Biostatistique
Analyse de la survie

Pr Pascal ROY

Remerciements : Pr Jacques ESTEVE

1. Introduction - Analyse de la survie

2. Méthode de Kaplan et Meier - Test du Log-Rank

3. Taux de Mortalité

4. Survie paramétrique / Survie exponentielle par intervalle

5. Modèle à Taux proportionnel

Dans ces diapos, « log » est le logarithme népérien.

1. Introduction

- Contexte Médical
  - Quantifier la *probabilité de décès*, rechute...
  - Evaluer les *facteurs pronostiques*
  - Comparer des *traitements*

....

Le risque (ou probabilité) de décès d'un individu est de toute façon égal à un !!!

La probabilité de décéder estimée sur une cohorte d'individus n'a de sens qu'à délai fixé.

La variable d'intérêt est la variable aléatoire T, que constitue la durée qui sépare le diagnostic de la mort (ou de la rechute), variable dont on étudie la distribution.

Contexte

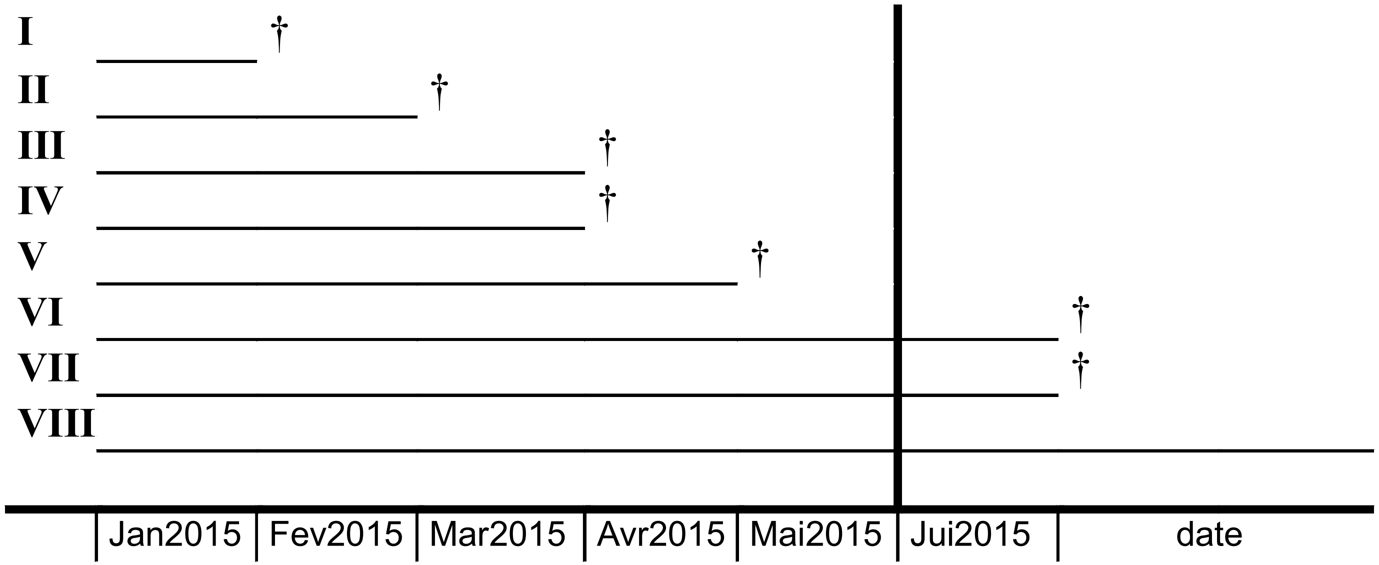
- Statistique
  - mesure du *temps* écoulé entre deux *événements*
  - distribution *non gaussienne*
  - utilisation de méthodes *non-paramétriques*
  - Prise en compte de données incomplètes: données *censurées* et *tronquées*

Temps entre événements

Origine	Evénement
Naissance	Maladie,décès
Infection HIV	SIDA
Diagnostic,rando-misation,traitement	*Décès* (ou décès *de la maladie*)
Entrée à l'hopital	Infection nosocomiale
Rémission	Rechute
Mise en service	Panne

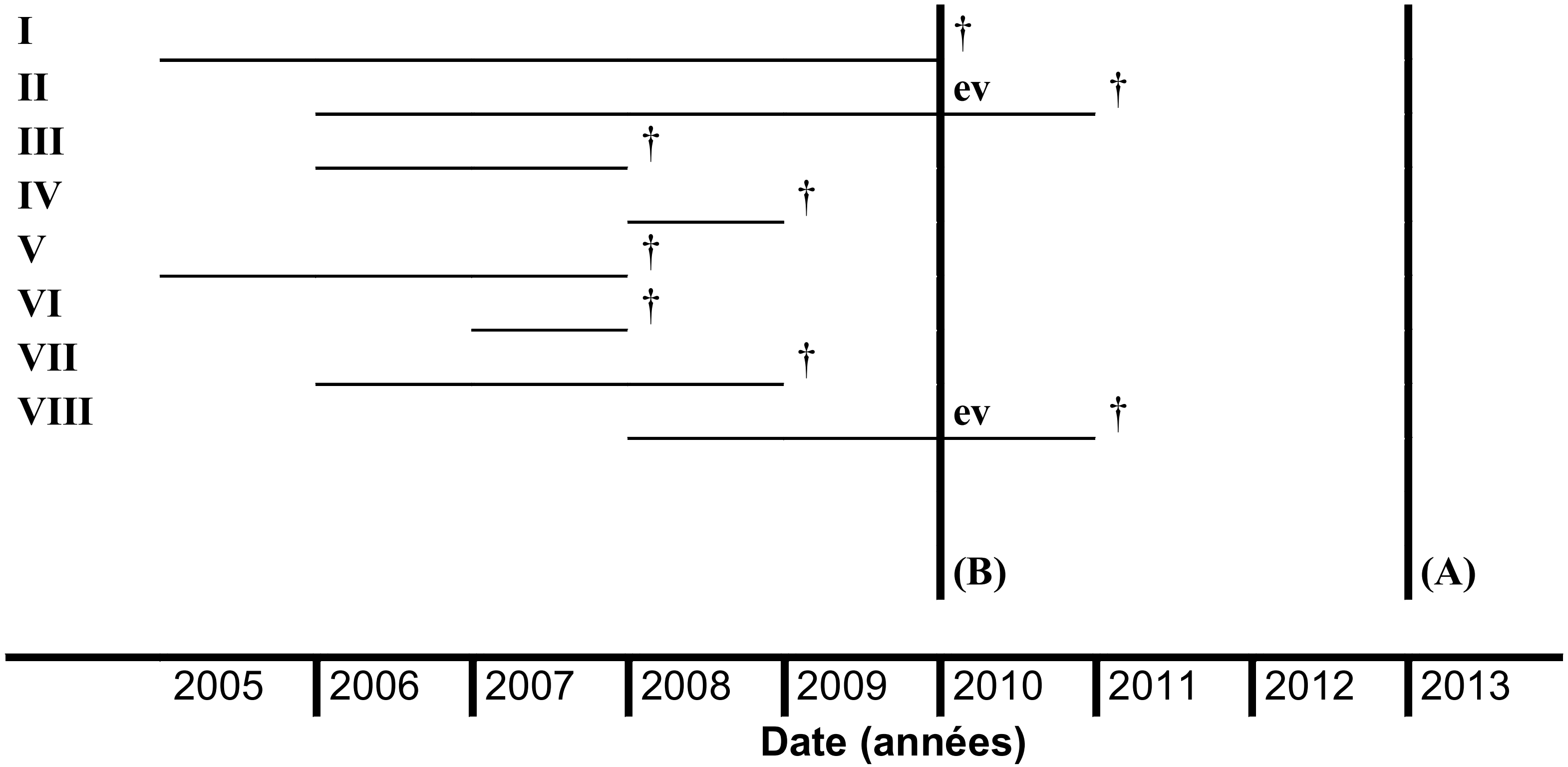
Censure

- L ’observation est censurée (à droite) si on sait seulement que $T > t$, *date* à la quelle l ’observation s ’est achevée.
  - l ’évènement est le *décès* et le sujet est en vie à la fin de l ’étude
  - l ’événement est le *décès par cancer* et l ’observation s’achève par un *accident*



Données fictives

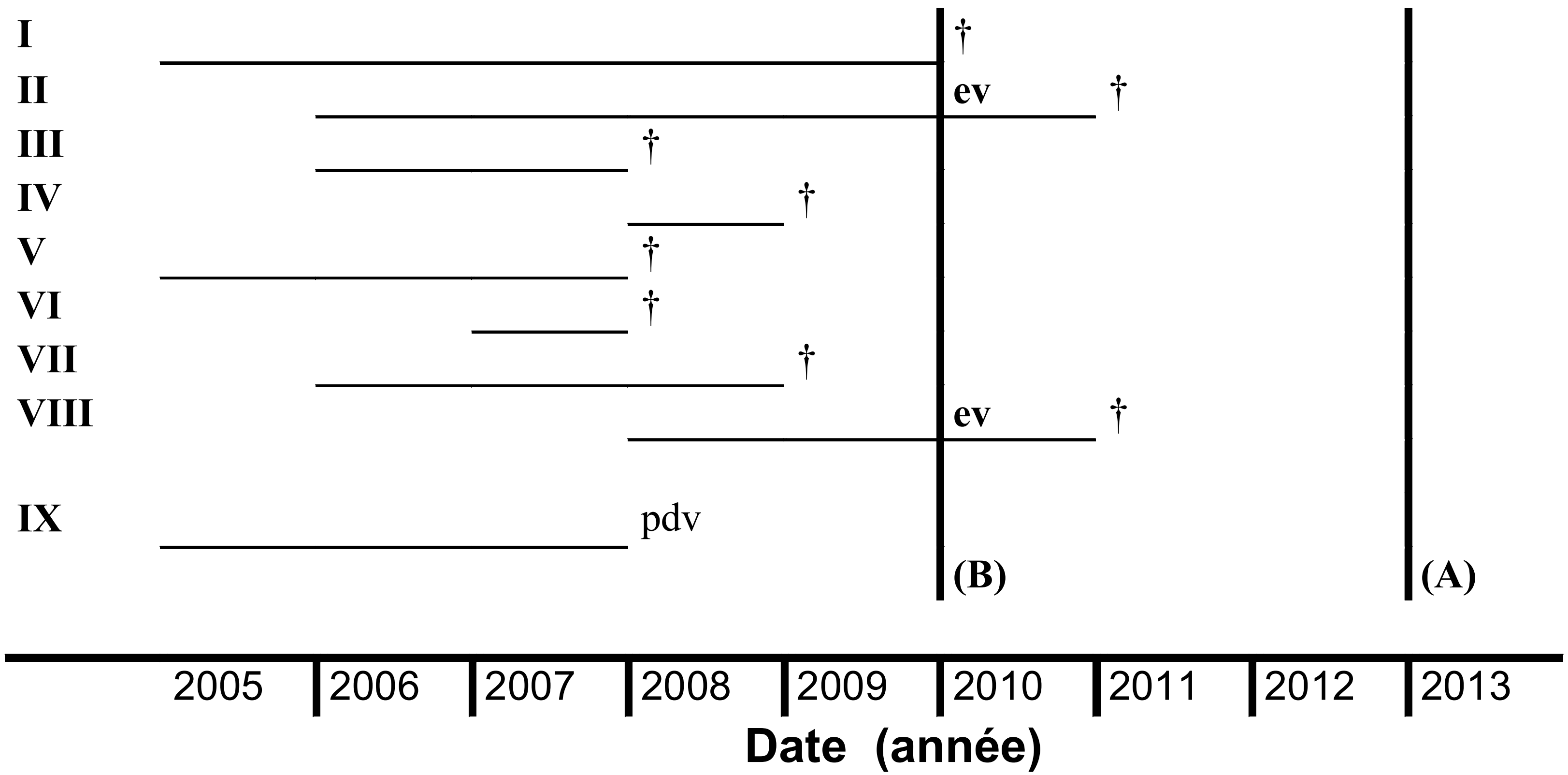
Date d’origine commune à tous les sujets



Données fictives

Dates d'origine échelonnées au cours du temps.

(A) et (B) sont deux dates possibles de fin de suivi

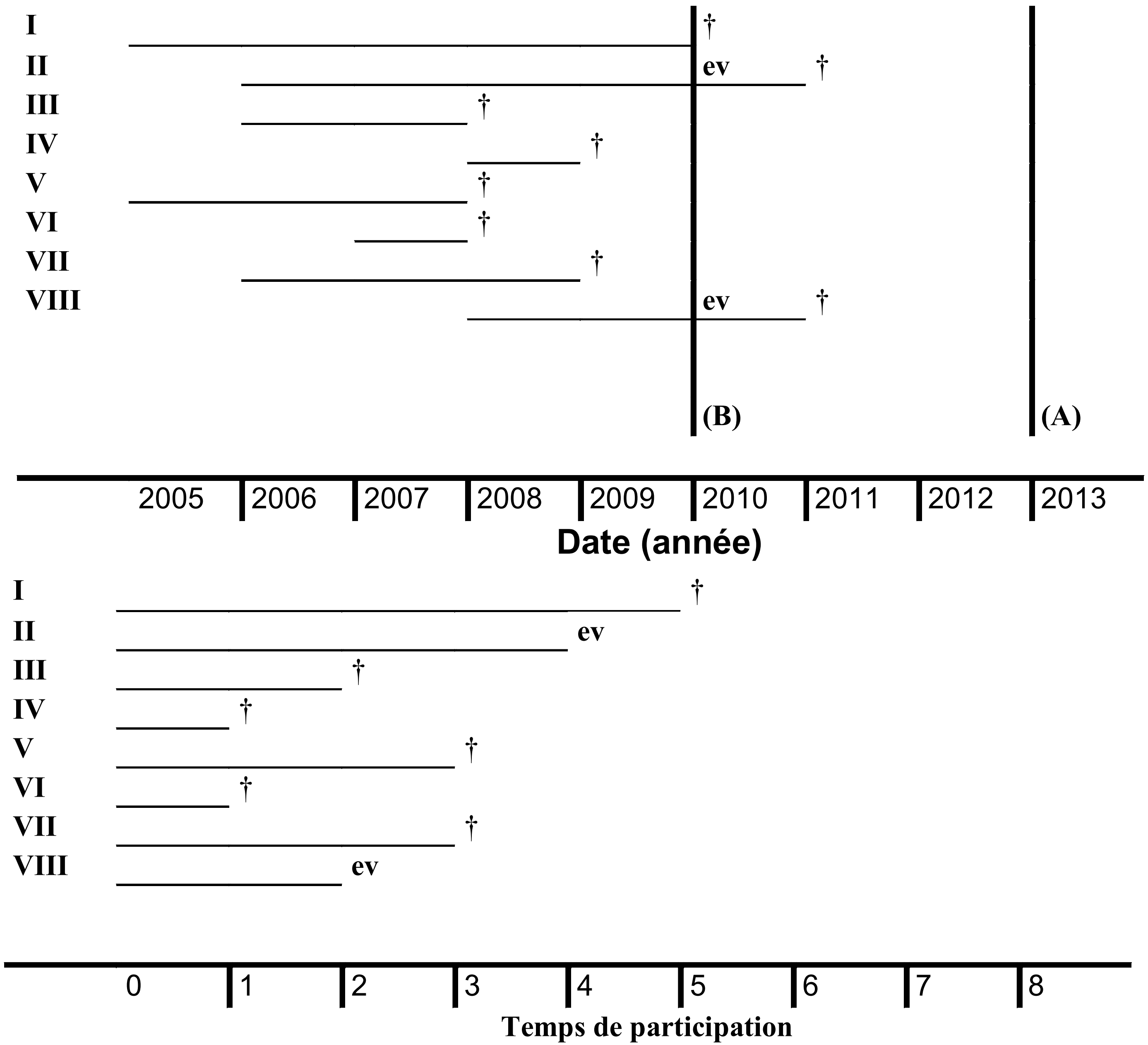


ev : exclu vivant

Données fictives

Dates d'origine échelonnées au cours du temps.

(A) et (B) sont deux dates possibles de fin de suivi


I
II
III
IV
V
VI
VII
VIII
ev
ev
(B)
(A)
2005
2006
2007
2008
2009
2010
2011
2012
2013
Date (année)
I
II
III
IV
V
VI
VII
VIII
ev
ev
0
1
2
3
4
5
6
7
8
Temps de participation

En résumé, certains sujets ne sont pas suivis jusqu'à ce que l'événement se produise, pour eux on sait seulement que $T > t_i$. L'observation correspondante est une observation *censurée.* Les données de survie sont donc constituées de deux informations, la durée du suivi T, et l'indicateur d'événement δ ; $\delta = 1$ si l 'observation se termine par un *décès,* $\delta = 0$ si le sujet est *vivant* à la fin des dernières nouvelles.

Définitions

- **Risque de décès** = *probabilité* d 'être décédé à la date t = *fonction de répartition* de la variable T = $R(t)=Pr(T \leq t)$
- **Survie =** *probabilité* d 'être en vie à la date t =$S(t)=Pr(T > t) = 1\text{-}R(t)^{*}$

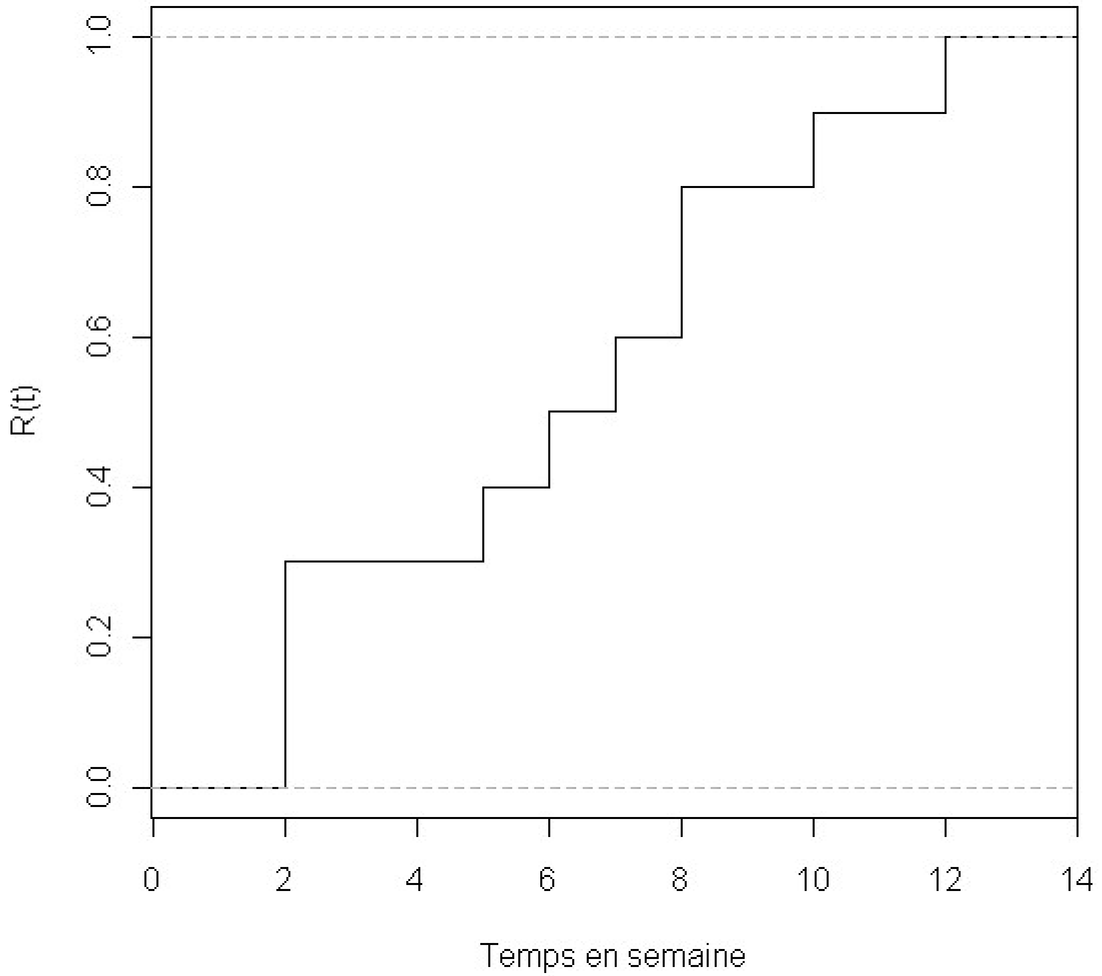
* pour une distribution continue

Rappel
Distribution cumulée

Durées de survie

obs:	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_{10}
	2	2	8	2	6	8	5	10	12	7
ord:	$t_{(1)}$	$t_{(2)}$	$t_{(3)}$	$t_{(4)}$	$t_{(5)}$	$t_{(6)}$	$t_{(7)}$	$t_{(8)}$	$t_{(9)}$	$t_{(10)}$
	2	2	2	5	6	7	8	8	10	12
événmt:			≤ 2	≤ 5	≤ 6	≤ 7		≤ 8	≤ 10	≤ 12
freq.			0.3	0.4	0.5	0.6		0.8	0.9	1

Fonction de répartition empirique



Données complètes*

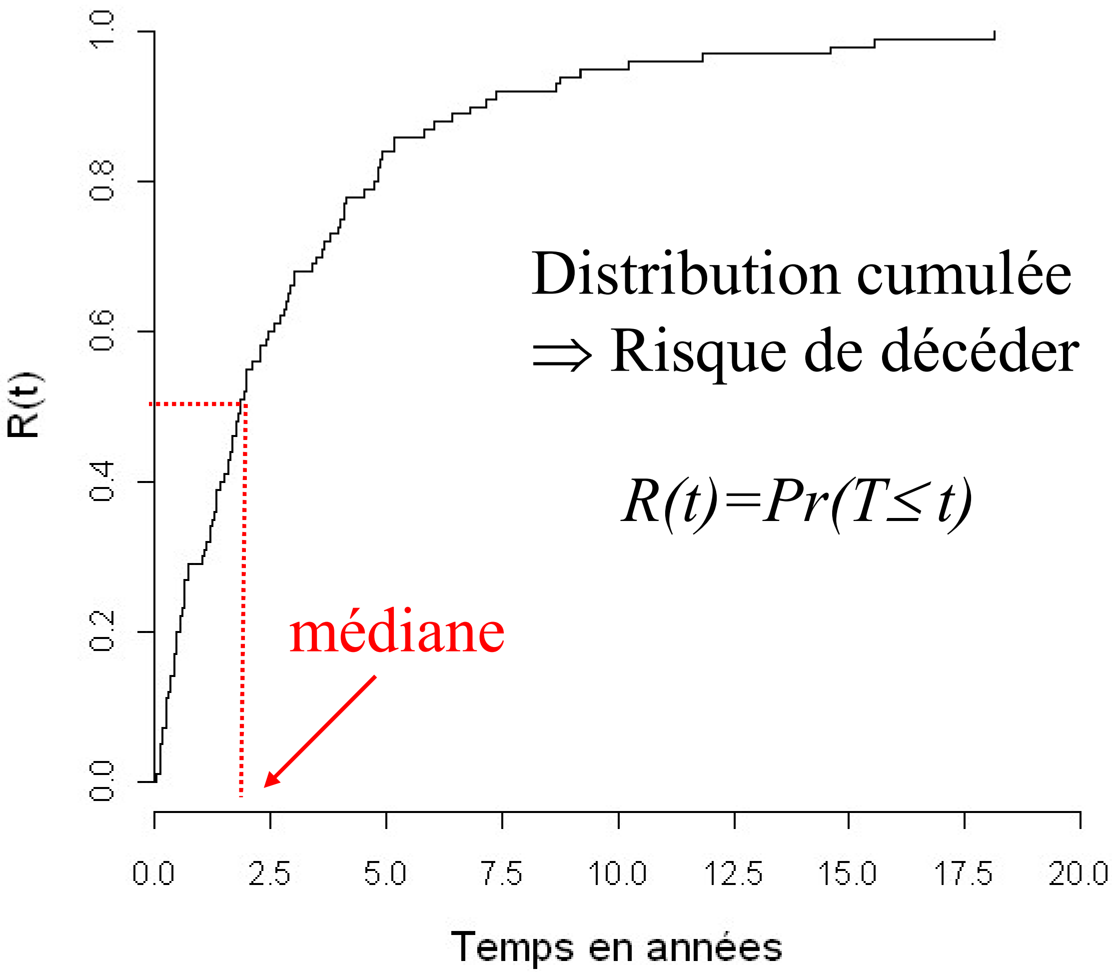
- Si $t_{(i)}$ est le *ième* temps de survie de l 'échantillon
- Le ***Risque*** est estimé par:
  - $t_{(i)} \longrightarrow i/n_0$
- La ***Survie*** est estimée par:
  - $t_{(i)} \longrightarrow (n_0 - i)/n_0 = n_i/n_0 = 1 - i / n_0$

*et sans ex-aequo

Exemple 1: temps de survie en années
(données complètes)

1.22	1.86	1.35	3.81	0.47	1.63	2.47	0.42	0.24	0.45
2.83	5.17	18.14	4.92	0.30	2.87	9.21	1.32	2.09	3.48
1.58	5.16	2.43	0.05	3.00	0.24	4.01	0.34	1.59	7.39
0.42	1.08	0.63	0.56	1.79	0.63	2.60	0.11	1.43	3.97
1.22	7.18	3.66	0.10	2.29	3.40	4.13	2.78	4.82	4.87
5.81	1.25	4.08	6.81	6.02	0.15	1.66	3.60	0.18	0.12
2.72	15.56	0.72	1.35	0.72	8.68	1.96	1.97	4.75	0.62
2.94	2.27	4.53	10.22	1.03	1.84	4.81	1.66	0.34	1.78
0.55	0.25	3.00	0.25	0.48	8.74	0.64	6.41	14.59	1.93
0.10	4.08	11.82	1.78	1.51	1.96	0.41	1.12	1.28	0.60

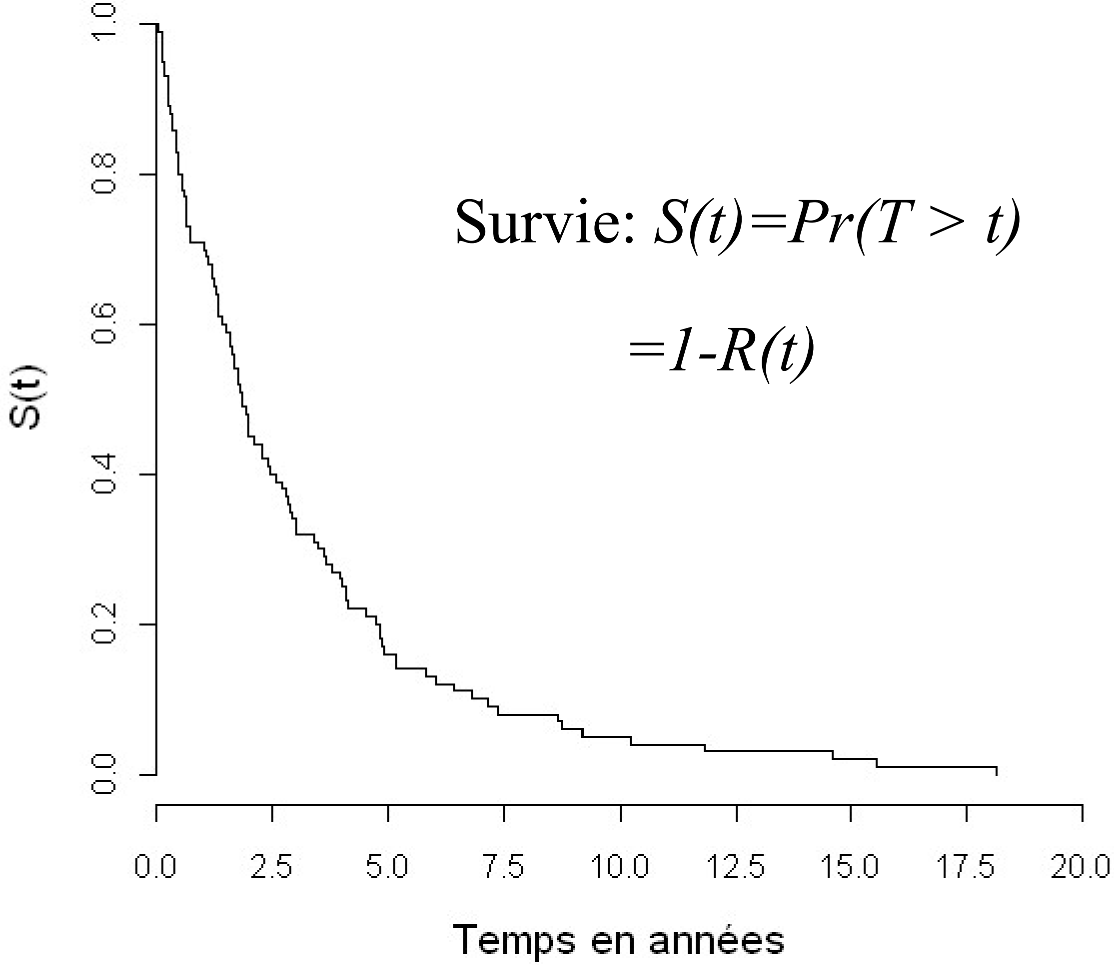
moyenne: 2.994 ans déviation standard: 3.363 ans



Distribution cumulée

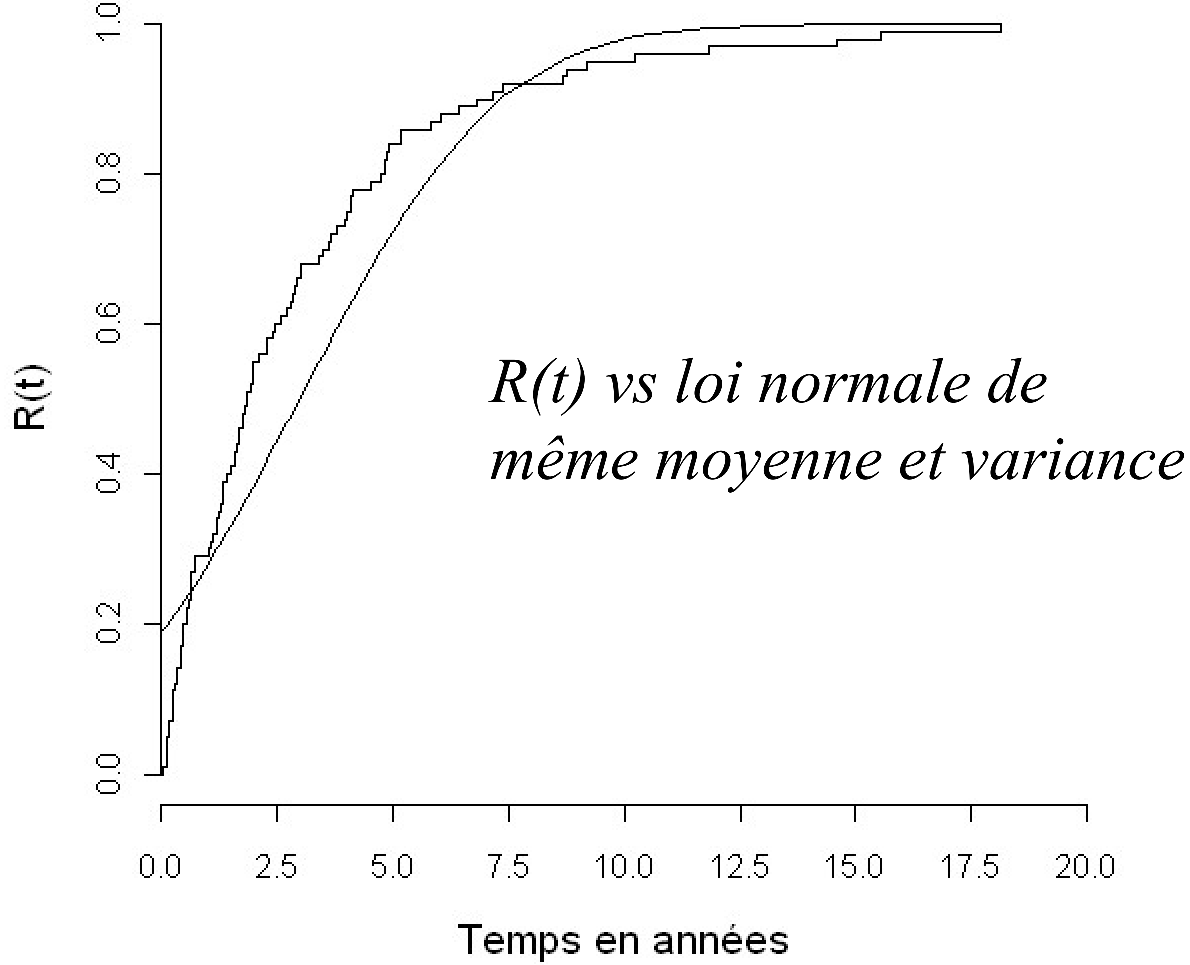
$\Rightarrow$ Risque de décéder

$$R(t)=Pr(T \leq t)$$



Survie: $S(t)=Pr(T > t)$

$=1-R(t)$


R(t)
1.0
0.8
0.6
0.4
0.2
0.0
0.0
2.5
5.0
7.5
10.0
12.5
15.0
17.5
20.0
Temps en années
R(t) vs loi normale de
même moyenne et variance

Données incomplètes

- Certains sujets ne sont pas suivis jusqu 'à ce que l 'événement se produise
  - pour eux on sait seulement que $T > t_i$
  - t_i est une observation *censurée*
- Les données censurées sont résumées par
  - ***T*** durée du suivi, δ indicateur de l 'événement
    - $\delta = 1$ si l 'observation se termine par un ***décès***
    - $\delta = 0$ si le sujet est ***vivant*** à la fin du suivi

1. Introduction - Analyse de la survie

2. Méthode de Kaplan et Meier - Test du Log-Rank

3. Taux de Mortalité

4. Survie paramétrique / Survie exponentielle par intervalle

5. Modèle à Taux proportionnel

2. Méthode de Kaplan-Meier

- Données complètes (*observation de i décès)*
  - $S(t_i) = (1- i / n_0)=$

$$\left(1-\frac{1}{n_0}\right)\times\left(1-\frac{1}{n_0-1}\right)\times\left(1-\frac{1}{n_0-2}\right)\times....\left(1-\frac{1}{n_0-i+1}\right)$$

  - s_1 s_2 s_3 …….. s_i
- Données incomplètes ($\Sigma\delta_i$ *décès*)

$$\left(1-\frac{\delta_1}{n_0}\right)\times\left(1-\frac{\delta_2}{n_0-1}\right)\times\left(1-\frac{\delta_3}{n_0-2}\right)\times....\left(1-\frac{\delta_i}{n_0-i+1}\right)$$

Kaplan-Meier (exemple)

- **Données complètes (*observation de 50 décès*)**
  - $S(t_{20}) = (1- 20 / 50)= 30 / 50$

$$\left(\frac{49}{50}\right)\times\left(\frac{48}{49}\right)\times\left(\frac{47}{48}\right)\times....\left(\frac{30}{31}\right)$$

- **Données incomplètes ($\Sigma\delta_i$ *décès*)**

$$\left(\frac{49}{50}\right)\times(1)\times\left(\frac{47}{48}\right)\times....\left(\frac{30}{31}\right)$$

Kaplan-Meier (exemple)

- **Données complètes (*observation de 50 décès*)**
  - $S(t_{20}) = (1 - 20/50) = 30/50$

$$\left(\frac{\cancel{49}}{50}\right)\times\left(\frac{48}{\cancel{49}}\right)\times\left(\frac{47}{48}\right)\times\ldots.\left(\frac{30}{31}\right)$$

- **Données incomplètes ($\Sigma\delta_i$ *décès*)**

$$\left(\frac{49}{50}\right)\times(1)\times\left(\frac{47}{48}\right)\times\ldots.\left(\frac{30}{31}\right)$$

Kaplan-Meier (exemple)

- **Données complètes (*observation de 50 décès)***
  - $S(t_{20}) = (1- 20 / 50) = 30 / 50$

$$\left(\frac{\cancel{49}}{50}\right)\times\left(\frac{\cancel{48}}{\cancel{49}}\right)\times\left(\frac{47}{\cancel{48}}\right)\times\ldots.\left(\frac{30}{31}\right)$$

- **Données incomplètes ($\Sigma\delta_i$ *décès*)**

$$\left(\frac{49}{50}\right)\times(1)\times\left(\frac{47}{48}\right)\times\ldots.\left(\frac{30}{31}\right)$$

Kaplan-Meier (exemple)

- **Données complètes (*observation de 50 décès)***
  - *$S(t_{20}) = (1 - 20 / 50) = 30 / 50$*

$$\left(\frac{\cancel{49}}{50}\right) \times \left(\frac{\cancel{48}}{\cancel{49}}\right) \times \left(\frac{\cancel{47}}{\cancel{48}}\right) \times \ldots . \left(\frac{30}{31}\right)$$

- **Données incomplètes ($\Sigma\delta_i$ *décès*)**

$$\left(\frac{49}{50}\right) \times (1) \times \left(\frac{47}{48}\right) \times \ldots . \left(\frac{30}{31}\right)$$

Kaplan-Meier (exemple)

- **Données complètes (*observation de 50 décès)***
  - $S(t_{20}) = (1 - 20/50) = 30/50$

$$\left(\frac{\cancel{49}}{50}\right)\times\left(\frac{\cancel{48}}{\cancel{49}}\right)\times\left(\frac{\cancel{47}}{\cancel{48}}\right)\times....\left(\frac{30}{\cancel{31}}\right)$$

- **Données incomplètes (*$\Sigma\delta_i$ décès*)**

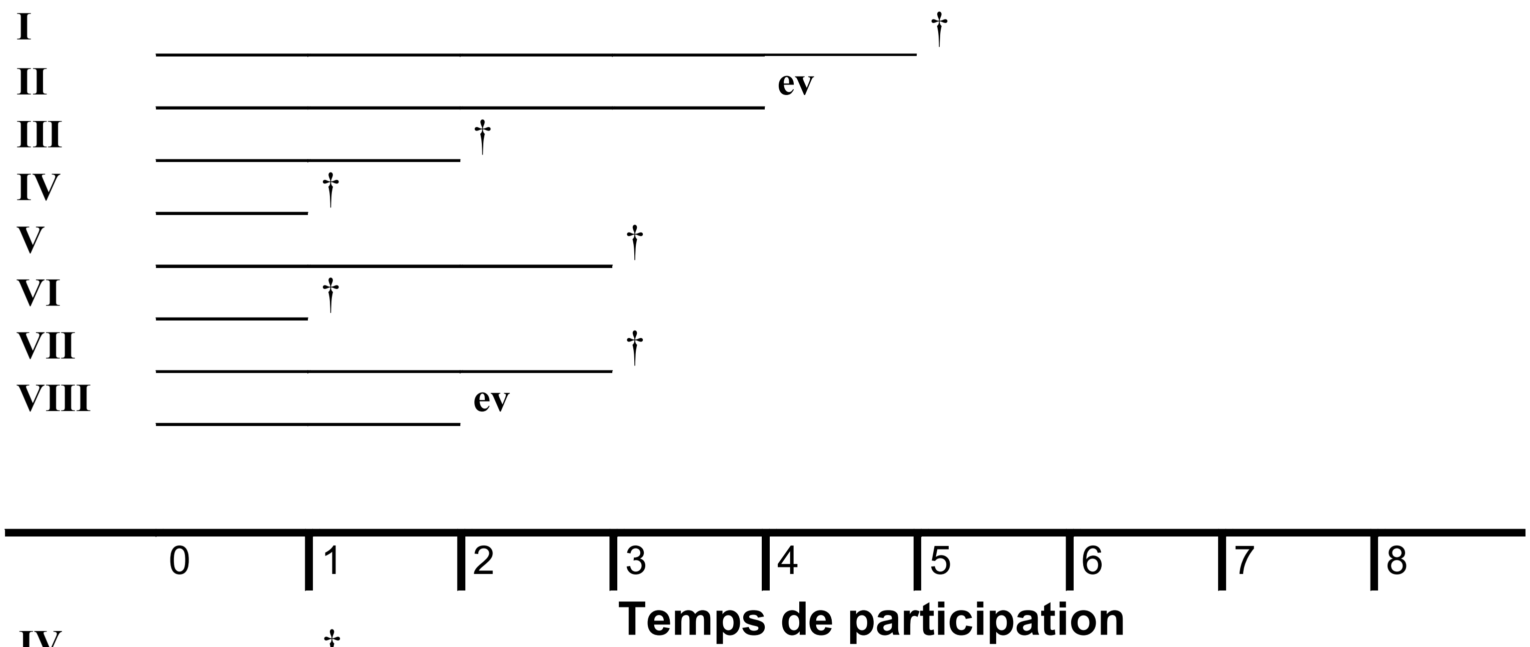
$$\left(\frac{49}{50}\right)\times(1)\times\left(\frac{47}{48}\right)\times....\left(\frac{30}{31}\right)$$

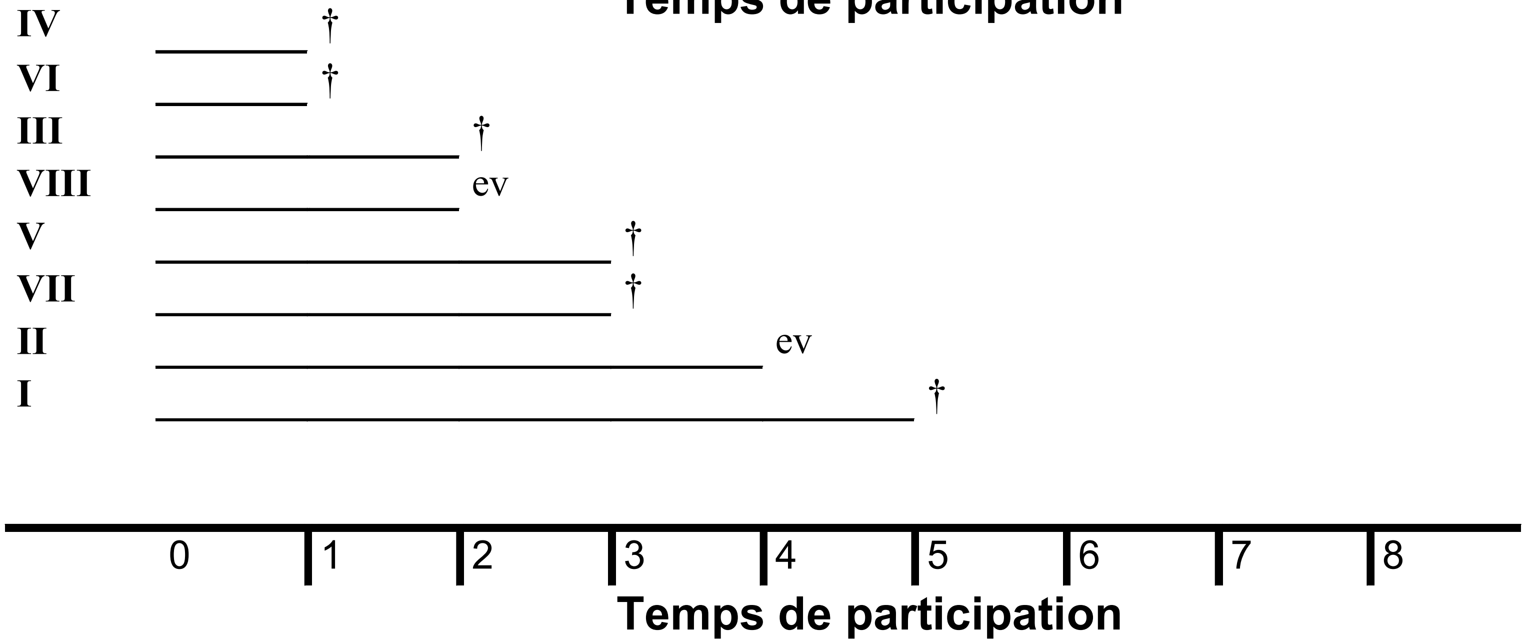
Erreur type de la survie en t_i

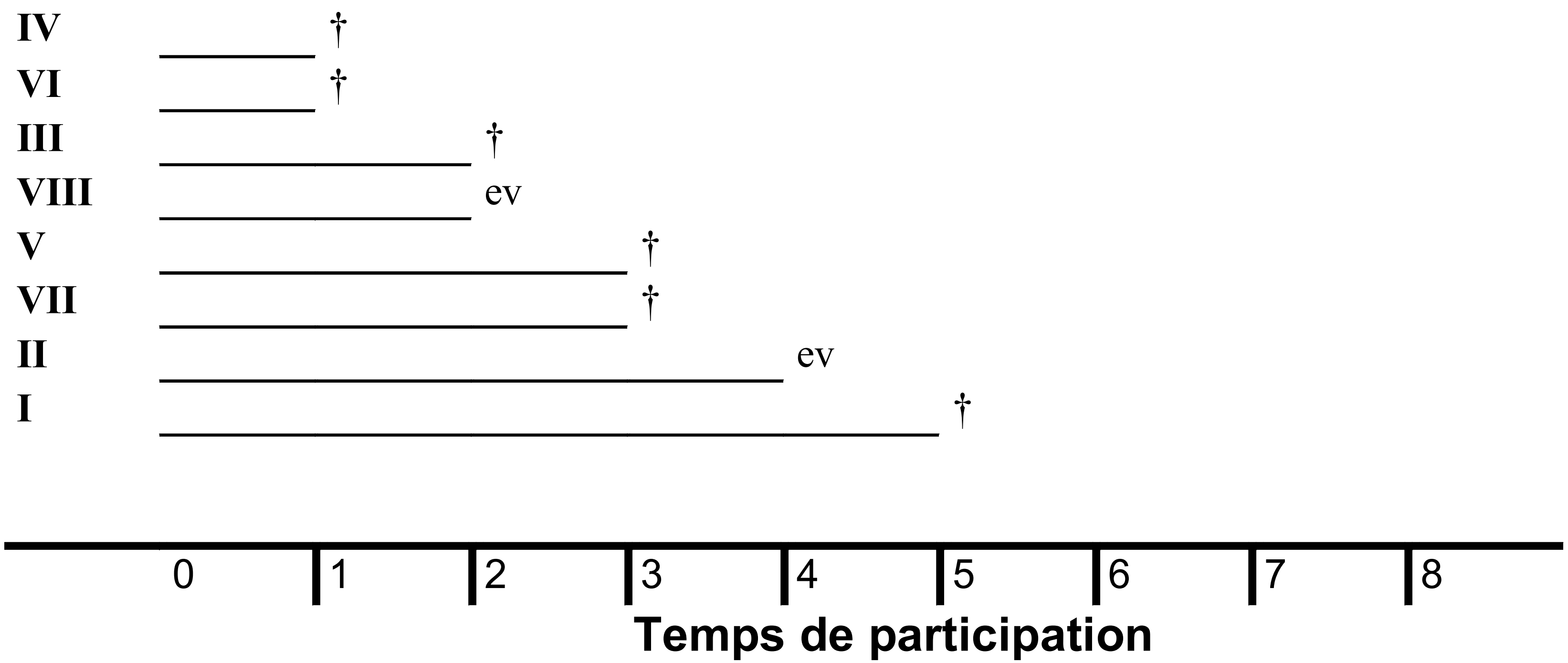
- **Si les données sont complètes (pas de censure) l 'erreur type est celle d'une proportion**

$$\hat{S}_i = n_i / n_0 \qquad SE[\hat{S}_i] = \sqrt{\frac{\hat{S}_i (1-\hat{S}_i)}{n_0}}$$

- **En cas de censure l'erreur type est**
  - plus élevée que celle-ci
    - Variance de Greenwood

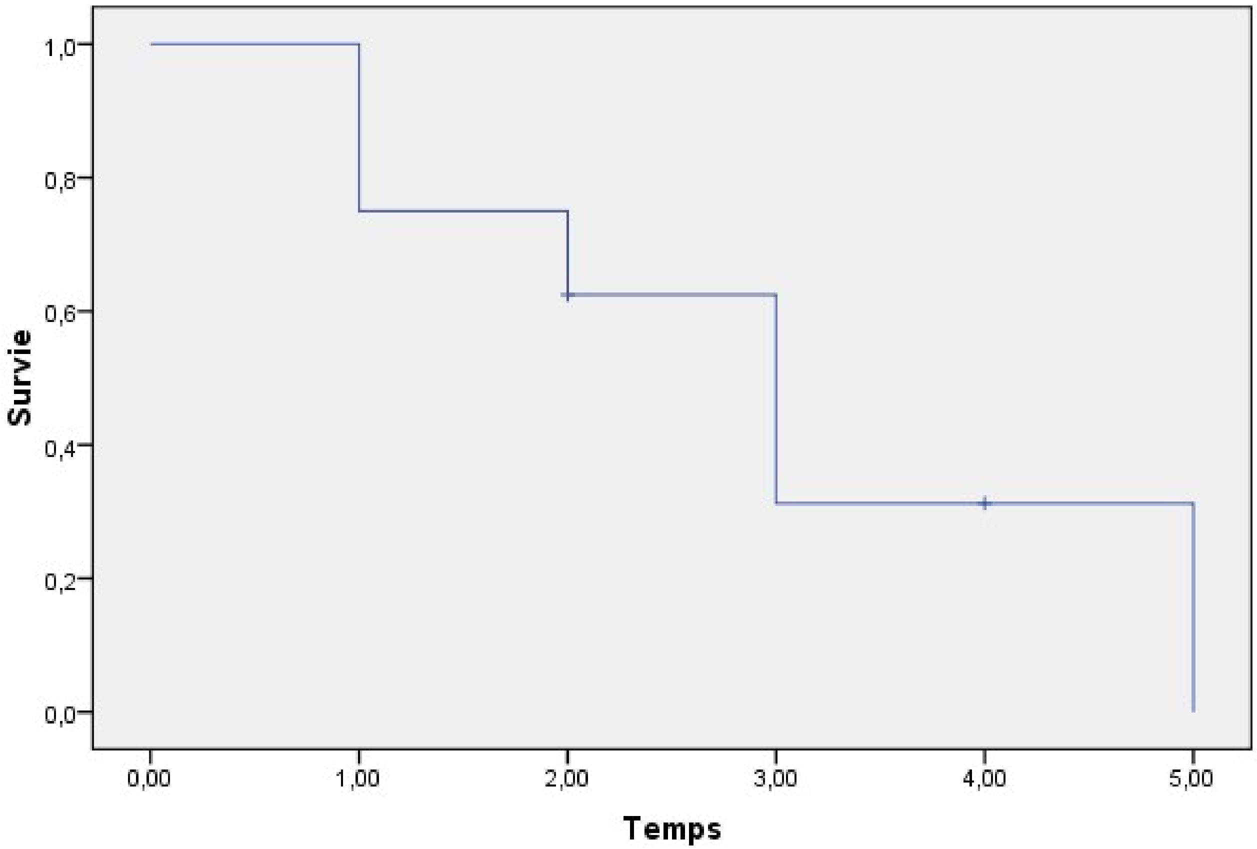

I †
II ev
III †
IV †
V †
VI †
VII †
VIII ev
0 1 2 3 4 5 6 7 8
Temps de participation


IV †
VI †
III †
VIII ev
V †
VII †
II ev
I †
0 1 2 3 4 5 6 7 8
Temps de participation



t_i	N_i	ev	d_i	s_i	S_i
0	8				1
1	8	0	2	6/8	6/8
2	6	0	1	5/6	5/8
2	5	1	0	1	5/8
3	4	0	2	2/4	5/16
4	2	1	0	1	5/16
5	1	0	1	0	0

t_i	N_i	ev	d_i	s_i	S_i
0	8				1
1	8	0	2	6/8	6/8
2	6	0	1	5/6	5/8
2	5	1	0	1	5/8
3	4	0	2	2/4	5/16
4	2	1	0	1	5/16
5	1	0	1	0	0



Comparaison de la survie dans deux groupes

- Comparer quoi?
  - la survie moyenne
  - la survie médiane
  - la probabilité de survie à un "délai" donné
- La troisième option est banale: utiliser le test classique:

$$\chi = \frac{\hat{S}_{1i} - \hat{S}_{2i}}{\sqrt{Var(\hat{S}_{1i}) + Var(\hat{S}_{2i})}}$$

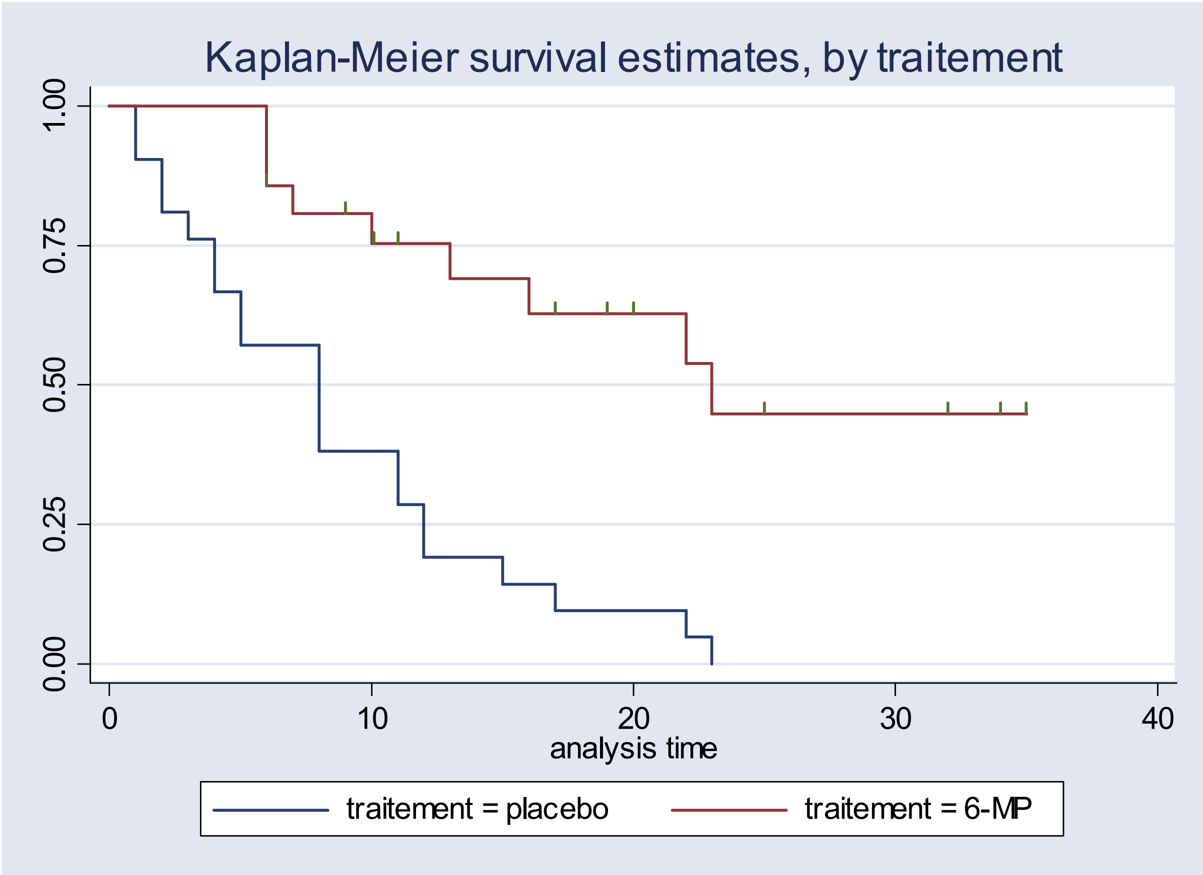
Comparaison de la survie dans deux groupes

- L'option précédente ne compare que la situation dans les deux groupes ***à un délai donné***
- On souhaite comparer les **distributions** des durées de survie
- La spécificité des données de survie suggère l'utilisation de ***méthodes non-paramétriques***

Exemple 2, Anderson data

Une première illustration de ce qui précède est basé sur des données histroriques très utilisées pour illustrer la méthodologie.
Elle concerne la durée de rémission de la leucémie aigue obtenue par traitement stéroide si on traite en outre avec la 6-mercaptopurine
Ref:Freireich *et al*. *Blood* 1963;**21:**699-716

Rechute 6-MP vs Placebo



Comparaison des distributions de survie

- Généralisation des test de ***rang*** aux données censurées: les ***rangs*** des observations *triées globalement* se distribuent ils au hasard entre les deux groupes?
  - ***le test du "log-rank"*** est le test efficace contre l'alternative de taux proportionnels dans les deux groupes
  - C'est formellement le principe du test de ***Mantel-Haenszel-Cochran***

Test du “Log-rank”

- Calcul pratique
  - Calculer pour chaque date de décès le ***nombre attendu*** de décès qui serait observé si les taux de décès étaient les mêmes dans les (deux) groupes
  - Cumuler les différences entre ***observés*** et ***attendus*** jusqu ’à la date du dernier décès observé
  - Evaluer la ***signification*** de la différence cumulée en la comparant à son erreur type
- Noter que ce test ne prend en compte que les rangs d’apparition des décès non les dates

Test du Log-rank (technique)

- **Pour chaque date de décès construire la table:**

	Groupe 1	*Groupe 2*	*Total*
décès	d_{1i}	d_{2i}	d_{+i}
effectifs	n_{1i}	n_{2i}	n_{+i}

- **Sous l'hypothèse nulle H_0, les d_{+i} décès se distribuent proportionellement aux effectifs**
  - Le nombre attendu dans le groupe 2 est donc

$$e_{2i} = d_{+i} \times n_{2i} / n_{+i}$$

  - d_{+i} étant fixé, la variance de $\Delta_i = d_{2i} - e_{2i}$ est:

$$Var(\Delta_i) = \frac{n_{1i} n_{2i} d_{+i} (n_{+i} - d_{+i})}{n_{+i}^2 (n_{+i} - 1)}$$

  - test $= (\Sigma_i \Delta_i)^2 / \Sigma_i Var(\Delta_i)$

Si k est grand, ou si les marges de chaque tableau sont grandes,

$$U = \sum_{i=1}^{k} w_i \left(d_{2i} - e_{2i}\right) = \sum_{i=1}^{k} w_i \left(d_{2i} - d_{+i}\frac{n_{2i}}{n_{+i}}\right)$$

Suit une loi asymptotiquement normale. Sous H_0, le test s'écrit

$$\chi^2 = \frac{\left[\sum_{i=1}^{k} w_i \left(d_{2i} - d_{+i}\frac{n_{2i}}{n_{+i}}\right)\right]^2}{\sum_{i=1}^{k} w_i^2 d_{+i}\frac{\left(n_{+i} - d_{+i}\right)}{\left(n_{+i} - 1\right)}\frac{n_{1i}n_{2i}}{n_{+i}^2}} \quad \chi^2 \, à \, 1 \, ddl$$

$$\chi^2 = \frac{\left[\sum_{i=1}^{k} w_i \left(d_{2i} - d_{+i}\frac{n_{2i}}{n_{+i}}\right)\right]^2}{\sum_{i=1}^{k} w_i^2 d_{+i}\frac{\left(n_{+i}-d_{+i}\right)}{\left(n_{+i}-1\right)}\frac{n_{1i}n_{2i}}{n_{+i}^2}}$$

Si w_i = 1, test de Mantel-Haenszel, de Cox, ou du Logrank

comparant

$$O_2 = \sum_{i=1}^{k} d_{2i} \qquad \textbf{et} \qquad E_2 = \sum_{i=1}^{k} e_{2i}$$

$$\chi^2 = \frac{\left(O_2 - E_2\right)^2}{\sum_{i=1}^{k} v_i}$$

$$\chi^2 = \frac{(O_2 - E_2)^2}{\sum_{i=1}^{k} v_i}$$

La formule approchée du logrank

$$\chi_a^2 = \frac{(O_2 - E_2)^2}{E_2} + \frac{(O_1 - E_1)^2}{E_1}$$

est conservative !

Exemple 2:6-MP vs Placebo

T	n1	d1	n2	d2	n+	d+	e2	var	e1
1	21	0	21	2	42	2	1.000	0.488	1.000
2	21	0	19	2	40	2	0.950	0.486	1.050
3	21	0	17	1	38	1	0.447	0.247	0.553
4	21	0	16	2					
5	21	0	14	2					
6	21	3	12	0					
7	17	1	12	0					
8	16	0	12	4					
10	15	1	8	0					
11	13	0	8	2					
12	12	0	6	2					
13	12	1	4	0					
15	11	0	4	1					
16	11	1	3	0					
17	10	0	3	1					
22	7	1	2	1					
23	6	1	1	1					
Total		**9**		**21**			**10.749**	**6.257**	19.251

Log-rank test $\chi^2 = (21-10.749)^2 / 6.257 = 16.794$

Log rank test, extension

- Le test s'étend simplement dans deux directions:
  - **Comparaison de plus de deux groupes**: la table 2x2 est remplacée par une table 2xk. La variance par une matrice de covariance. Le test est à k-1 ddl.
  - **Contrôle pour des facteurs de confusion:** on stratifie les données selon les valeurs des variables à contrôler. On calcule les observés, les attendus et les variances dans chaque strate. On cumule les résultats pour construire le test. *(les attendus sont calculés à "facteur de confusion constants")*

Conclusion

- On dispose de nombreux outils pour estimer la *distribution de durée de survie* éventuellement censurée
- Des *tests de comparaison* de distributions existent
- On aimerait avoir des *modèles de régression* pour gérer simultanément plusieurs co-variables.

1. Introduction - Analyse de la survie

2. Méthode de Kaplan et Meier - Test du Log-Rank

3. Taux de Mortalité

4. Survie paramétrique / Survie exponentielle par intervalle

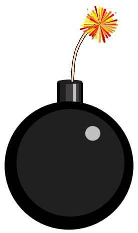
5. Modèle à Taux proportionnel

3. Taux de mortalité

- Objectif: Quantifier le risque de décéder *maintenant* si on était en vie *jusqu'à maintenant*
  - Le taux de décès *λ(t)* mesure la *« force de mortalité »* appliquée à la population des patients *survivants* à la date *t* après le diagnostic

$$\lambda(t) = \lim_{dt \to 0} \frac{\Pr\left(t \leq T < t + dt \middle| T > t\right)}{dt}$$

- Le *taux* de décès **n 'est pas** une *probabilité* en particulier il est éventuellement >1



Relation entre *R, S et λ*

dérivée de R (densité)

$$\lambda(t) = \frac{R'(t)}{1 - R(t)}$$

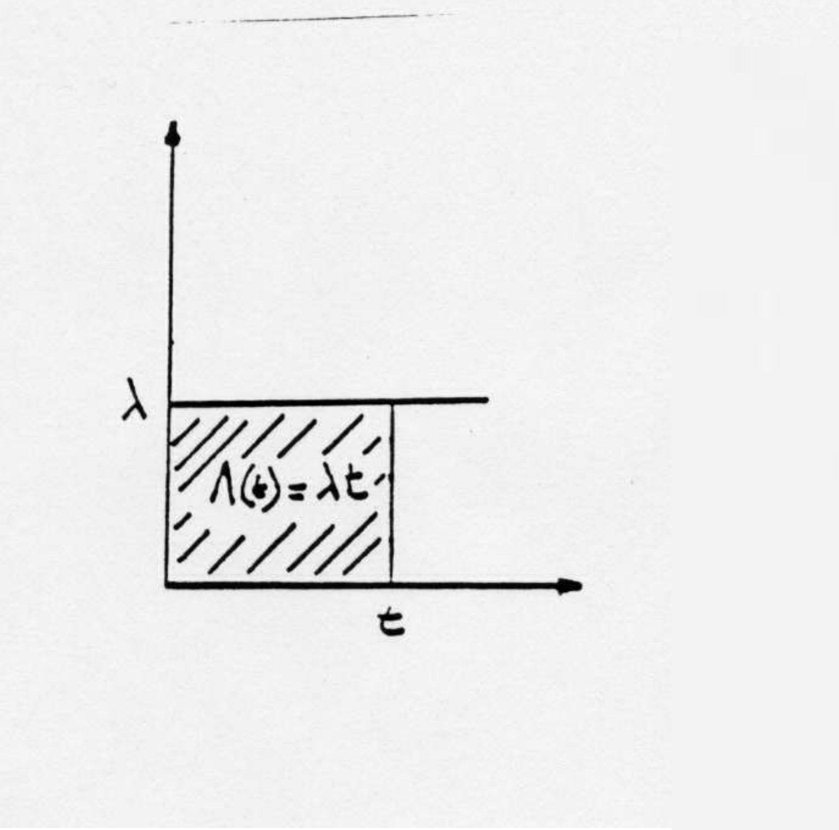
taux cumulé

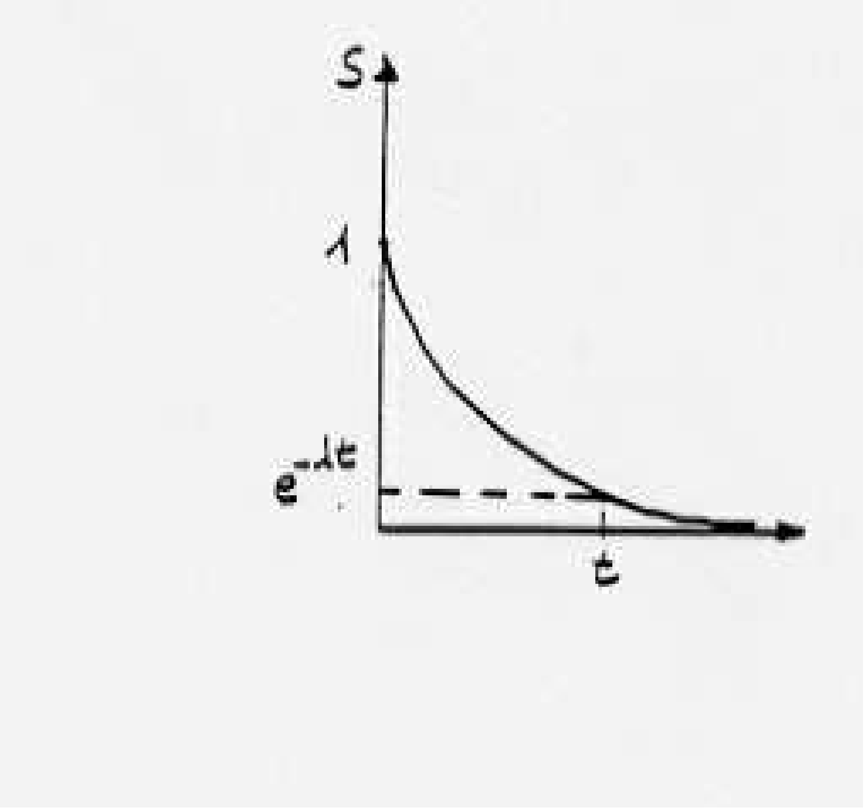
$$Log(1 - R(t)) = -\int_0^t \lambda(u)du = -\Lambda(t)$$

$$S(t) = e^{-\Lambda(t)}$$

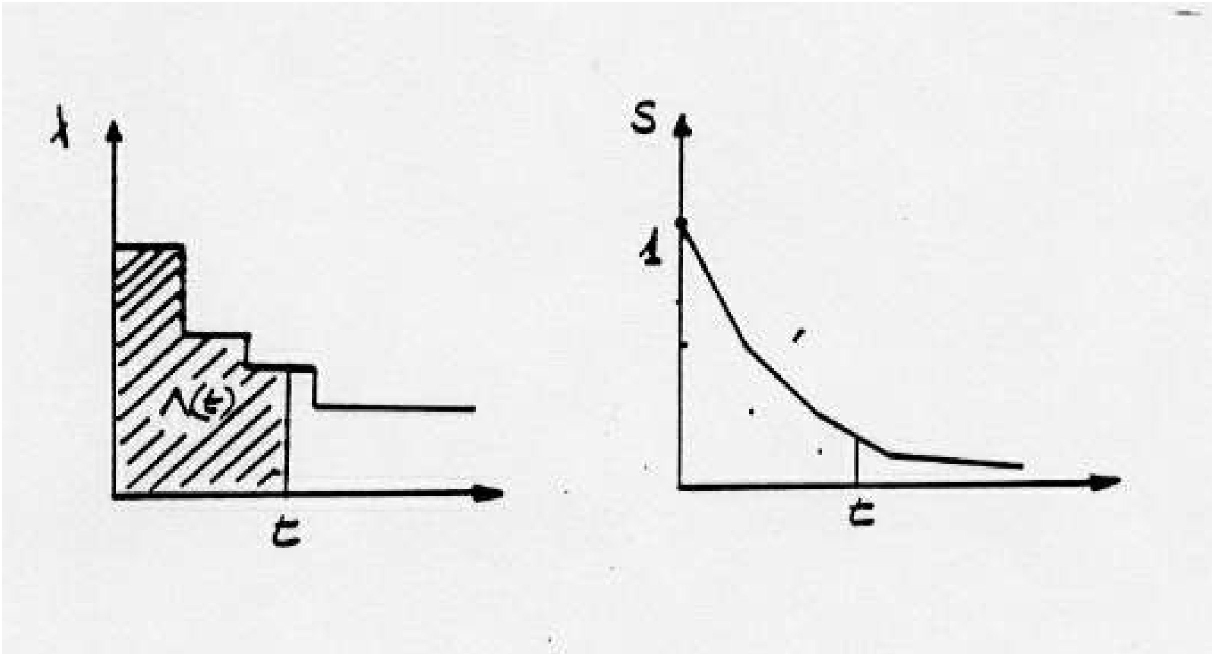
lorsque $\lambda(t)$ est constant (= λ) la survie est dite ***exponentielle:*** $S(t) = e^{-\lambda t}$

4. Survie exponentielle

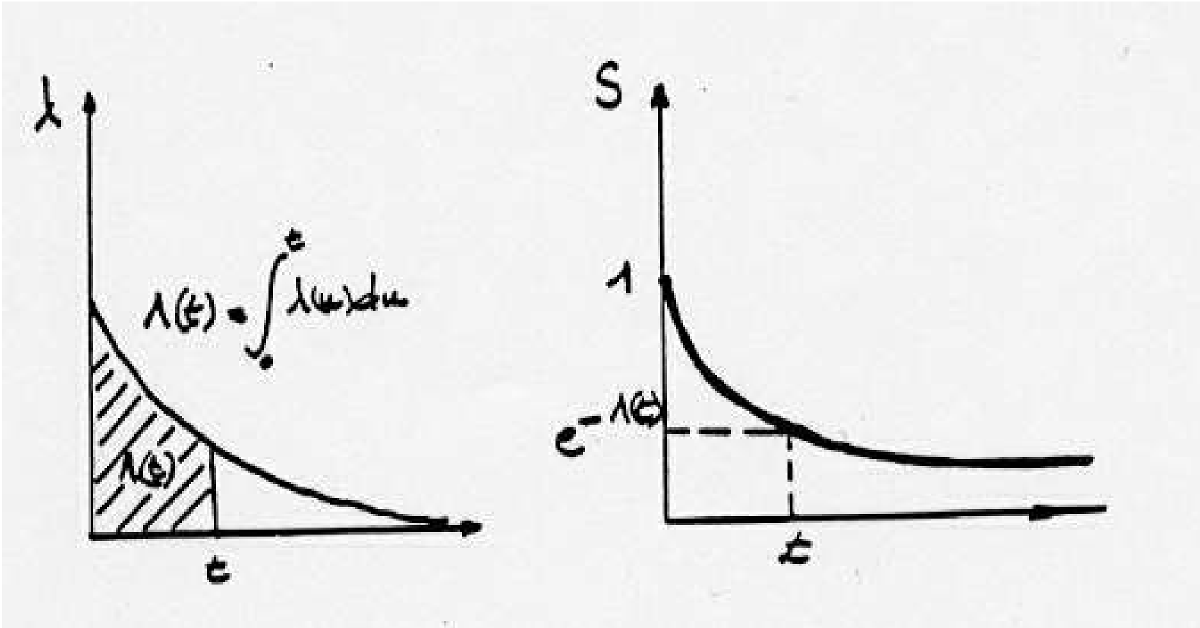




Survie exponentielle par intervalle



Cas général



1. Introduction - Analyse de la survie

2. Méthode de Kaplan et Meier - Test du Log-Rank

3. Taux de Mortalité

4. Survie paramétrique / Survie exponentielle par intervalle

5. Modèle à Taux proportionnel

Exemple 1: temps de survie en années

100 données complètes

1.22	1.86	1.35	3.81	0.47	1.63	2.47	0.42	0.24	0.45
2.83	5.17	18.14	4.92	0.30	2.87	9.21	1.32	2.09	3.48
1.58	5.16	2.43	0.05	3.00	0.24	4.01	0.34	1.59	7.39
0.42	1.08	0.63	0.56	1.79	0.63	2.60	0.11	1.43	3.97
1.22	7.18	3.66	0.10	2.29	3.40	4.13	2.78	4.82	4.87
5.81	1.25	4.08	6.81	6.02	0.15	1.66	3.60	0.18	0.12
2.72	15.56	0.72	1.35	0.72	8.68	1.96	1.97	4.75	0.62
2.94	2.27	4.53	10.22	1.03	1.84	4.81	1.66	0.34	1.78
0.55	0.25	3.00	0.25	0.48	8.74	0.64	6.41	14.59	1.93
0.10	4.08	11.82	1.78	1.51	1.96	0.41	1.12	1.28	0.60

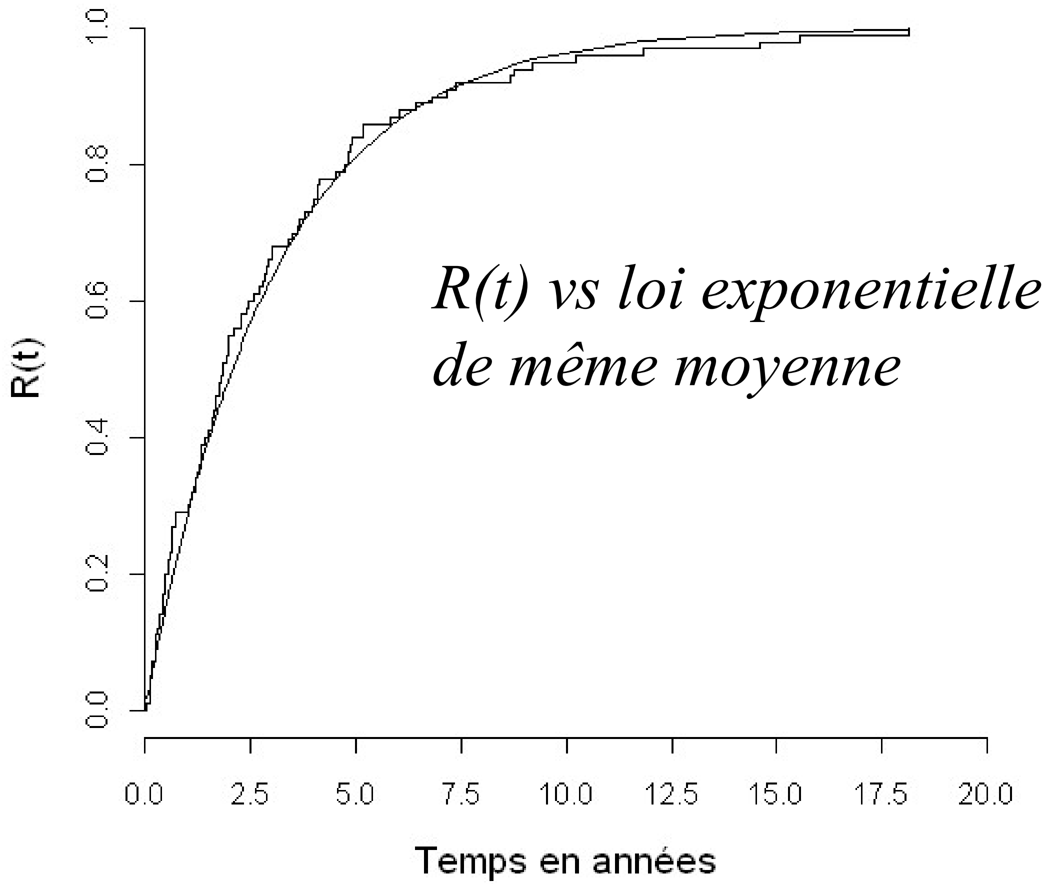
moyenne: 2.994 ans   déviation standard: 3.363 ans

Les données sont ici parfaitement décrites par une distribution exponentielle ayant la même moyenne que les données du tableau 1. La distribution exponentielle de moyenne $1/\lambda$ est le plus simple des modèles de survie paramétrique avec

$$S(t) = \exp(-\lambda t)$$
$$R(t) = 1\text{-}\exp(-\lambda t)$$

Le paramètre λ est le taux de mortalité, ou nombre de décès par unité de temps, estimé par $n_0 /(\Sigma t_i)$. Les données étant complètes, on a en effet observé n_0 décès pour une durée totale d'observation de ces sujets égale à $(\Sigma\ t_\iota)$. Cette quantité s'exprime en personnes années d'observation : années d'observation accumulées par l'ensemble des personnes étudiées.

La fonction de répartition exponentielle qui s’ajuste le mieux aux données est la fonction 1- exp (-1/2.994 x t) où 2.994 est la moyenne des temps de survie des données.



Modèles paramétriques I

- Définir $\lambda(t)$ à l'aide d'une fonction connue aux paramètres près, par exemple:

Nom	Taux de mortalité	Survie
exponentie lle	λ	$e^{-\lambda t}$
Weibull	$\gamma\lambda^{\gamma} t^{\gamma-1}$	$e^{-(\lambda t)^{\gamma}}$
Gompertz	...	....
Lognormal	...	....

Modèle exponentiel Relation entre *R, S et* λ

$$S(t) = e^{-\lambda t}$$

$$R(t) = 1 - e^{-\lambda t}$$

$$R'(t) = \lambda e^{-\lambda t}$$

$$\lambda(t) = \frac{R'(t)}{1 - R(t)} = \lambda$$

$$Log\big(S(t)\big) = Log(1 - R(t)) = -\int_0^t \lambda \, du = -\lambda t = -\Lambda(t)$$

Modèle de Weibull Relation entre *R, S et* λ

$$S(t) = e^{-(\lambda t)^{\gamma}}$$

$$R(t) = 1 - e^{-(\lambda t)^{\gamma}}$$

$$R'(t) = \gamma \lambda^{\gamma} t^{\gamma-1} e^{-(\lambda t)^{\gamma}}$$

$$\lambda(t) = \frac{R'(t)}{1 - R(t)} = \gamma \lambda^{\gamma} t^{\gamma-1}$$

$$Log\left(S(t)\right) = Log(1 - R(t)) = -\int_0^t \lambda(u)\,du = -(\lambda t)^{\gamma} = -\Lambda(t)$$

Modèle à 2 paramètres. Le modèle exponentiel correspond au cas où $\gamma = 1$. Si $\gamma > 1$, le taux est une fonction croissante du temps. Si $\gamma < 1$, le taux est une fonction décroissante du temps. Le taux instantané est une puissance du temps.

Estimation par intervalle(I)

- On subdivise la durée de suivi en intervalles dans lesquels le taux de décès est supposé constant:
  - *$t_0=0$, t_1, t_2, ...t_I=fin du suivi*
- On compte le nombre de personnes n_{i-1} exposées au risque de décès en t_{i-1} pour chaque i
- ....

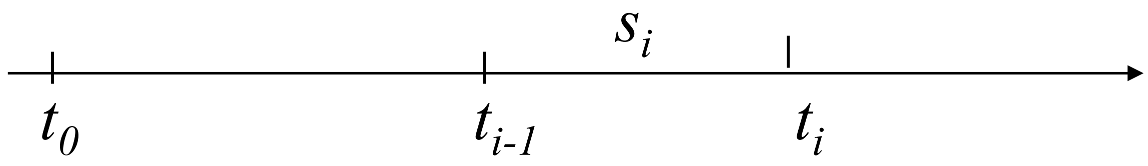
Estimation par intervalle(II)

- On compte le nombre de ***décès*** $\boldsymbol{d_i}$ et de ***sortis vivants*** $\boldsymbol{c_i}$ dans l'intervalle $\boldsymbol{[t_{i-1}, t_i[}$

- On estime
  - *le taux de décès* λ_i dans chaque intervalle
  - la probabilité conditionnelle $\boldsymbol{s_i = e^{-\lambda_i(t_i - t_{i-1})}}$ de survivre en $\boldsymbol{t_i}$ si on était ***vivant en*** $\boldsymbol{t_{i-1}}$

Estimation par intervalle(III)

$S(t_{i-1})$ $\quad S(t_i) = S(t_{i-1})*si$



Pour survivre en t_i ($S(t_i)$) il faut

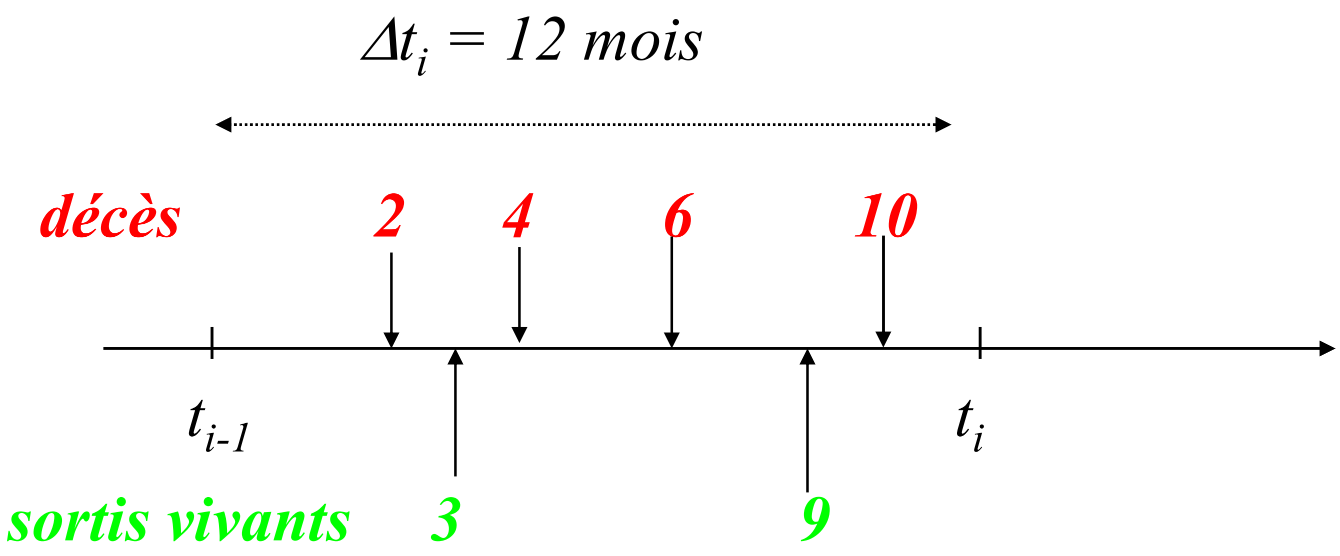
•***survivre jusqu'en t_{i-1} ($S(t_{i-1})$)***

•***survivre dans l'intervalle $]t_{i-1}\ t_i]$ (s_i)***

La survie en t_i peut ainsi être calculée comme le produit de la survie dans des intervalles successifs....

$$S(t_i) = s_1 * s_2 * .....s_i$$

Estimation de λ_i (Exemple)



$n_{i-1} = 40$ $n_i = 34$

$$\hat{\lambda}_i = \frac{4}{34 \times 12 + (2+3+4+6+9+10)} = \frac{4}{442} = 9.050 / 1000\, p.m$$

5. Modèles à taux proportionnels

- **Dans le cas d 'un facteur dichotomique le modèle s 'écrit:**
  - » $\lambda_1(t)=\alpha\lambda_0(t)$; α est appelé le ***taux relatif*** du groupe *1* par rapport au groupe *0*

- **L'idée du modèle à *taux de mortalité proportionnels* est implicite dans le test du « log-rank » qui évalue l'hypothèse *"α=1 ?"***

- **Si on modélise linéairement le paramètre λ des distributions *exponentielles et de Weibull* on obtient des modèles à taux proportionnels**

- **Le *modèle de Cox* est un modèle à taux proportionnels dans lequel la forme analytique du taux n'est pas spécifiée.**

Modèles à taux proportionnels II

- Le modèle s 'écrit $\lambda(t,z)= \lambda (t,0)exp(\beta z)$, la fonction $\lambda (t,0)$ est *le taux de base (inconnu), z un vecteur de covariable (mesurée),* β *un vecteur de paramètre (à estimer)*
- Le *taux cumulé* et *la survie* sont donc:

$$\Lambda(t,\mathbf{z}) = \Lambda(t,0) \times e^{\beta \mathbf{z}}$$

$$S(t,\mathbf{z}) = e^{-\Lambda(t,\mathbf{z})} = S(t,0)^{\exp(\beta z)}$$

Modèles à taux proportionnels III

$$Log\left(\frac{\lambda(t,\mathbf{z})}{\lambda(t,\mathbf{0})}\right) = \sum_{i=1}^{m}\beta_i z_i = \beta\mathbf{z}$$

Le log du taux relatif est une fonction linéaire des covariables

$$Log\left(\frac{\lambda(t,z_1,..,z_i..z_m)}{\lambda(t,z_1,..,0,..z_m)}\right) = \beta_i z_i$$

$\alpha_i = e^{\beta_i}$ est le taux relatif des sujets pour lesquels z_i=1 par rapport à ceux pour lesquels z_i=0, *' toutes choses égales par ailleurs '*

Exemple 2:analyse avec le modèle de Cox

- Etudier le temps de remission *avec le nouveau traitement 6-MP*; comparer la distribution du temps de remission dans les deux groupes *'traitement standard (corticoïdes+placebo)* et *'nouveau traitement (corticoïdes et 6-MP)*
- le modèle est
  - $\lambda(t,z)=\lambda(t,0)exp(\beta z)$; $\boldsymbol{z=1}$ pour le nouveau traitement, et $\boldsymbol{z=0}$ pour le traitement standard
  - $\boldsymbol{\alpha=exp(\beta)}$ est le ***taux relatif*** de rechute ***dans le groupe 6-MP*** comparé au groupe placebo
  - le taux de rechute dans le groupe traitement est $\lambda(t,0)$

Ajustement pour une variable de confusion

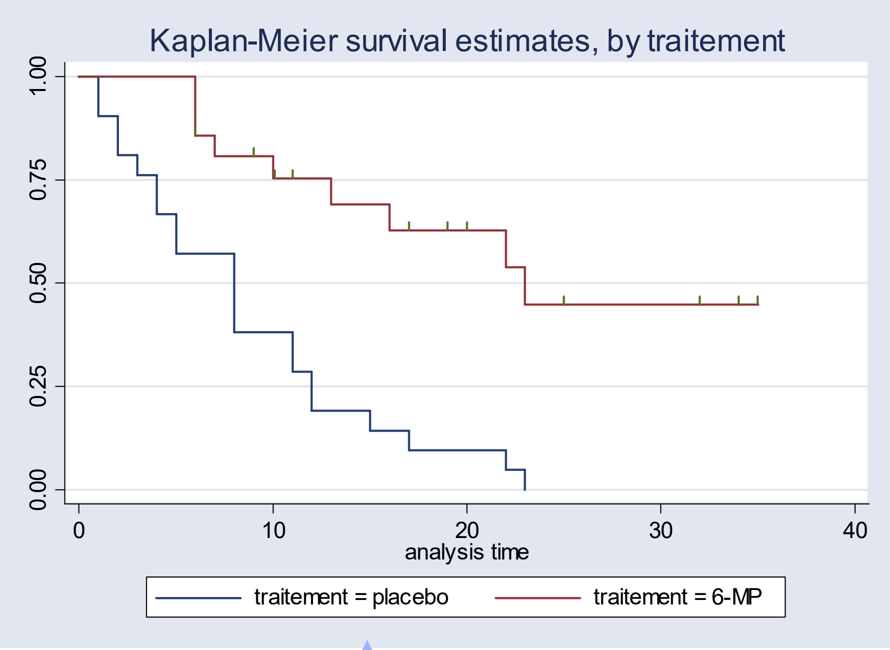
- Bien que l'étude soit randomisée la valeur de logwbc est différente pour les traités et les non traités. Il convient donc de prendre en compte cette variable dans l'évaluation de la différence.
- Le méthode de regression réalise cela avec le modèle:

λ(t, traitement, logwbc)=λ(t,0,0)exp(β*traitement + γ*logwbc)

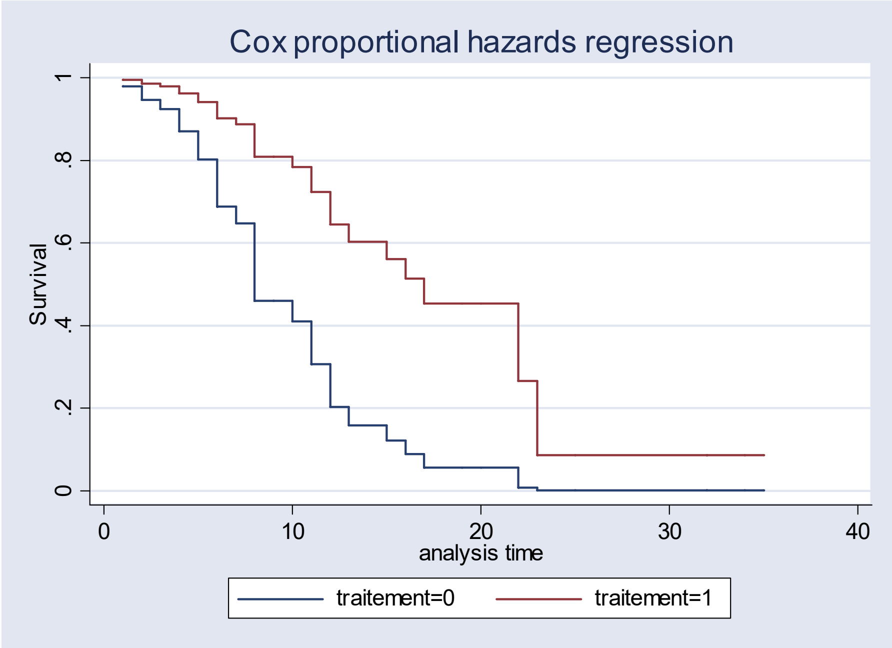
α=exp(β) **est alors le taux relatif des** ***"traités" vs "placebo" à wbc constant***

β=-1.2941 **exp(β) = *0.2742***

Survie ajustée



Comparaison ajustée
Pour Logwbc=2.930 (moyenne)



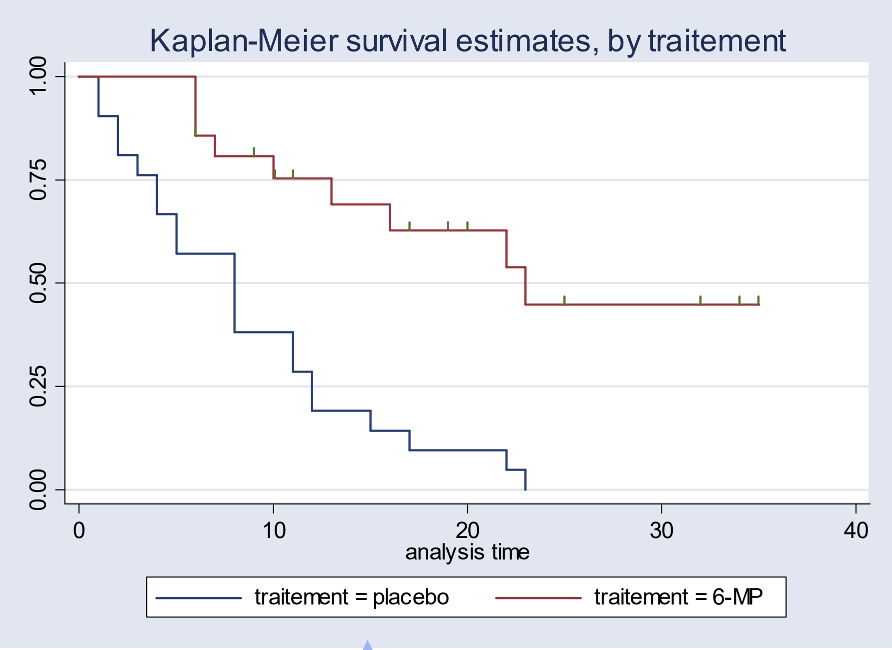
Comparaison KM, le groupe traité bénéficie de Wbc plus favorables

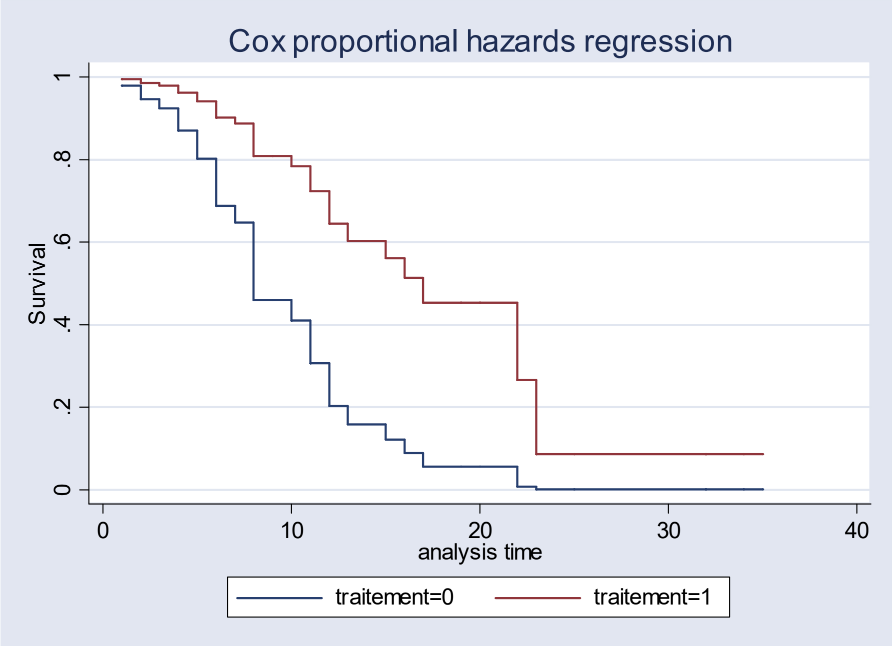
Survie ajustée

$$0.80^{0.2742} = 0.94$$

$$0.40^{0.2742} = 0.78$$

$$0.20^{0.2742} = 0.64$$





Comparaison KM, le groupe traité bénéficie de Wbc plus favorables