,

Analyse de survie

Vaianesthésie

Coucou !! Nouveau cours aujourd'hui, malheureusement mon dernier cours (🥲, je suis très triste). Il est assez long aussi mais je vous assure qu'il est cool ! En tout cas, je vais tout faire que vous l'appréciez ! C'est parti !

Mes explications/remarques seront en italique et de cette couleur. Je vous le signalerais aussi avec un emoji « 🦋 » et un encadré comme celui-ci pour les explications plus longues !

ANALYSE DE SURVIE

SOMMAIRE

I. Introduction

Les méthodes d'analyse de survie sont des méthodes de référence pour décrire les **données longitudinales** recueillies lors d'un **suivi** de sujets ou de groupes de sujets.

Une étude de survie est une étude :

- Longitudinale
- Prospective
- Observation d'un groupe de sujets : une **cohorte**

> **Remarque** : Les analyses de survie essaient de modéliser la survenue d'un évènement en fonction du temps.

On rencontre un très grand nombre de situations pratiques dans lesquelles le centre d'intérêt est la **survenue d'un évènement**. Il peut s'agir :

- D'un décès,
- De la survenue d'une complication après un geste opératoire,
- De la rechute d'une maladie après une période de rémission,
- De la disparition de symptômes sous traitement, etc.

La méthodologie introduite dans ce cours s'appliquera sans modification à tout type d'évènement à la survenue duquel on s'intéresse. Cependant, pour la commodité de l'expression, on parlera généralement dans la suite de survie, considérant ainsi que l'évènement d'intérêt est le **décès**.

L'évènement considéré doit être défini de la même manière pour TOUS les sujets.

S'intéresser à la survenue, dans le temps, d'un évènement, c'est s'intéresser au **délai** de survenue de cet évènement, délai compté à partir de **l'instant de référence (ou date d'origine)** :

- Dans le cas d'un décès pris comme évènement d'intérêt, dire d'un patient qu'il survit au moins un certain temps c'est dire que le délai de survenue du décès est supérieur à ce temps.

Aloooooors, j'explique au cas où ce n'est pas clair ! Si t'as compris, tu passes !

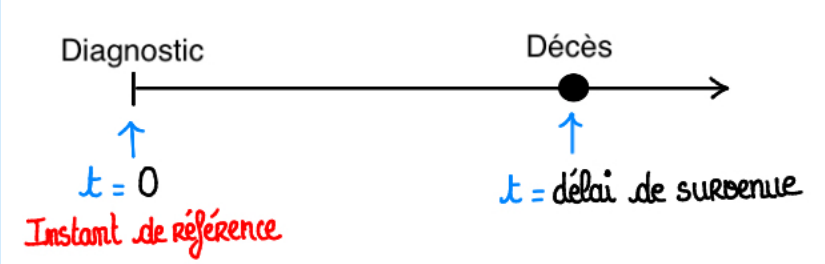
*S'intéresser à la survenue d'un évènement, c'est regarder **au bout de combien de temps** il arrive à partir d'un **moment de départ** (appelé* instant de référence*).*

Je prends un exemple pour vous expliquer :
*On part du jour où un patient est diagnostiqué malade (*instant de référence*), et on observe **combien de temps il vit avant son décès** (*évènement d'intérêt*).*

➔ *Dire qu'un patient **survit au moins 3 ans**, cela veut dire que **pendant ces 3 ans il n'est pas encore décédé**. Donc le **délai avant décès** est **supérieur à 3 ans**.*

En résumé :

- *On mesure un **temps entre deux moments** (début → évènement).*
- *Si l'évènement n'est pas encore arrivé, on sait juste que ce **temps est plus grand** que la durée observée.*



Petit mémo qui peut aider : On cherche le temps entre un point de départ et un évènement.

- *Instant de référence : le départ (ex : diagnostic, début de suivi)*
- *Évènement d'intérêt : ce qu'on attend (ex : décès, rechute, guérison)*
- *Délai de survenue : temps écoulé entre les deux !*

*Si l'évènement n'est pas encore arrivé, le délai est **plus grand** que le temps déjà passé.*

Petite analogie si jamais :

Imagine que tu retourner un sablier le jour où le patient est diagnostiqué.

➔ Tant que le patient est en vie, le sable continue de couleur → donc le délai avec décès augmente.

➔ Le jour où il décède, le sablier s'arrête → on connait enfin le temps exact entre le départ et l'évènement.

Donc dire « il a survécu au moins 5 ans, c'est comme dire « le sablier n'est toujours pas vide après 5 ans ».

Voilà, j'espère que c'est clair au cas où la phrase te semblait bizarre ! Désolé, j'ai beaucoup parlé mais je veux être sûre que vous ayez bien compris ! Je veux que vous soyez des machines !

Les objectifs d'une analyse de survie sont **d'estimer** et **d'expliquer** la **durée de survie** en fonction de certains **facteurs** (on parle de **facteurs pronostiques**) et, souvent, de **comparer la survie** entre 2 groupes de sujets ou plus.

- Un **facteur pronostique** est un facteur susceptible d'expliquer la survenue du décès (ou d'un autre évènement) au cours du temps. Ils influencent de manière positive ou négative la survie.

> **Remarque** : Les facteurs pronostiques sont à **différencier des facteurs de risques +++**

Au total on s'intéresse à :

➔ La **probabilité de survivre au moins un certain temps t** à compter d'un instant de référence, ou encore à

➔ La **probabilité pour que l'évènement d'intérêt survienne après un délai t** à compter de l'instant de référence.

Exemple :

On s'intéresse aux complications post opératoires en chirurgie digestive. On observe les patients pendant une semaine après leur opération. Au bout d'une semaine, on connaîtra le nombre de personnes avec complications et le nombre de patients sans complications. On pourra ensuite se demander quand les complications surviennent, leur dynamique, leur répartition. Ainsi, on pourrait trouver qu'au bout de 48h, 35% des patients ont eu des complications. C'est le fameux « time to event » : délai jusqu'à l'apparition de l'événement.

II. Définitions

a. Cohorte

Une **cohorte** correspond à un ensemble de sujets qui vivent les **mêmes évènements** au **même moment**. En matière de recherche médicale, c'est un ensemble de sujets inclus dans une étude au même moment, et suivis dans des **conditions standardisées** pendant une **durée prédéfinie**.

Une **cohorte « incipiente »** (néologisme « inception cohort ») doit inclure des **sujets observés au début de leur affection** à un point uniforme de l'évolution de leur maladie (« cas incidents »).

Une **cohorte idéale** correspond au fait que tous les patients sont inclus au même moment, tous les patients sont « alignés ».

b. Évènement d'intérêt

Contrairement à ce que le terme « survie » laisse penser, **l'évènement d'intérêt n'est pas** forcément le décès, mais peut-être aussi la survenue d'une maladie, la récurrence de symptômes après traitement, ou encore en dehors d'un contexte médical, la durée de vie des composants électroniques etc.

En pratique, **les méthodes d'analyse de « survie » doivent donc être appliquées à chaque fois qu'il existe une notion de durée jusqu'à l'évènement d'intérêt**. Dans ce cours, la terminologie « survie » sera utilisée quel que soit le type d'évènement d'intérêt.

Lorsque l'évènement d'intérêt est le décès :

- On peut s'intéresser **aux décès toutes causes** et, dans ce cas, chaque décès de patient compte comme un évènement.

- On peut également ne s'intéresser qu'au **décès pour une cause spécifique** (par exemple, décès par accident coronaire) et, dans ce cas, les décès d'autres causes (***par exemple, décès par cancer***) ne comptent pas comme un évènement, mais comme une **censure**. Ceci n'est possible que lorsque les « autres causes de décès » sont indépendantes du phénomène étudié.

> **Remarque** : Une cohorte est un groupe d'individus qui suit les mêmes contraintes et qu'on suit dans le temps. Cela peut durer des années !

Exemple :

On peut s'intéresser à l'apparition de maladies cardiovasculaires dans la promo des LAS 2024-2025. C'est une cohorte, tout le monde arrive au même moment, est suivi et soumis aux mêmes contraintes.

c. Durée de survie

La **durée de survie** est le délai entre les deux dates :

Prenons un exemple schématique :

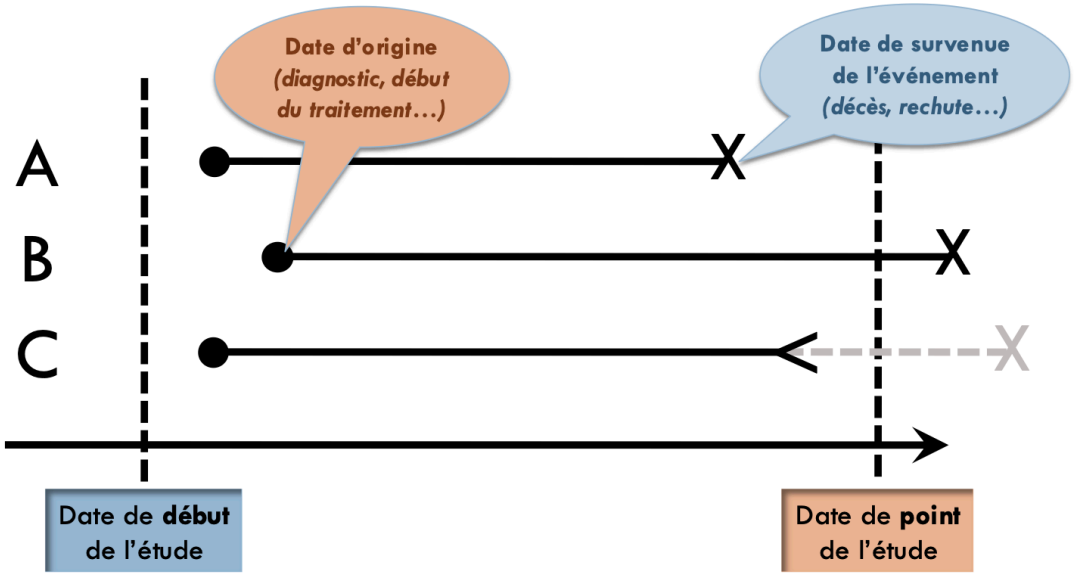
Trois patients A, B et C sont inclus dans une étude de suivi longitudinal (étude de cohorte …).

Il y a divers paramètres à prendre en compte :

1. Date d'origine

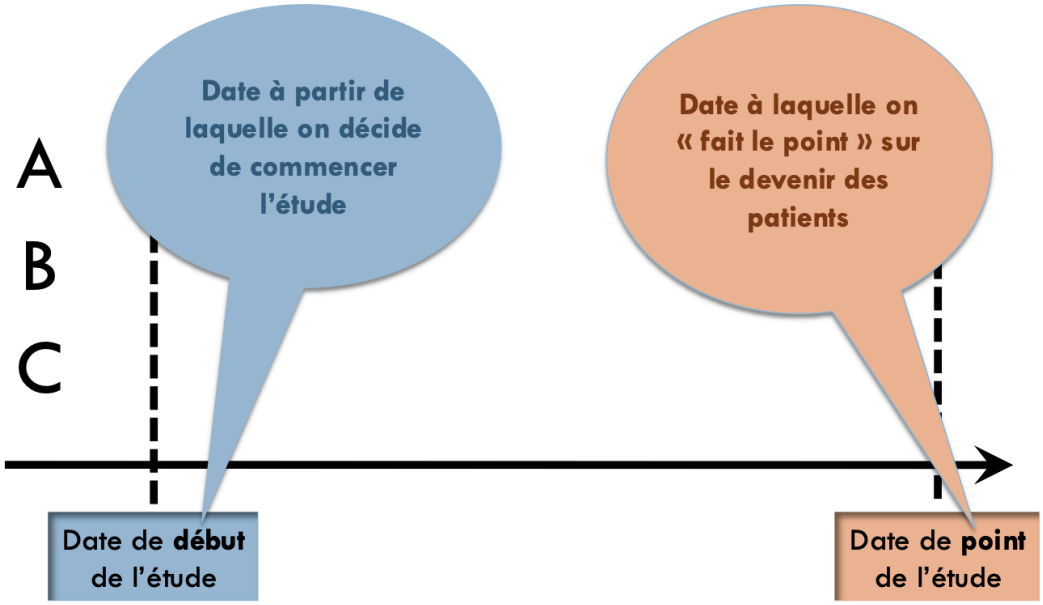
La **date d'origine** est une date calendaire indiquant **le point de départ de la surveillance** : par exemple, la date de randomisation dans un essai thérapeutique.

Cette date d'origine peut être identique ou différente pour chaque sujet en fonction des modalités d'inclusion des sujets.



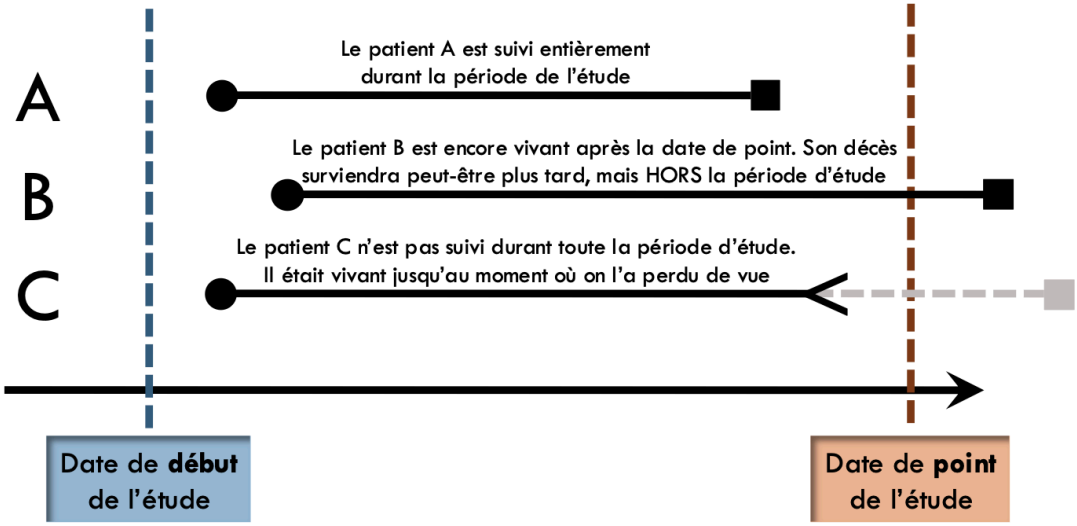
2. Date de point (end-point)

La **date de point** est une date fixe calendaire et correspond à la **date choisie pour faire le bilan**, au-delà de laquelle les informations recueillies ne sont plus considérées dans l'analyse.



3. Date de dernières nouvelles

La **date de dernières nouvelles** est la date la plus récente à laquelle on a recueilli des informations sur le patient, notamment sur la survenue ou non de l'évènement étudié.

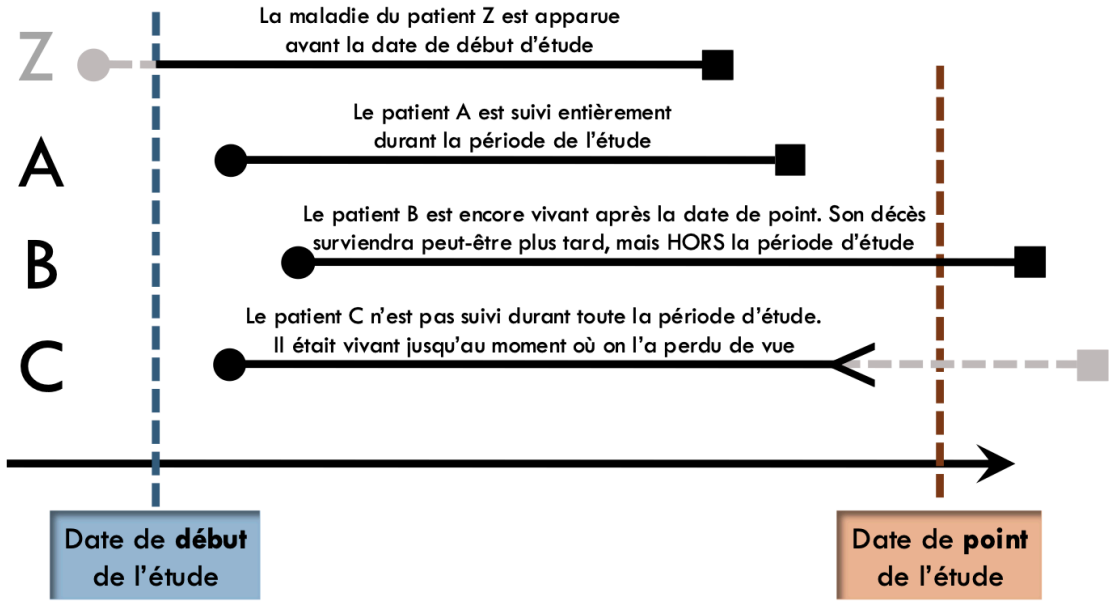


4. Cas particuliers

Dans certains cas, la date d'origine peut être **antérieure** à l'inclusion dans l'étude (on parle alors de cohorte « historique »).

Exemple :

Il peut s'agir de la date de découverte d'une hypertension artérielle dans une étude de cohorte portant sur les facteurs de risque de mortalité cardio-vasculaire.



Dans certaines études, le point de départ du suivi est évident.

Par exemple, pour les complications d'une intervention chirurgicale, le temps 0 correspond au jour de l'opération, et on suit ensuite les patients dans les jours ou mois qui suivent.

Cependant, dans d'autres situations, le point d'origine du temps n'est pas l'inclusion dans l'étude. Prenons l'exemple de la ***survenue d'un AVC chez des patients hypertendus****. Ces patients peuvent être hypertendus depuis plusieurs années avant leur inclusion dans l'étude. Si un patient fait un AVC deux mois après son inclusion, peut-on dire que le délai de survenue est de 2 mois ?* ***Non.***

*En effet, le risque d'AVC dépend de la **durée réelle de l'hypertension**, et non du temps écoulé depuis l'entrée dans l'étude. Dans ce cas, il faut donc remonter à **l'origine de la maladie (début de l'hypertension)** pour définir correctement le temps à risque. C'est compris ?*

d. Perdus de vue (lost of follow-up)

Un sujet est dit **perdu de vue** lorsque sa surveillance est interrompue avant la date de point et que l'évènement ne s'est pas produit (cf. patient C).

Un cas particulier concerne les sujets inclus dans l'étude mais n'ayant fait l'objet d'aucun suivi. Ces sujets ne seront pas comptabilisés dans l'analyse. On parle alors de « perte de vue » d'emblée.

Dans tous les cas, il est d'usage de vérifier que le processus de perte de vue pour l'ensemble des sujets (perte de vue d'emblée, ou après une durée de suivi) n'est pas lié à l'évènement d'intérêt.

Exemple :

En comparant les caractéristiques de ces patients à celles des sujets ayant fait l'objet d'un suivi complet.

e. Censure (censored data)

Une durée de survie d'un individu est dite **censurée** lorsque l'évènement d'intérêt n'a pas été observé pour cet individu. Elle concerne donc :

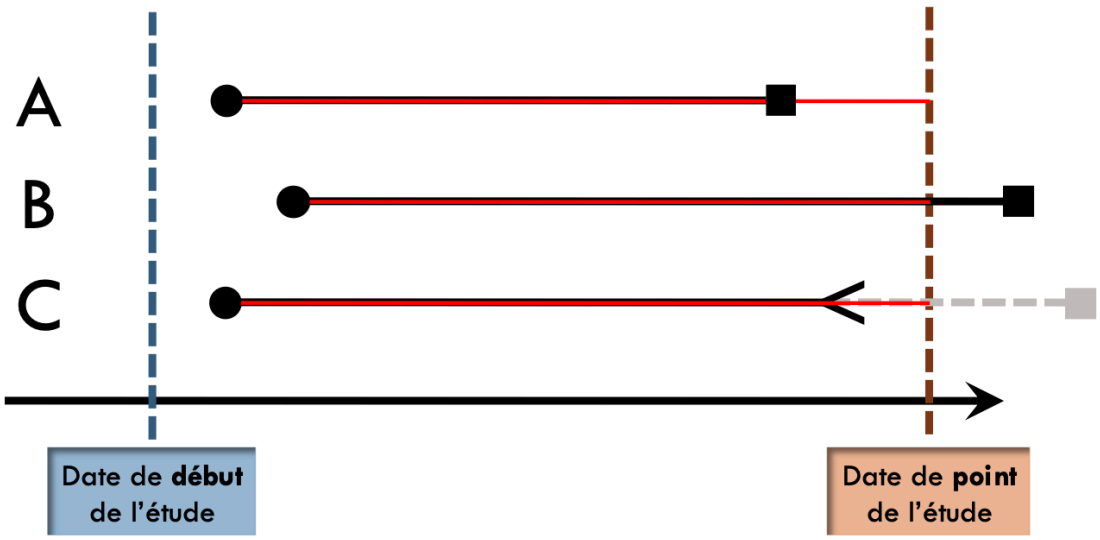
- Les sujets perdus de vue (patient C)
- Les sujets vivants à la date de point (souvent appelés exclus-vivants) (patient B)

Ces deux mécanismes de censure sont de nature différente.

En effet, on ne peut assimiler les perdus de vue aux exclus-vivants, car la raison de leur « disparition » peut être liée à l'évolution de la maladie (décès méconnu de l'investigateur par exemple).

f. Temps de recul

Le **recul** est le délai entre la date d'origine et la date de point, c'est-à-dire le délai maximum potentiel de suivi pour un sujet. Les reculs minimum et maximum d'une série de sujets définissent donc l'ancienneté de cette série.



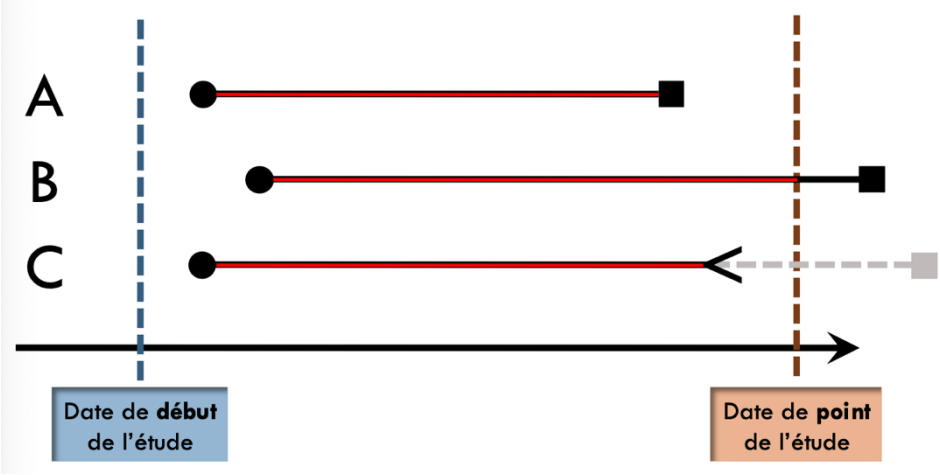
g. Temps de participation

Le **temps de participation** correspond à la durée de surveillance pour chaque sujet utilisé dans l'estimation de la survie. Trois situations peuvent se produire :

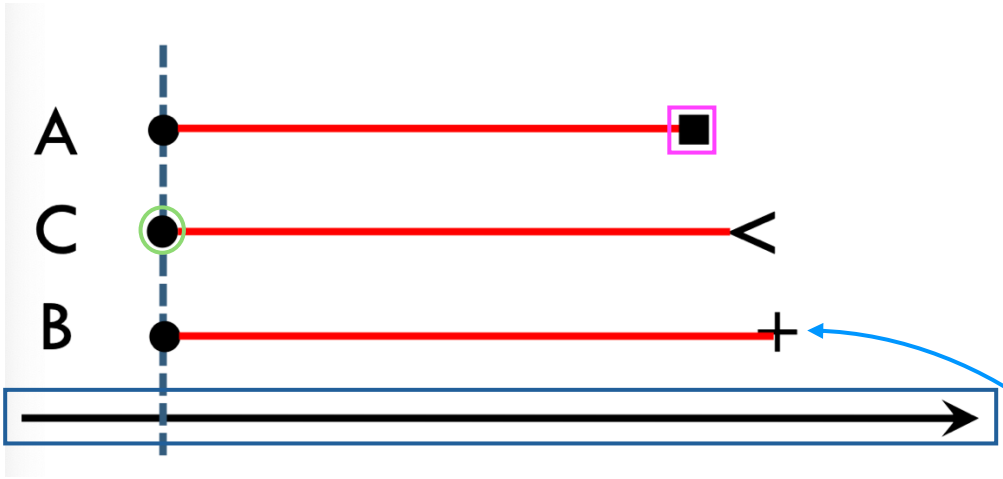
- L'évènement s'est produit au cours de la surveillance : le temps de participation est le délai entre la **date d'origine** et la survenue de l'évènement (patient A)

- Le sujet est vivant à la date de point : son temps de participation est le délai entre la **date d'origine** et la date de point (patient B)

- Le sujet est perdu de vue : dans ce cas, son temps de participation est défini par le délai entre la **date d'origine** et la date de dernières nouvelles (patient C)



Ce type de graphique est important pour comprendre que tout le monde n'a pas le même temps de participation. Dans une cohorte idéale, on voudrait que tous les individus aient le même temps de participation.



Vaïana, comment je dois lire un graphique de survie ? Regarde mon poussin !

Sur ce type de graphique :

- *L'axe horizontal représente le **temps**.*
- *Le point noir correspond à la **date d'origine (temps 0)**.*
- *Le carré noir indique la **survenue de l'événement** (ex : décès).*
- *Le signe + correspond à un **patient censuré** (vivant sans événement à la date de point).*
- *Le symbole « < » indique un **patient perdu de vue**.*
- *La ligne rouge représente la **durée de suivi** du patient.*

L'événement étudié n'est pas forcément un décès : cela peut être toute issue d'intérêt (rechute, AVC, complication…). On va donc observer la survenue d'un évènement.

Prenons un exemple avec trois patients : A, B et C : *Chaque patient a une évolution différente.*

- 👩 ***Patient A***
  - *Il entre dans l'étude à la date d'origine.*
  - *On le suit.*
  - *L'événement survient pendant la période d'étude.*

  <u>On connaît précisément le délai de survenue de l'événement</u>.
- 👱 ***Patient B***
  - *Il est inclus à la date d'origine.*
  - *À la date de point (fin d'étude), il n'a pas présenté l'événement*
  - *Il est **censuré**. On sait qu'il n'a pas eu l'événement jusqu'à cette date, mais il pourrait survenir plus tard.*

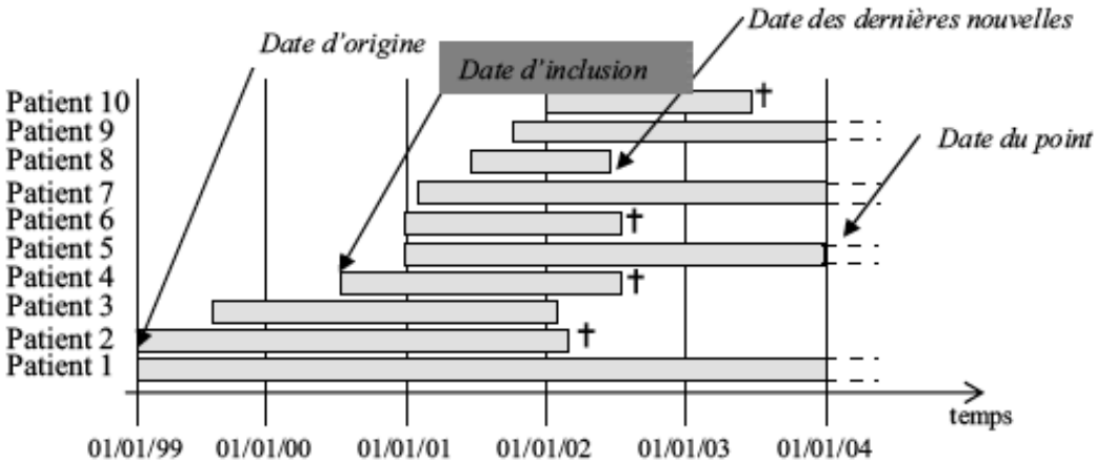
  <u>On ne dispose pas d'information au-delà.</u>
- 👧 ***Patient C***
  - *Il est inclus puis suivi.*
  - *À un moment, il est **perdu de vue** (il ne revient plus en consultation).*
  - *On ne sait pas si l'événement est survenu après sa dernière visite.*

  <u>Il est donc également censuré, mais pour perte de suivi.</u>

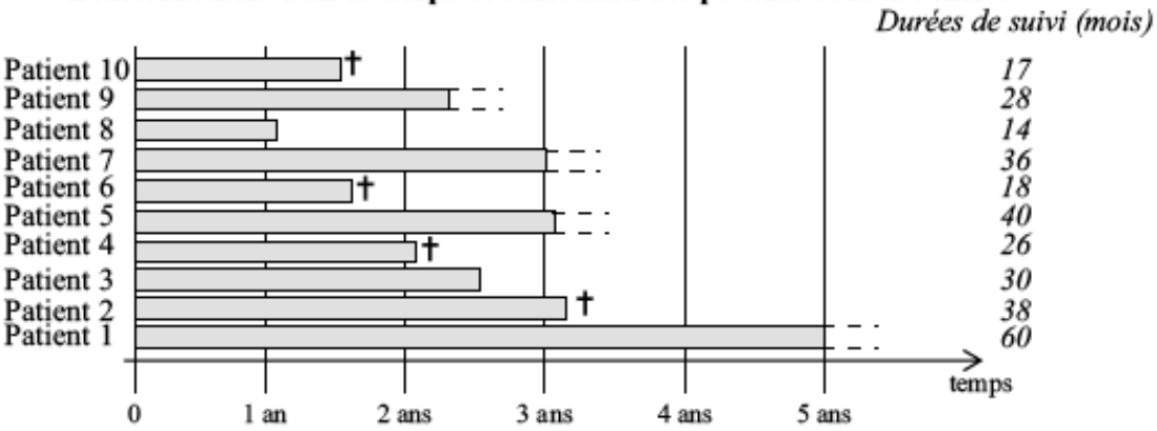
Dans tout ça, je veux que vous reteniez qu'*une* ***censure*** *signifie que l'on ne connaît pas la date exacte de l'événement. Elle peut être due :*

- *À la fin de l'étude,*
- *À une perte de vue,*
- *À un décès d'une autre cause (selon l'événement étudié).*

Les données sont utilisées ***jusqu'à la dernière information disponible****.*



Echelonnement dans le temps de l'inclusion des patients dans la cohorte



Description des durées de suivi

Le graphique du haut présente un calendrier de type grégorien, celui du bas un calendrier relatif. Pour le calendrier relatif, tous les patients sont alignés sur 0. On voit directement la durée de participation de chaque individu avec la survenue de l'événement (ici le décès) ou bien les perdus de vue.

Prenez une pause si besoin mes poussins, il reste encore pas mal de notions à voir …

III. Fonctions de survie

a. Loi exponentielle

La **loi de Poisson** (*révisez le cours de Claudia*) régit la survenue d'un évènement par unité de mesure (temps, volume, surface …). On démontre que si un évènement se réalise selon une loi de Poisson (de paramètre $\lambda = \mu = \sigma^2$), le temps entre deux réalisations consécutives de l'évènement considéré est distribué selon une **loi exponentielle d'espérance** $\frac{1}{\lambda}$ (λ est appelé le taux de défaillance instantané).

Avec ces lois, on est dans le domaine/modèle du **paramétrique** car on fait une hypothèse sur le comportement des évènements (et variable).

La **loi exponentielle** est utilisée couramment pour représenter la durée de vie de composants ou d'équipements pour lesquels l'hypothèse d'un taux de défaillance constant au cours du temps peut être justifiée. Cela implique que les défaillances sont dues uniquement au hasard et qu'elles se produisent selon un processus de Poisson.

La **fonction de densité** de la loi exponentielle est : pour tout $\boldsymbol{x} \geq \mathbf{0}$, $f(x) = \lambda e^{-\lambda x}$

La **fonction de répartition** de la loi exponentielle est donnée par l'équation :

$$F(t) = P\ (X \leq t) = \int_0^t \lambda e^{-\lambda x} dx = 1 - e^{-\lambda t}$$

De ce fait : $F(t) = 1 - e^{-\lambda t}$

$F(t)$ représente la proportion d'équipements (de composants, etc …) qui tombent en panne avant le temps t (c'est la fonction de « défaillance »). Ainsi, la quantité $1 - F(t)$ représente la quantité d'équipements qui fonctionnent pendant une durée de temps au moins égale à t. Cette quantité est notée $S(t)$ et s'appelle la **fonction de survie** :

$$\boldsymbol{S(t) = 1 - F(t) = P(X > t) = e^{-\lambda t}}$$

🦋 ***Olala, Vaïana j'ai peur … Relax, ce que vous devez retenir de cette page selon moi :***

*Si le **taux de risque λ est constant dans le temps**, alors la durée jusqu'à l'événement suit une **loi exponentielle**. C'est un modèle **paramétrique** (on suppose une forme mathématique).*

Ainsi :

- *λ = taux instantané de survenue de l'événement*
- *Plus λ est grand → plus le risque est élevé*
- *Espérance de vie moyenne = **1/λ***

Si vous devez retenir 2 formules ici selon moi (on ne sait jamais) :

↳ ***Fonction de répartition***

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

*C'est la probabilité que l'événement arrive **avant** **t**.*

↳ ***Fonction de survie***

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

*C'est la probabilité de survivre **au moins** **jusqu'à** **t**.*

En gros, tout cela signifie que le risque ne change pas avec le temps, il n'y a pas d'usure, pas d'amélioration. Les évènements surviennent au hasard.

Ainsi, les évènements sont régis selon une loi de Poisson et les délais entre 2 évènements par une loi exponentielle. De plus, on peut voir la fonction de survie comme une fonction de répartition comme étant le complément de la fonction répartition qui s'occupe/identifie les décès. On regarde d'abord les décès puis on va s'intéresser à la survie.

b. La fonction de survie

En épidémiologie clinique, la **durée résiduelle de vie d'un patient**, à compter de l'instant de référence (date d'origine), est une caractéristique variable d'un patient à l'autre ; c'est donc une **variable aléatoire**, que nous noterons **T**. La probabilité pour que le décès (« la défaillance ») intervienne après un délai supérieur à t est donc la probabilité pour que T soit supérieur à t :

$$S(t) = \Pr(T > t) = 1 - F(t)$$

Avec :

- F : fonction de répartition de la durée de vie résiduelle (proportion de patients décédés au temps t).

En épidémiologie clinique, la fonction de survie est donc une fonction de répartition. On la note également S(t). Elle représente :

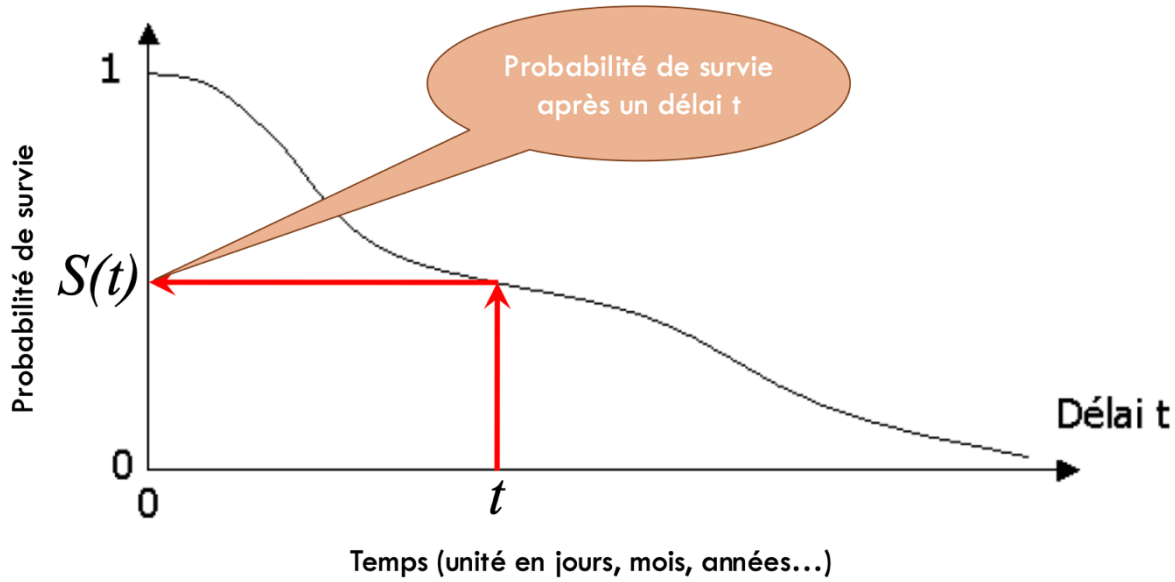
- La probabilité pour qu'un patient soit **encore vivant après un délai t**
- Ou encore la **proportion « vraie » des survivants après un délai t**

La fonction de survie, S(t), est la **probabilité que l'évènement d'intérêt ne survienne pas avant la date t.**

- $S(t) = Pr$ (délai de survenue de l'évènement d'intérêt à compter de l'instant de référence > t).

- Si l'évènement d'intérêt est le **décès**, c'est la probabilité de survivre au moins jusqu'à la date t.
- Si l'évènement d'intérêt est la **récurrence** de symptômes après traitement, c'est la probabilité de survivre sans symptômes jusqu'à la date t (on parle alors de *disease free survival*).

La fonction de survie est représentée graphiquement par une courbe de survie :



Remarque : *Face à ce graphique, on peut se demander quelle est la probabilité de survie au bout d'un certain temps mais aussi se demander au bout de combien de temps un certain nombre de patients sont décédés.*

La fonction de survie permet de **calculer la probabilité pour que le décès survienne après** un délai t1 et avant le délai t2 (avec t2 > t1). Il s'agit de calculer : $\Pr(T \in]t1; t2])$.

Or, $\Pr(T \in]t1; t2]) = F(t2) - F(t1) = S(t1) - S(t2)$

La fonction de survie donne aussi une information essentielle pour la suite : la **probabilité de survire encore après un délai t sachant que l'on est survivant après un** délai τ ($\tau < t$), que l'on notera S(t/ τ). On a :

$$S(t) = Pr(X > t) \quad \text{et} \quad S(\tau) = Pr(X > \tau)$$

$$t = \tau + s \quad \text{avec} \quad s > 0$$

Or nous avons l'égalité d'événements suivante :

$$\{X > \tau + s\} \cap \{X > \tau\} = \{X > \tau + s\}$$

En appliquant la formule des probabilités composées, il vient aisément que :

$$S(t/_{\tau}) = \frac{Pr((X > t) \cap (X > \tau))}{Pr(X > \tau)} = \frac{Pr((X > \tau + s) \cap (X > \tau))}{Pr(X > \tau)} = \frac{Pr(X > \tau + s)}{Pr(X > \tau)} = \frac{S(\tau + s)}{S(\tau)}$$

Finalement, on a :

$$S\left(\frac{t}{\tau}\right) = \frac{S(t)}{S(\tau)}$$

Sachez que la démonstration n'est pas à retenir, pas de stress !!! La seule formule à connaitre ici est la dernière qui est encadrée en bleu ! Qu'est-ce qu'on retient principalement dans tout ça ?

- ⇨ *La loi exponentielle sert à modéliser une survie quand : le taux de risque est constant dans le temps. C'est l'hypothèse essentielle.*

- ⇨ *Retenez cette formule : $S(t) = e^{-\lambda t}$. C'est la **fonction de survie**. Elle donne la probabilité de survivre au moins jusqu'au temps t.*

- ⇨ *Ce qui peut être utile : $F(t) = 1 - S(t)$. Donc :*
  - *$F(t)$ = probabilité que l'événement arrive avant t*
  - *$S(t)$ = probabilité de survivre au-delà de t*

- ⇨ *Le prof insiste surtout ici sur cette probabilité conditionnelle :*

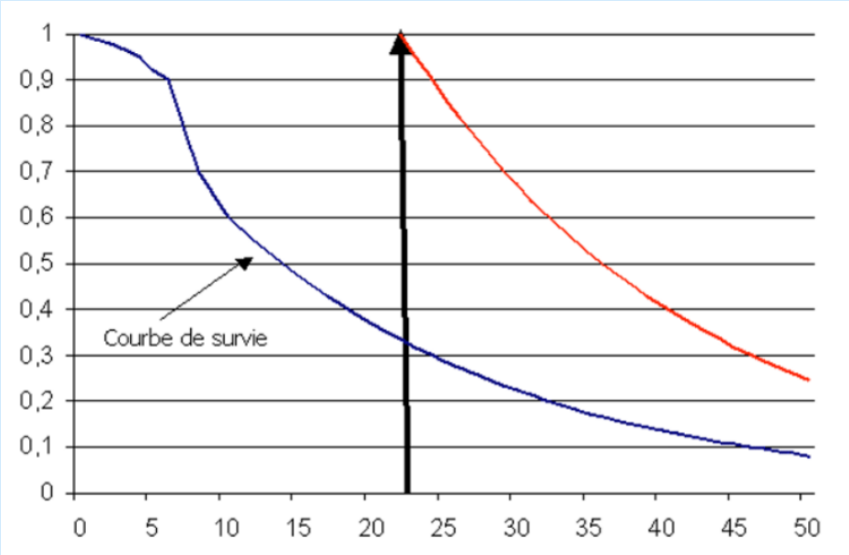
$$S(t \mid \tau) = \frac{S(t)}{S(\tau)}$$

C'est la probabilité de survivre jusqu'à t sachant qu'on est vivant à τ.

Allez, on respire et on continue !

Exemple :

Supposons que l'on veuille calculer la probabilité de survivre après (un délai de) t = 33 ans sachant que l'on est vivant à t = 23 ans.



À la lecture de la courbe de survie, on remarque qu'il y a 33% de survivants à 23 ans. On lit également que à 33 ans, la proportion de survivants de la population initiale est de 20%.

Mais, ne nous intéressant qu'aux survivants à 23 ans, ces 20% représentent 0,2/0,33 de la population d'intérêt, c'est-à-dire $\frac{S(33\ ans)}{S(23\ ans)}$.

IV. Estimation de la survie

Ok, partie quand même importante ! On reste concentré !

a. Recueil des données

Bon quelques petits rappels !

La **date d'origine** correspond à la date à laquelle a débuté l'observation.

Exemple : Date de diagnostic du cancer broncho-pulmonaire. Cette date doit avoir un sens clinique, afin que la « survie » analysée puisse être interprétée facilement par les lecteurs.

La **date des dernières nouvelles** correspond à la date de décès pour les patients décédés ou la date à laquelle on dispose des dernières données relatives à l'état du patient qu'il n'est pas décédé.

La **date de point** est la date à laquelle on fait le point ou la date de fin d'observation.
Tout patient chez qui l'évènement d'intérêt n'a pas été observé à la date de point est censuré à cette date. Un sujet perdu de vue à la date de point sera censuré à la date de des dernières nouvelles.

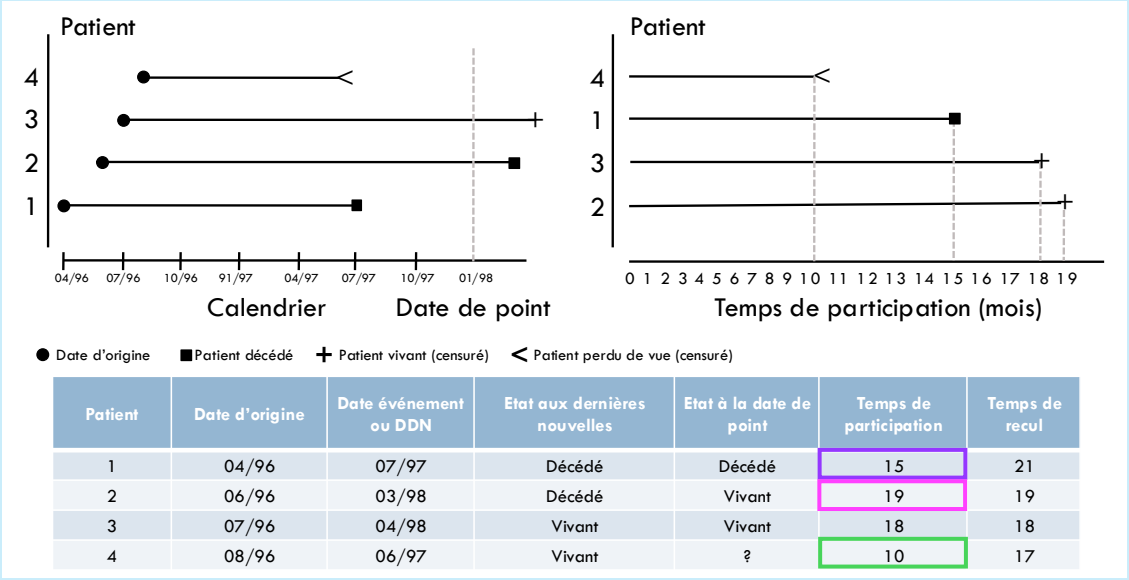
Un **évènement « en tout ou rien »** (binaire) correspond à l'état du patient en deux éventualités (vivant ou décédé) à la date des dernières nouvelles. Tout évènement binaire autre que le décès à un délai de survenue peut être analysé en délai de survie. ***Exemple*** : On peut étudier la survenue de la rechute ou de la récidive tumorale après un traitement ou la survenue de métastases.

b. Calcul des durées de suivi

À partir de ces données, les **durées de suivi** (ou temps de participation) de chaque patient sont calculées par différence. Elles correspondent au délai entre la **date d'origine** et la **date des dernières nouvelles** qui sera :

- La date de décès en cas de décès
- La date de point pour les patients vivants pour lesquels le suivi est assuré
- Ou la date de perte de vue pour les patients vivants n'étant plus suivi dans la cohorte à la date de point.

Exemple :



Patient	Date d'origine	Date événement ou DDN	Etat aux dernières nouvelles	Etat à la date de point	Temps de participation	Temps de recul
1	04/96	07/97	Décédé	Décédé	15	21
2	06/96	03/98	Décédé	Vivant	19	19
3	07/96	04/98	Vivant	Vivant	18	18
4	08/96	06/97	Vivant	?	10	17

🦋 *À gauche, on a un calendrier grégorien (durées absolues) et à droite un calendrier statisticien (durées relatives) où on regarde le temps de participation. À partir de leurs données, on peut construire le tableau ci-dessus → DDN = date des dernières nouvelles*

Ok décortiquons ensemble le tableau de valeurs mes poussins !

- ⇨ *Concernant le **temps de participation :***
- ⇨ *La "date des dernières nouvelles" dépend de la situation du patient.*

Temps de participation = Date des dernières nouvelles – Date d'origine

On peut avoir 3 situations possibles :

- ⇨ ***Patient décédé pendant l'étude***

Temps de participation = Date de décès – Date d'origine

Exemple patient 1 : $04/96 \rightarrow 07/97 = \mathbf{15}\ \boldsymbol{mois}$ *entre les deux dates !*

- ⇨ ***Patient vivant à la date de point***

*Temps de participation = **Date de point – Date d'origine***

Même s'il décède après, on ne regarde pas.

Exemple patient 2 : $06/96 \rightarrow 03/98 = \mathbf{19}\ \boldsymbol{mois}$ *entre les deux dates !*

Il est vivant à la date de point → censuré.

- ⇨ ***Patient perdu de vue***

*Temps de participation = **Date des dernières nouvelles – Date d'origine***

On s'arrête à la dernière info connue.

Exemple patient 4 : $08/96 \rightarrow 06/97 = \mathbf{10}\ \boldsymbol{mois}$ *entre les deux dates !*

Cependant, faites attention à ne pas confondre avec le temps de recul

Dans le tableau du prof (colonne de droite) :

- ***Temps de recul** = Date de point – Date d'origine*
- ***Temps de participation** = durée réellement utilisée dans l'analyse*

Ce n'est pas toujours la même chose !

c. Calcul de la survie

Si aucune variable n'est censurée, la fonction de survie se calcule par le pourcentage de survivants en fonction du temps. Cependant, cela ne se produit jamais, car un certain nombre de sujets seront perdus de vue, et un certain nombre seront encore vivants à la date de point.

Deux méthodes d'analyse de survie sont de préférence utilisées : **l'analyse actuarielle** et la **méthode de Kaplan-Meïer**, qui sont deux méthodes non paramétriques (*non-parametric ou distribution-free*), puisqu'elles **ne nécessitent aucune hypothèse sur la distribution des temps de survie**.

- L'analyse actuarielle est moins utilisée que la méthode de Kaplan-Meïer, et s'applique principalement lorsqu'il y a un **grand nombre de sujets ++** (plus de 200 par groupe) et de nombreux évènements.

- La méthode de Kaplan-Meïer est donc la méthode de choix pour les échantillons de **taille plus réduite ++**.

Ces deux méthodes supposent une **hypothèse forte** : les probabilités de survie sont supposées indépendantes du calendrier.

Exemple :

On suppose que la survie à 1 an d'un groupe de patients inclus en 1970 est identique à celle d'un groupe de patients inclus en 1990. Cette hypothèse n'est pas forcément vérifiée pour les études disposant d'un recul maximum très important, notamment en raison des progrès thérapeutiques vis-à-vis de la maladie étudiée. Elles partent du principe qu'il n'y a pas de progrès thérapeutique tout au long de l'étude.

La **fonction de survie** estimée peut être résumée soit par le taux de survie à un délai fixé (1 an, 5 ans, …) ; soit par une valeur de durée : médiane de survie (*median survival time*) et quantiles (*percentiles*).

d. Analyse actuarielle

La fonction de survie est calculée sur **des intervalles de temps fixés à priori** **+++** (mois, trimestre, semestre, année, …). Schématiquement, le mode de calcul est le suivant. Pour chaque intervalle de temps (par exemple $[0{,}1\ an]$, $[1{,}2\ ans]$, …), on définit :

- Le nombre de sujets vivants au début de l'intervalle : **V**
- Le nombre de sujets décédés dans l'intervalle : **D**
- Le nombre de sujets vivants aux dernières nouvelles, dont le temps de participation s'arrête dans l'intervalle (censure) : **C**

L'hypothèse actuarielle (*actuarial assumption*) suppose que ces sujets sont exposés au risque d'évènement sur la moitié de l'intervalle (*6 mois dans notre exemple*).

- Nombre de sujets exposés au risque d'évènement (ex : décès) sur l'intervalle est :

$$N = V - \left(\frac{C}{2}\right)$$

- La probabilité d'évènements durant l'intervalle est simplement estimée par le rapport du nombre d'évènements sur le nombre de sujets à risque :

$$Probabilité = \frac{D}{N}$$

- Survie sur l'intervalle est :

$$Survie = \frac{(N - D)}{N}$$

Cette probabilité est appelée **survie instantanée**.

La fonction de survie est obtenue en faisant le produit des survies instantanées sur l'ensemble des intervalles.

Exemple :

La survie à 3 ans = (survie instantanée entre 2 et 3 ans) x (survie instantanée entre 1 et 2 ans) x (survie instantanée entre 0 et 1 an).

Instants	V	C	D	N = V – C/2	(N - D) / N	S(t)
0	-	-	-	-	-	1
3	210	0	0	210	**1**	1x1 = **1**
6	210	10	40	210 - 5 = 205	(205-40)/205 = **0,805**	0,805 x 1 = **0,805**
9	160	30	10	160 - 15 = 145	(145-10)/145 = **0,931**	0,931 x 0,805 = **0,749**
12	120	10	20	120 – 5 = 115	(115-20)/115 = **0,826**	0,826 x 0,749 = **0,619**
15	90	20	0	90 – 10 = 80	**1**	1 x 0,619 = **0,619**
18	70	0	20	70	(70-20)/70 = **0,714**	0,714 x 0,619 = **0,442**
21	50	18	3	50 – 9 = 41	(41-3)/41 = **0,927**	0,927 x 0,442 = **0,410**
24	29	8	2	29 – 4 = 25	(25-2)/25 = **0,920**	0,920 x 0,410 = **0,377**

On rappelle que :

- V : nombre de sujets vivants au début de l'intervalle
- C : nombre de sujets vivants censurés dans l'intervalle
- D : nombre de sujets décédés dans l'intervalle
- N : nombre de sujets exposés au risque de décès

> Remarque : *Ce tableau s'appelle une **table de mortalité** si l'évènement d'intérêt est la mort.*

La première colonne du tableau est une échelle de temps (0, 3, 6, …) qui va permettre de faire la différence (en partie) entre la méthode actuarielle où les intervalles sont toujours les mêmes (réguliers) et la méthode de Kaplan-Meïer qui va prendre comme intervalle les moments où les évènements surviennent.

> **Remarque** : *À 6 mois, on a 210 vivants, il y a eu 40 décès dans l'intervalle et 10 sujets sont censurés.*

Pour calculer la probabilité de survie, il faut connaitre le dénominateur des survies instantanées (N). À 6 mois N=205, cela va permettre de calculer le taux de survie instantanée qui va être de 0,805. Enfin, on fait le produit des survies instantanées sur les intervalles précédents. C'est toujours la même mécanique, être de vivant à 6 mois c'est aussi être vivant à 3 mois donc on fait le produit de la survie instantanée à 6 mois et de la probabilité de survie à 3 mois : 0,805 x 1 (*cf. flèches*).

Comment on lit ce gros tableau ?

On découpe le temps en intervalles fixes (ici 0–3, 3–6, 6–9 mois…). Pour chaque intervalle, on calcule une survie instantanée, puis on multiplie tout pour obtenir S(t).

⇨ ***Colonne 1 : V = vivants au début de l'intervalle***

Exemple à 6 mois : V = 210 → 210 personnes vivantes au début de l'intervalle 6–9.

⇨ ***Colonne 2 : C = censurés dans l'intervalle***

À 6 mois : C = 10 → 10 personnes quittent l'étude (perdus de vue ou vivants à la date de point). En actuariel, on suppose qu'ils sont exposés en moyenne ***la moitié de l'intervalle****.*

⇨ ***Colonne 3 : D = décès dans l'intervalle***

À 6 mois : D = 40 → 40 décès entre 6 et 9 mois.

⇨ ***Colonne 4 : N = sujets réellement exposés au risque***

Formule actuarielle :

$$N = V - \frac{C}{2}$$

Pourquoi C/2 ? Parce qu'on suppose que les censurés sont exposés en moyenne la moitié du temps. Exemple à 6 mois :

$$N = 210 - \frac{10}{2} = 210 - 5 = 205$$

⇨ ***Colonne 5 : Survie instantanée de l'intervalle***

$$\frac{N - D}{N}$$

- *À 6 mois :*

$$\frac{205 - 40}{205} = 0{,}805$$

Cela signifie que 80,5 % survivent pendant cet intervalle.

⇨ ***Colonne 6 : S(t) = survie cumulée***

On multiplie les survies instantanées successives.

- *À 6 mois :*

$$S(6) = 1 \times 0{,}805 = 0{,}805$$

- *À 9 mois :*

$$S(9) = 0{,}931 \times 0{,}805 = 0{,}749$$

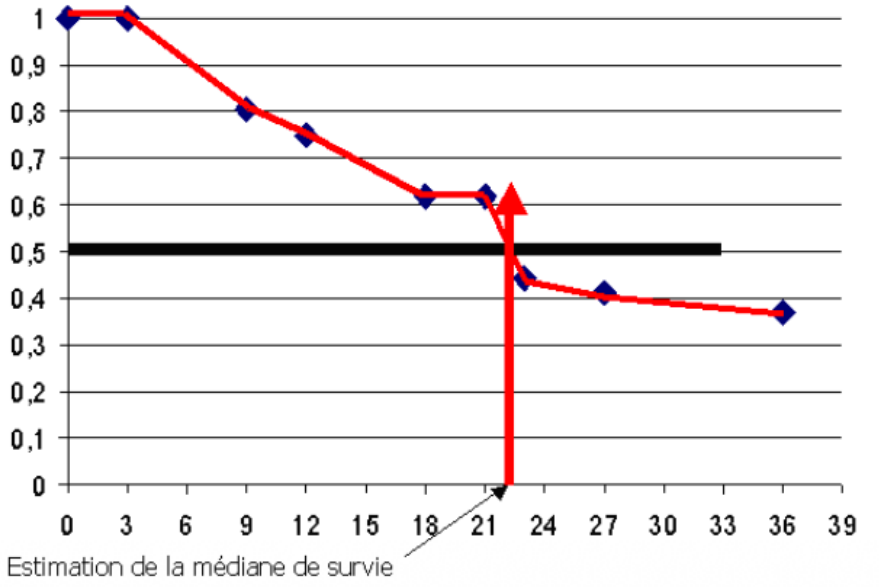
Donc : 74,9 % sont encore vivants à 9 mois.

Ce que vous devez vraiment comprendre est que chaque intervalle donne :

- *Combien étaient au départ (V)*
- *Combien sortent (C)*
- *Combien meurent (D)*
- *On ajuste avec C/2*
- *On calcule une survie locale*
- *On multiplie pour obtenir la survie cumulée*

Est-ce que ça va ? 😊

Pour chaque intervalle de temps, on représente **l'estimation de la survie S(t)** par un point. Les coordonnées du premier point sont 0 (à j0) en abscisse, et 1 (100 %) en ordonnée. Tous les points consécutifs sont reliés par un segment de droite.



L'inconvénient majeur de cette méthode est qu'elle estime la survie à chaque borne supérieure des intervalles constitués a priori, et considère chaque censure, survenant dans un intervalle, de manière équivalente, c'est-à-dire qu'un sujet suivi pendant 21 jours apporte la même information qu'un sujet suivi pendant 29 jours pour la survie à 30 jours dans l'exemple présenté. C'est la raison pour laquelle cette méthode est à réserver à de **grands échantillons**.

e. Méthode de Kaplan-Meïer

Contrairement à l'analyse actuarielle, les intervalles **ne sont pas fixés à priori +++,** mais sont définis par les instants auxquels les évènements sont observés (ex : on change d'intervalle à chaque décès). Ces intervalles sont donc **inégaux**, débutant à l'instant d'un évènement et s'arrêtent juste avant l'évènement suivant.

Pour chaque intervalle entre deux évènements, on définit V, D et C comme précédemment (avec la particularité que D vaut souvent 1, sauf dans le cas où plusieurs évènements surviennent au même temps de participation).

Dans l'analyse de **Kaplan-Meïer** : $N = V - C$

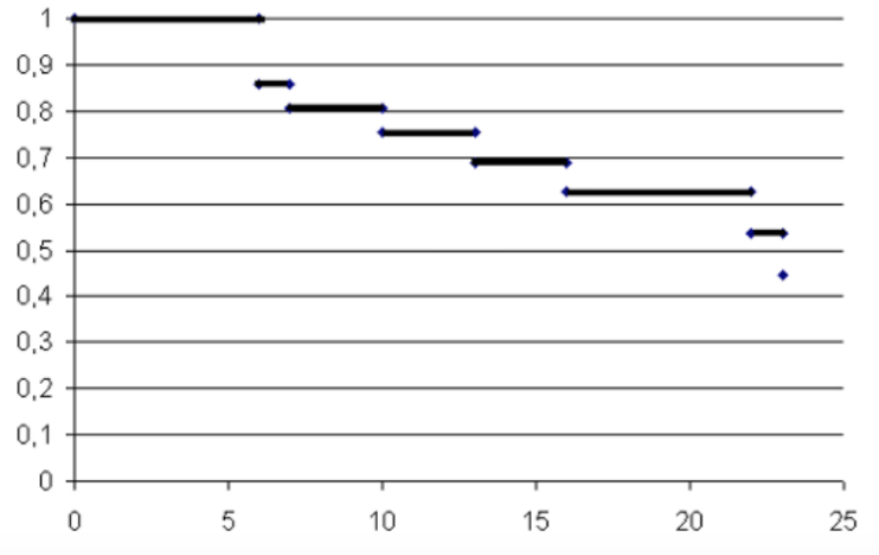
La probabilité de survie instantanée calculée sur cet intervalle vaut : $\frac{(N-D)}{N}$

L'estimation de Kaplan-Meïer de la fonction de survie s'obtient, comme dans l'analyse actuarielle, en faisant le produit des survies instantanées.

Instants	V	C	D	N = V - C	(N - D) / N	S(t)
0	21	-	-	-	-	1
6	21	0	3	21	**0,857**	**0,857**
7	18	1	1	17	**0,941**	**0,807**
10	16	1	1	15	**0,933**	**0,753**
13	14	2	1	12	**0,917**	**0,690**
16	11	0	1	11	**0,909**	**0,627**
22	10	3	1	7	**0,857**	**0,537**
23	6	0	1	6	**0,833**	**0,448**

- *V, C, D et N comme tout à l'heure*.

La courbe de survie se compose de **paliers successifs**, où les probabilités de survie sont constantes entre deux temps d'événements consécutifs. Le **premier palier vaut 1** depuis l'origine jusqu'au délai de survenue du premier événement. Il s'abaisse ensuite à la première valeur calculée pour constituer un second palier jusqu'au délai de survenue de l'événement suivant, etc. Il est possible de relier les paliers successifs par des segments verticaux, mais il n'est pas correct de les relier par des segments obliques. La courbe ainsi obtenue présente une **allure en « marches d'escalier »**.



Tut'Récap

⇨ Les intervalles de temps

○ Méthode actuarielle

Imagine un **immeuble avec des étages fixes**.

- 0–3 mois
- 3–6 mois
- 6–9 mois

Peu importe ce qu'il se passe dedans. Les étages sont **définis à l'avance**.

Le temps est découpé en blocs réguliers.

○ Kaplan-Meier

Imagine un escalier qui descend à chaque décès.

Les marches apparaissent quand un événement survient. Les intervalles ne sont **PAS** fixes. Ils dépendent du moment des décès.

⇨ Comment on traite les censures ?

○ Actuarielle

On dit : "Bon… on ne sait pas exactement quand ils sont partis…

On va dire qu'ils ont été là en moyenne la moitié du temps." Donc :

$$N = V - C/2$$

Les censurés comptent **à moitié**.

○ Kaplan-Meier

Ici on connaît les temps exacts. Donc :

$$N = V - C$$

Les censurés comptent **entièrement jusqu'au moment où ils partent**.

⇨ L'allure des courbes

○ Actuarielle

Courbe **lissée**, segments obliques. Imagine un **toboggan**.

- **Kaplan-Meier**

Courbe en **marches d'escalier**. Imagine des marches nettes qui descendent à chaque décès. **On ne relie PAS les marches en diagonale**.

En gros !

- Actuarielle = immeuble régulier + censures à moitié + toboggan
- Kaplan-Meier = escalier irrégulier + censures exactes + marches nettes

Sinon, autre petite astuce :

- **Kaplan-Meïer** = **K** comme **K**rash → La courbe descend **pile au moment du crash (décès)**.
- **Actuarielle** = **Actuarielle = Assurance** → On travaille par **périodes fixes** comme les contrats d'assurance.

J'espère que ça vous aide un peu !

f. Choix d'une valeur résumée

1. Médiane de survie

La courbe de survie apporte des renseignements importants, mais il est utile de disposer d'indicateurs synthétiques ou résumés de cette courbe. La **moyenne de survie** n'est pas un bon indicateur, pour des raisons d'ordre statistique, notamment liées à l'existence de censures.

La **médiane de survie** lui est préférée. Elle représente la **durée t** pour laquelle la probabilité de survie **S(t) est de 50 %.** À cause de la distribution par paliers de la fonction de survie, il est souvent impossible de connaitre la durée correspondant à une survie exacte de 50 %. En pratique, la médiane est estimée par la plus petite durée pour laquelle la survie est inférieure à 50 %.

Il arrive que la fonction de survie soit toujours supérieure à 50 %. Dans ce cas, la médiane ne peut être estimée. On estime alors les **quantiles** (NB : un quartile = 25 %) : pour le p-ième quantile on estime la durée pour laquelle la probabilité de survie est de 100-p. **Par exemple**, le 25e quantile (ou 1er quartile) correspond à la plus petite durée pour laquelle la survie est inférieure à 75 %.

2. Survie à date fixée

Un autre indicateur fréquemment utilisé pour résumer l'information d'une courbe de survie est **l'estimation de la survie à un temps donné** (survie à 5 ans par exemple...)

Je sais que ça commence à faire long, tenez bon, c'est bientôt terminé !

V. Comparaison de deux fonctions de survie

a. Contexte

Il arrive fréquemment que l'on souhaite montrer qu'une action (intervention, traitement) ou une classification ont un lien avec la survie. Il s'agira de conduire une étude comparative et de mettre en œuvre un test d'hypothèses.

Le principe du **test du log-rank** (*ou test de Mantel-Cox ou de Peto-Mantel- Haenszel*) est de comparer, dans chaque groupe, le nombre observé et le nombre attendu d'événements si la survie était identique dans les deux groupes, sur l'ensemble de la période étudiée.

Une erreur importante, et souvent retrouvée, consiste à assimiler l'efficacité du traitement à la réponse des patients à ce traitement, et à comparer la survie, non plus entre les patients traités et les patients non traités, mais entre les sujets qui répondent au traitement et les sujets qui ne répondent pas (*comparison of survival by response*).

Cette méthode est à proscrire et peut provoquer des biais et des conclusions fausses :

- Les sujets répondeurs sont en général en meilleure santé que les sujets non répondeurs et sont donc susceptibles - indépendamment de tout traitement - de vivre plus longtemps ;
- La comparaison de la survie par la réponse au traitement peut être biaisée puisque les patients doivent vivre suffisamment longtemps pour avoir la possibilité de répondre au traitement (*guarantee-time bias*)

b. Principe du test log-rank

Pour chaque intervalle de temps (qu'il s'agisse de l'analyse actuarielle ou de Kaplan-Meier), le nombre attendu d'événements, sous l'hypothèse nulle d'égalité de la survie entre les deux groupes, s'obtient en appliquant, au nombre de sujets exposés au risque d'événements, la proportion d'événements observés sur l'ensemble des deux groupes.

Le test du log-rank, évaluant l'écart entre le **nombre observé** et le **nombre attendu d'événements** sur les deux groupes, est un Khi2 à 1 degré de liberté (ddl). Ce test est généralisable au cas de k groupes et permet de tester si globalement la survie est différente entre les groupes.

On pose deux hypothèses :

- **H0** : les fonctions de survie sont les mêmes dans les deux populations d'où sont issus les groupes A et B ➔ $S_A(t) = S_B(t)$.
- **H1** : les deux fonctions de survie diffèrent.

Exemple :

Imaginons que l'on souhaite faire la preuve qu'un traitement adjuvant à la chirurgie dans le carcinome hépatocellulaire améliore la survie des patients. Les grands traits de l'étude sont les suivants :

⇨ La survie sera comptée à partir de la date de la chirurgie.

- ⇨ Des patients ont été inclus pendant une année dans une étude qui a duré 3 ans et répartis par tirage au sort dans un des deux groupes de traitement : chirurgie seule (groupe A) ou chirurgie + traitement adjuvant (groupe B).
- ⇨ La durée de suivi des patients (durée de participation à l'étude ou recul) varie d'un patient à l'autre

A la fin de l'étude on dispose pour chaque patient :

- ⇨ Du groupe auquel il a appartenu, A ou B
- ⇨ Des temps de suivi pour chaque patient selon son groupe et selon le fait que le patient soit décédé ou bien que le patient soit censuré, qu'il soit encore vivant ou perdu de vue.

Supposons que l'on dispose des observations suivantes :

- Dans le groupe A,les tAi et tAi*sont : 1;1;2;2;3;4;4;5;5;8;8;8;8;11; 11; 12; 12; 15; 17; 22; 23
- Dans le groupe B, les tBi et tBi* sont : 6; 6; 6; 6,1*; 7; 9*; 10; 10,1*; 11,2*; 13; 16; 17,3*; 19*; 20*; 22; 23; 25*; 32*; 32*; 34*; 35*

Les ensembles des tAi et tBi (patients décédés) constituent l'ensemble des temps de décès observés, quel que soit le groupe ; on les notera ti et on les considérera ordonnés par valeurs croissantes. Ici les ti sont : 1; 2; 3; 4; 5; 6; 7; 8; 10; 11; 12; 13; 15; 16; 17; 22; 23

🦋 *À quoi sert ces exemples ? Je comprends pas ? Bouge pas, je t'explique !*

On cherche à construire une liste des temps de décès communs aux deux groupes ! On a =

- *Groupe A : liste de temps (avec * pour censure)*
- *Groupe B : liste de temps (avec * pour censure)*

Mais pour comparer les survies, on ne travaille PAS séparément.

On doit :

- ⇨ *Prendre uniquement les **temps de décès réels***
- ⇨ *Les mélanger (A + B)*
- ⇨ *Les classer par ordre croissant*

Ça donne la liste des t_i. Exemple dans la diapo :

$$t_i = 1; 2; 3; 4; 5; 6; 7; 8; 10; 11; 12; 13; 15; 16; 17; 22; 23$$

*C'est indispensable parce que le **test du log-rank** fonctionne comme ça :*

À chaque temps de décès t_i :

- *On regarde combien de personnes sont encore à risque dans chaque groupe*
- *On compare le nombre de décès observés au nombre attendu*

*Donc on doit connaître **les temps communs d'événements**.*

Imaginez deux escaliers (courbes Kaplan-Meïer) :

- *Groupe A*
- *Groupe B*

*Le test du log-rank va comparer les deux escaliers **marche par marche**. Mais pour ça, il faut savoir à quels moments exacts on regarde. Ces moments sont les t_i.*

C'est plus clair ?

c. Estimation des décès

Le principe est d'abord d'estimer, tout groupe confondu, **la probabilité de décéder à t_i sachant que l'on est vivant à t_{i-1}**, c'est-à-dire estimer $1 - S(t_i - t_{i-1}))$ et ceci pour chacun des temps de décès observés t_i.

On utilise ici l'estimateur de Kaplan-Meïer de $S(t_i - t_{i-1})$.

On obtient ainsi la dernière colonne du tableau ci-après.

t_i	V	C	N = V - C	D	$S(t_i / t_{i-1}) = (N - D) / N$	$1 - S(t_i / t_{i-1})$
1	42		42	2	0,952	0,048
2	40		40	2	0,950	0,050
3	38		38	1	0,974	0,026
4	37		37	2	0,946	0,054
5	35		35	2	0,943	0,057
6	33		33	3	0,909	0,091
7	30	1	29	1	0,966	0,034
8	28		28	4	0,857	0,143
10	24	1	23	1	0,957	0,043
11	22	1	21	2	0,905	0,095
12	19	1	18	2	0,889	0,111
13	16		16	1	0,938	0,062
15	15		15	1	0,933	0,067
16	14		14	1	0,929	0,071
17	13		13	1	0,923	0,077
22	12	3	9	2	0,778	0,222
23	7		7	2	0,714	0,286

d. Calcul des décès attendus

On estime ensuite le nombre de décès que l'on attend dans chacun des groupes A et B, à chaque t_i, en supposant que la probabilité conditionnelle de décès estimée s'applique identiquement à chacun des deux groupes. Pour cela on évalue à chaque ti l'effectif à risque à cette date. On obtient les deux dernières colonnes du tableau suivant.

t_i	V	C	N = V - C	D	$S(t_i / t_{i-1}) = (N - D) / N$	$1 - S(t_i / t_{i-1})$	N_A	N_B	E_A	E_B
1	42		42	2	0,952	0,048	21	21	1,000	1,000
2	40		40	2	0,950	0,050 x 19		21	0,950	1,050
3	38		38	1	0,974	0,026	17	21	0,447	0,553
4	37		37	2	0,946	0,054 x 16		21	0,864	1,136
5	35		35	2	0,943	0,057	14	21	0,799	1,201
6	33		33	3	0,909	0,091	12	21	1,092	1,988
7	30	1	29	1	0,966	0,034	12	17	0,408	0,579
8	28		28	4	0,857	0,143	12	16	1,714	2,286
10	24	1	23	1	0,957	0,043	8	15	0,344	0,656
11	22	1	21	2	0,905	0,095	8	13	0,760	1,240
12	19	1	18	2	0,889	0,111	6	12	0,666	1,334
13	16		16	1	0,938	0,062	4	12	0,249	0,751

Ces nombres sont notés **EAi** et **EBi**. On remarque que l'on utilise ici, comme toujours, la justesse supposée de l'hypothèse nulle (H0) puisque les probabilités de décès, et donc la survie, sont supposées ne pas dépendre du groupe.

Sous l'hypothèse nulle (H0) ces nombres doivent être voisins des nombres de décès réellement observés. En particulier le total de ces nombres de décès au cours du temps (noté E_A et E_B selon le groupe) doit être voisin du nombre total de décès observés (noté DA et DB selon le groupe), et ceci dans chacun des groupes.

Dans l'exemple, on obtient : EA=10,74 ; EB=19,26 ; DA=21 ; DB=9.

Rembobinons pour éclaircir certaines choses !

⇨ ***D : il vient d'où ?***

Dans la table Kaplan-Meier (colonne D) : D_i = nombre total de décès au temps t_i

- *Di = décès **tous groupes confondus** au temps t_i.*
- *D est déjà donné par la liste des temps de décès !*

⇨ ***N_A et N_B : ils viennent d'où ?***

À chaque temps t_i, on compte :

N_A = *nombre de sujets du groupe A encore à risque juste avant t_i*

N_B = *nombre de sujets du groupe B encore à risque juste avant t_i*

Ça signifie :

- *Ils sont vivants*
- *Ils ne sont pas encore censurés*
- *Ils n'ont pas encore eu l'événement*

Ce sont les effectifs à risque dans chaque groupe. Donc eux aussi sont déjà connus à ce stade.

⇨ ***Donc E_A et E_B ne sont PAS donnés***

Ce sont eux qu'on calcule ! Sous l'hypothèse nulle (même survie), la probabilité de décès au temps t_i est :

$$p_i = \frac{D_i}{N_A + N_B}$$

Puis :

$$E_{A,i} = p_i \times N_A$$

$$E_{B,i} = p_i \times N_B$$

Regardez les flèches colorées plus haut et vous comprendrez les calculs !

Élément	Donné ?	Vient d'où ?
D	Oui ! 😊	Comptage des décès au temps t_i
N_A	Oui ! 😊	Nombre à risque dans A
N_B	Oui ! 😊	Nombre à risque dans B
E_A	Non ! 😔	On le calcule
E_B	Non ! 😔	On le calcule

Dites-vous qu'à chaque marche de l'escalier :

- *On regarde combien de gens sont encore debout dans A et B*
- *On voit combien tombent (D)*
- *On répartit ces chutes proportionnellement*

Désolé, vous devez vous dire que la fiche ne finit pas mais j'essaye de vous expliquer au maximum car je sais que ce cours n'est pas très simple ! Mais tenez bon, on y est presque.

Quelques petites remarques ! :

- NA : effectif du groupe A au temps t avant le décès
- NB : effectif du groupe B au temps t avant le décès
- N : effectif global au temps t avant le décès (-les censurés) : N = NA + NB
- DA : nombre de décès observés dans le groupe A au temps t
- DB : nombre de décès observés dans le groupe B au temps t
- D : nombre de décès observés global au temps t : D = DA + DB
- EA : nombre de décès attendus dans le groupe A au temps t : EA = $D \times \frac{NA}{N}$
- EB : nombre de décès attendus dans le groupe B au temps t : EB = $D \times \frac{NB}{N}$

EA et EB impliquant que les fonctions de survie soient les mêmes pour les 2 groupes. (H0 acceptée = pas de différence entre les groupes A et B) : EA = EB = D.

e. Test du Khi2

Le paramètre du test est construit à partir de ces quatre valeurs/paramètres (aléatoires normalement à ce stade de la construction) :

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

Sous H0, Q suit une distribution de $\chi 2$ à un degré de liberté.

Condition de validité : EA et EB $\geq$ 5, et 1 ddl

On construit l'intervalle de pari de niveau 0,95 : $IP_{0,95}$ = [0 ; 3,84]

On met en place la règle de décision. Si la valeur calculée $Q_C \in$ [0 ; 3,84], on ne pourra conclure à une différence entre les fonctions de survie dans les deux populations considérées. Si la valeur Q_C excède 3,84 on conclura au risque de 5% que les fonctions de survie diffèrent.

Dans l'exemple traité, on obtient Q_C = 15,26. On rejette donc l'hypothèse d'égalité des fonctions de survie. La survie est meilleure dans le groupe dans lequel $D < E$, c'est le groupe B. La preuve est faite (au risque d'erreur de 5%) que le traitement adjuvant améliore la survie des patients à compter de la date de chirurgie.

OMG, ENFIN FINI ??? Passons juste aux exercices de son diapo !

VI. Exercices

Exercice 1 :

On a suivi le devenir d'un grand groupe de malades atteints d'une maladie M, à partir de la date de diagnostic. On considère alors que l'on dispose des probabilités suivantes : au bout d'un an, 20 % des malades sont morts ; au bout de 2 ans, 50 % des malades sont morts ; au bout de 3 ans, 70 % des malades sont morts ; au bout de 4 ans, 80 % des malades sont morts ; au bout de 5 ans, 80 % des malades sont morts. **La probabilité qu'un malade ayant déjà survécu 2 ans survive moins de 3 ans est :**

- A. 20 %
- B. 30 %
- C. 40 %
- D. 50 %
- E. 60 %

La réponse C est **vrai** !

On détaille un peu ? On cherche à savoir : Quelle est la probabilité qu'un malade ayant déjà survécu 2 ans meure avant 3 ans ? Donc ce n'est PAS une probabilité simple. C'est une **probabilité conditionnelle** :

$$P(2 < T < 3 \mid T > 2)$$

Premièrement, il faut transformer les données

On nous donne les pourcentages cumulés de décès :

- 1 an → 20 % morts → survie = 80 %
- 2 ans → 50 % morts → survie = 50 %
- 3 ans → 70 % morts → survie = 30 %

Donc : $S(2) = 0{,}5$ et $S(3) = 0{,}3$

Deuxièmement, il faut comprendre la logique :

On cherche : parmi ceux encore vivants à 2 ans, combien meurent entre 2 et 3 ans ?

On utilise la formule :

$$P(2 < T < 3 \mid T > 2) = \frac{S(2) - S(3)}{S(2)}$$

Troisièmement, application numérique

$$= \frac{0{,}5 - 0{,}3}{0{,}5} = \frac{0{,}2}{0{,}5} = 0{,}4$$

Donc : 40% → Réponse C. □ **C. 40 % : (1-0,2) x (1-0,5) = 0,8 x 0,5 = 0,4**

Ainsi, à 2 ans, 50 % sont encore vivants. À 3 ans, 30 % sont encore vivants.

Donc 20 % sont morts entre 2 et 3 ans. Mais ces 20 % sont à rapporter aux 50 % encore vivants à 2 ans. 20 / 50 = 40 %.

ATTENTION : On peut avoir tendance à répondre 20%. Mais 20 %, c'est la baisse absolue, pas la probabilité conditionnelle. Imaginez 100 patients :

- À 2 ans → 50 vivants.
- À 3 ans → 30 vivants.

Donc 20 des 50 sont morts entre 2 et 3 ans. 20 / 50 = 40 %.

⇨ **Exercice 2 :**

On s'intéresse à une population de femmes atteintes d'un cancer du sein. Le taux de survie 5 ans après la découverte du cancer est de 65 %. Lors de la découverte du cancer, on peut définir la gravité du cancer par son stade (1 à 4). 45 % des femmes sont de stade 1, 30 % de stade 2, 15 % de stade 3, et 10 % de stade 4. La probabilité qu'une femme de cette population soit de stade 4 et survive au moins 5 ans est 0,03. **Quel est le % de femmes de stade 4 qui décèdent dans les 5 ans après la découverte de leur cancer ?**

- □ A. 30%
- □ B. 50%
- □ C. 70%
- □ D. 90%
- □ E. Les propositions A, B, C, et D sont fausses

Qu'est-ce que l'on a comme données :

- $P(\text{survie} \geq 5 \text{ ans}) = 0{,}65$
- $P(\text{stade 4}) = 0{,}10$
- $P(\text{stade 4 ET survie} \geq 5 \text{ ans}) = 0{,}03$

On cherche le % **de femmes de stade 4 qui décèdent dans les 5 ans**, donc :

$$P(\text{décès} < 5 \text{ ans} \mid \text{stade 4})$$

Premièrement, il faut trouver la survie à 5 ans chez les stades 4 :

$$P(\text{survie} \geq 5 \text{ ans} \mid \text{stade 4}) = \frac{P(\text{stade 4 ET survie} \geq 5 \text{ ans})}{P(\text{stade 4})} = \frac{0{,}03}{0{,}10} = 0{,}30$$

Donc **30%** des stades 4 survivent au moins 5 ans.

Ensuite, on en déduit le % qui décèdent avant 5 ans

$$P(\text{décès} < 5 \text{ ans} \mid \text{stade 4}) = 1 - 0{,}30 = 0{,}70$$

Soit 70%. Donc la réponse **C est vraie.**

⇨ **Exercice 3 :**

On s'intéresse à une population de femmes atteintes d'un cancer du sein. Le taux de survie 5 ans après la découverte du cancer est de 65 %. Lors de la découverte du cancer, on peut définir la gravité du cancer par son stade (1 à 4). 45 % des femmes sont de stade 1, 30 % de stade 2, 15 % de stade 3, et 10 % de stade 4. La probabilité qu'une femme de cette population soit de stade 4 et survive au moins 5 ans est 0,03. **En cas de décès dans les 5 ans, quelle est la probabilité que la femme ait été de stade 4 ?**

- ☐ A. 20%
- ☐ B. 30%
- ☐ C. 50%
- ☐ D. 70%
- ☐ E. Les propositions A, B, C et D sont fausses

On demande : **En cas de décès dans les 5 ans**, quelle est la probabilité que la femme ait été **stade 4** ? Donc une proba conditionnelle :

$$P(\text{stade 4} \mid \text{décès} < 5 \text{ ans})$$

Premièrement, on doit trouver $P(\text{décès} < 5 \text{ ans})$

On sait : survie à 5 ans = 65%. Donc décès avant 5 ans = 35% → $P(D) = 1 - 0{,}65 = 0{,}35$

On doit trouver ensuite $P(\text{stade 4 ET décès} < 5 \text{ ans})$. On donne :

- $P(\text{stade 4}) = 0{,}10$
- $P(\text{stade 4 ET survie} \geq 5 \text{ ans}) = 0{,}03$

Donc, pour les stades 4, ceux qui meurent avant 5 ans :

$$P(\text{stade 4 ET décès}) = P(\text{stade 4}) - P(\text{stade 4 ET survie}) = 0{,}10 - 0{,}03 = 0{,}07$$

Troisièmement : Bayes (conditionnelle)

$$P(\text{stade 4} \mid D) = \frac{P(\text{stade 4 ET } D)}{P(D)} = \frac{0{,}07}{0{,}35} = 0{,}20$$

On obtient donc 20%, ainsi la réponse A est **vrai**.

C'est avec émotion que je clôture ma dernière fiche 🥹.

J'espère qu'elles vous auront toutes plu. Je m'excuse pour la longueur de celle-ci. J'ai essayé de détailler au maximum et de tout vous expliquer car je sais que ce cours est assez complexe ! J'espère que vous avez compris et si ce n'est pas le cas → GO FOFO. On fait des petites dédis quand même ?

- ⇨ Dédi à Marwa, votre super tut de kiné !
- ⇨ Dédi à Gev, votre tut gigiment goatesque
- ⇨ Dédi à Maude !
- ⇨ Anti-dédi au temps que j'ai pris pour faire cette fiche (même si elle est rempli d'amour)
- ⇨ Dédi à mes co-tuts qui sont des amours.
- ⇨ Re-dédi à Maïssou mon binôme de pharma, à Sirine superbe amie, Laure, Jeanne, etc.
- ⇨ Dédi à vous mes poussins, je suis tellement fière de vous sachez le 🥹