}^K l_k(\beta)$$

- Le vecteur score devient

$$U(\beta) = \sum_{k=1}^K U_k(\beta)$$

- Matrice d'information de Fisher

$$I(\beta) = \sum_{k=1}^K I_k(\beta)$$

Modèle de Cox Stratifié (3)

- Avantages

- Produit un ajustement naturel pour une variable de confusion
- Ne repose plus sur l'hypothèse de proportionnalité puisque l'on a 1 modèle de Cox par strates

- Inconvénients

- Pas d'estimation directe de l'importance de l'effet de la strate
- La précision dans l'estimation des coefficients et la puissance de l'analyse diminue avec le nombre de strate

Modèle Ajusté - Modèle Stratifié

x : variable quantitative

z : variable qualitative à 3 classes ($z = z_1, z_2$)

Modèle Ajusté

$$\lambda(t, x, z) = \lambda(t, 0, 0) \exp(\beta x + \gamma z)$$

$$\lambda_1(t, x) = \lambda(t, 0, 0) \exp(\beta x)$$

$$\lambda_2(t, x) = \lambda(t, 0, 0) \exp(\beta x) \exp(\gamma_1)$$

$$\lambda_3(t, x) = \lambda(t, 0, 0) \exp(\beta x) \exp(\gamma_2)$$

$\exp(\gamma_i)$: taux relatifs du modèle
(Proportionnalité)

Modèle Stratifié

$$\lambda(t, x) = \lambda(t, 0) \exp(\beta x)$$

$$\lambda_1(t, x) = \lambda_1(t, 0) \exp(\beta x)$$

$$\lambda_2(t, x) = \lambda_2(t, 0) \exp(\beta x)$$

$$\lambda_3(t, x) = \lambda_3(t, 0) \exp(\beta x)$$

Un modèles de Cox dans
chaque strate

Exemple (Peto, 1979) : Ajusté - Stratifié

Modèle ajusté

```
      coef exp(coef) se(coef)      z      p
TRT  1.22      3.4    0.598  2.05 0.04100
FR   4.28     72.1    1.191  3.59 0.00033

      exp(coef) exp(-coef) lower .95 upper .95
TRT      3.4    0.2945     1.05     11
FR     72.1    0.0139     6.98    745

Log Vraisemblance initiale = -46.9
Log Vraisemblance finale   = -35.2

Likelihood ratio test= 23.5 on 2 df,  p=7.86e-006
Wald test              = 13.9 on 2 df,  p=0.000982
Score (logrank) test = 28.3 on 2 df,  p=7.32e-007
```

Modèle stratifié sur la FR

```
      coef exp(coef) se(coef)      z      p
TRT  1.46      4.32    0.66  2.22 0.026

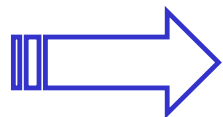
      exp(coef) exp(-coef) lower .95 upper .95
TRT      4.32    0.231     1.19    15.8

Log Vraisemblance initiale = -33.9
Log Vraisemblance finale   = -30.2

Likelihood ratio test= 6.07 on 1 df,  p=0.0137
Wald test              = 4.93 on 1 df,  p=0.0265
Score (logrank) test = 5.79 on 1 df,  p=0.0161
```


Résidus

- Permettent d'étudier la relation fonctionnelle entre une variable et les taux de décès
- On recherche une transformation de la variable z telle que la variable résultante – **résidus** – obéisse à la même loi pour tous les sujets
- Si z a un effet de la forme $f(z)$, alors ses résidus sont approximativement proportionnels à $f(z)$



Méthodes graphiques

Résidus Martingales

- Ils comparent, pour un sujet i , l'observation de l'événement (δ_i) à celle qui est attendue, conditionnellement au modèle testé, et mesurée par le taux cumulé de la population

$$res_i = \delta_i - \hat{\Lambda}(t_i)$$

- Évaluation de la forme : représentation du nuage de points ($z_i ; res_i$) et superposition de $f(z)$

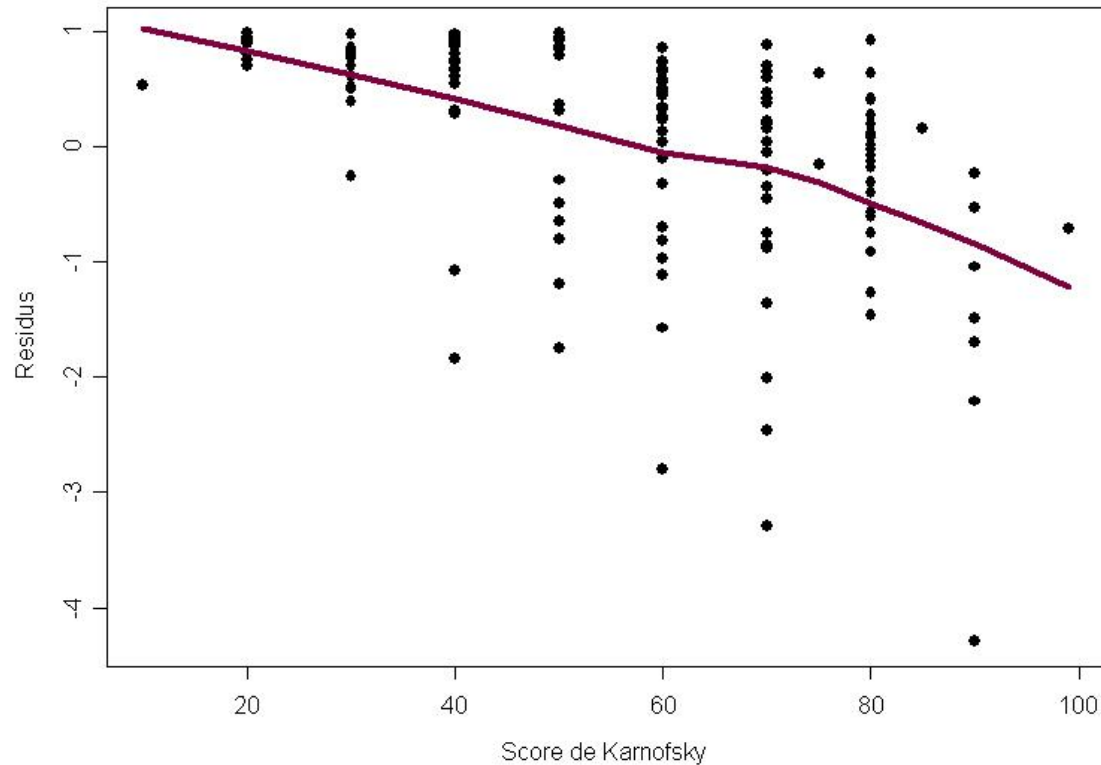
Exemple : Cancer du Poumon (1)

Données : Veteran's Administration lung cancer trial, Kalbfleisch-Prentice, 1980

- Traitement : standard ou test
- Score pronostic de Karnofsky : 0-100 (- ; +)
- Age : années
- Traitement antérieur : oui – non
- TDiagnostic : temps entre le diagnostic et l'entrée dans l'étude
- Type de cellules : 4 types différents

Exemple : Cancer du Poumon (2)

- Résidus martingales



Résidus de Schoenfeld

- Correspondent à la contribution au score de chaque décès
- Une covariable, pas d'ex-aequo : contribution du $i^{\text{ème}}$ décès (si ex-aequo, somme des résidus)

$$z_i - \bar{z}_i(\hat{\beta})$$

- Plusieurs covariables : chaque covariable fournit son résidu

$$z_{ki} - \bar{z}_{ki}(\hat{\beta})$$

Résidus Standardisés de Schoenfeld

- Résidus de Schoenfeld divisés par leur variance
- Permettent de vérifier l'hypothèse de proportionnalité des taux de décès
 - Test
 - Représentation graphique

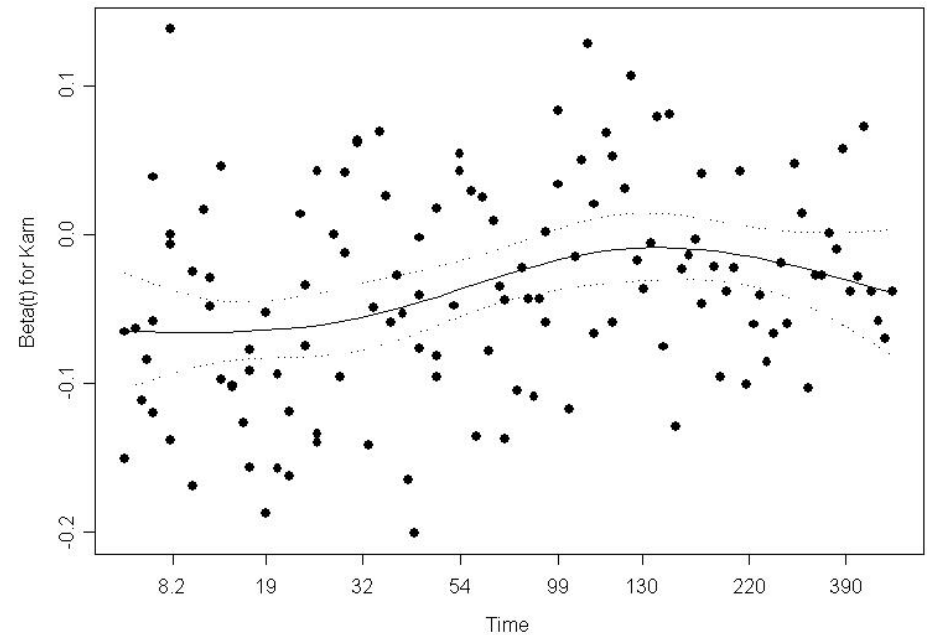
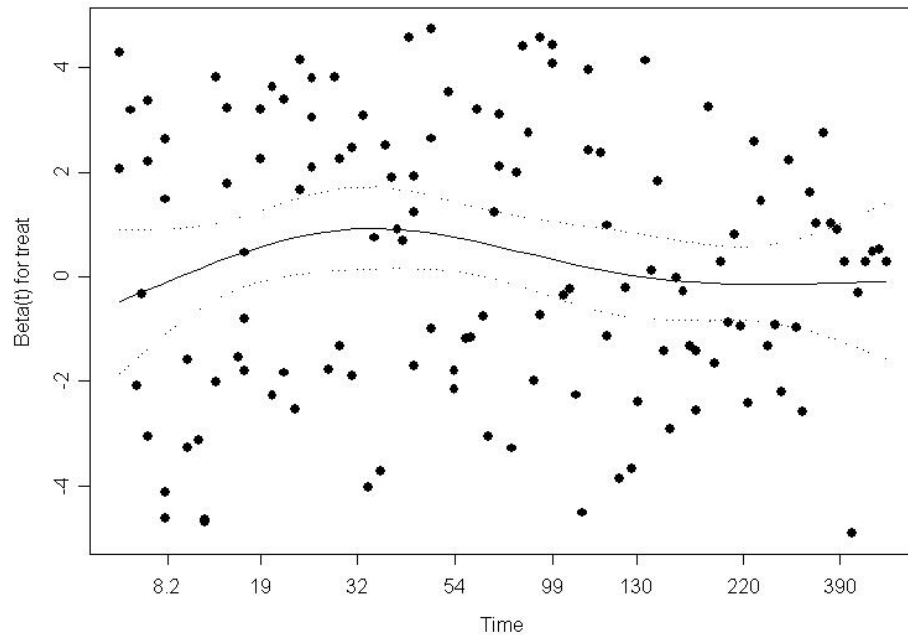
Exemple : Cancer du Poumon (3)

- Test des résidus standardisés de Schoenfeld

	rho	chisq	p
Traitement	-0.0607	0.545	0.46024
Age	0.1734	4.634	0.03134
Karnofsky	0.2568	9.146	0.00249
Diag.time	0.1542	2.891	0.08909
priorTRT	-0.1574	3.476	0.06226
GLOBAL	NA	13.488	0.01921

Exemple : Cancer du Poumon (4)

- Représentation graphique des résidus standardisés de Schoenfeld



Taux Relatifs Dépendants du Temps (1)

- Le modèle de Cox impose que l'effet estimé soit le même au cours du temps (proportionnalité) ; pas toujours vérifié

- Dans ce cas

$$\lambda(t, \mathbf{z}) = \lambda_0(t) \exp(\boldsymbol{\beta} \mathbf{z})$$

- Devient

$$\lambda(t, \mathbf{z}) = \lambda_0(t) \exp\left(\sum_{i=1}^p \beta_i(t) z_i\right)$$

Taux Relatifs Dépendants du Temps (2)

Diagnostic

- Méthode graphique : $\log(-\log(S))$
- Résidus de Schoenfeld

Taux Relatifs Dépendants du Temps (3)

Prise en compte

- Stratifier sur la variable
 - + Simple
 - Pas de test, variables qualitatives seulement, puissance
- Partitionner le temps (proportionnel par morceaux)
- Estimer les paramètres du modèles avec des fonctions dépendantes du temps

Références

- Cox DR. Regression models and life tables. *Journal of the Royal Statistical Society, Series B* 1972; 34: 187-220.
- Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. Wiley, New York, 1980.
- Therneau T, Grambsch PM. Modeling survival data: Extending the Cox model. New York: Springer-Verlag 2000.