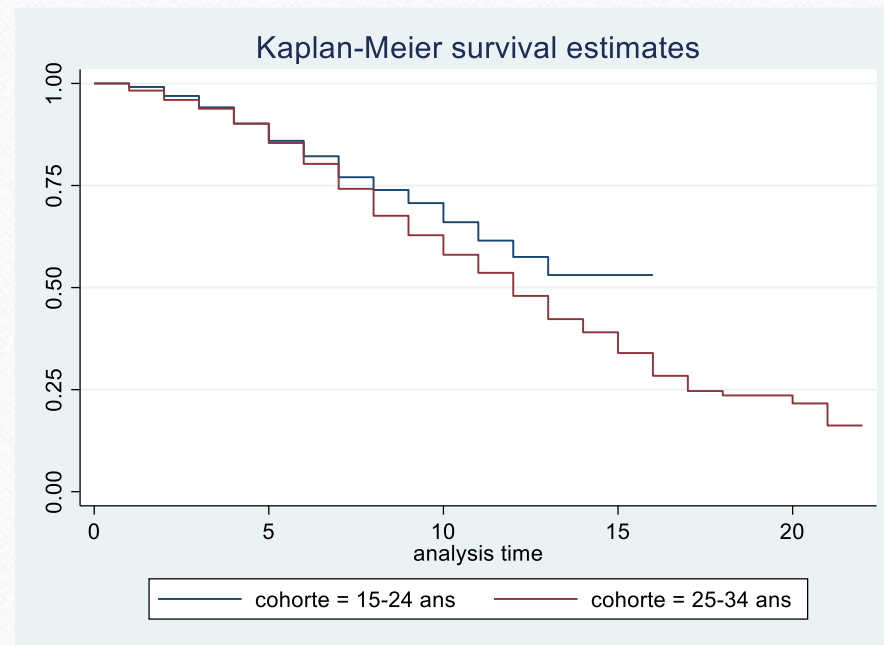
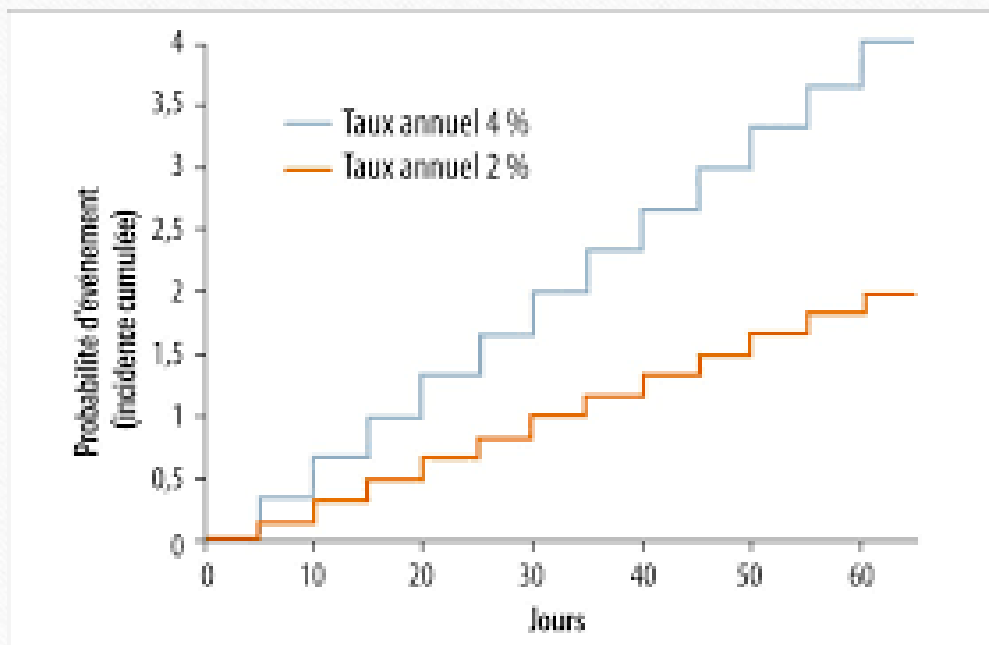
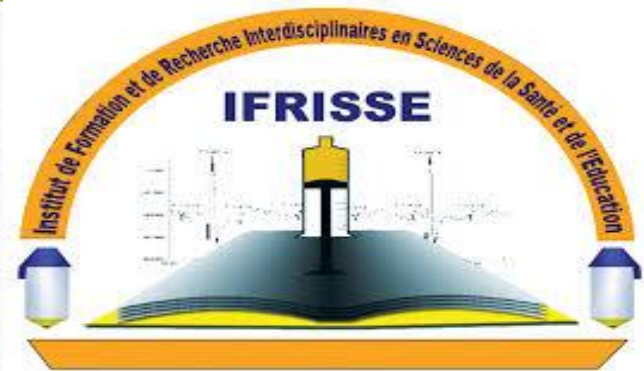


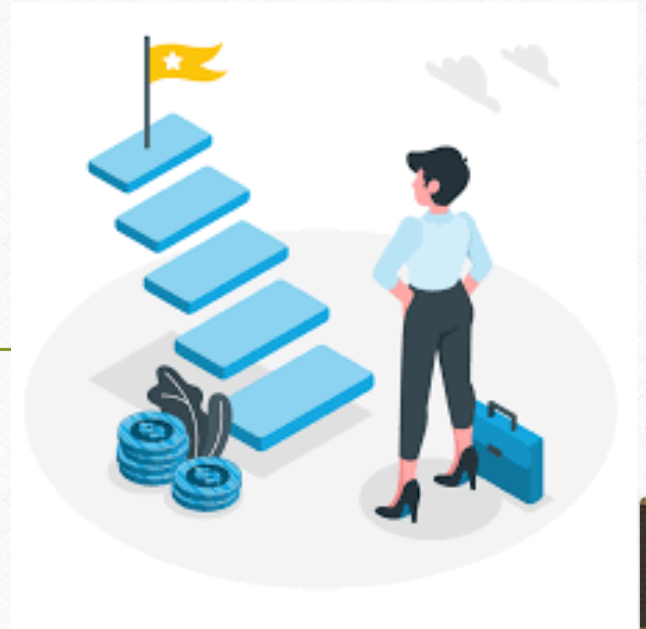
Introduction à l'Analyse de survie



Année académique_2021-2022

Objectifs

- Comprendre le principe d'analyse des données biographiques
- Décrire les étapes d'analyse des données biographiques



Plan de présentation

Généralités

Notions d'événement et de troncatures

Données biographiques

Méthode d'analyse descriptive: estimateurs de Kaplan-Meier

Méthode d'analyse explicative : régression semi-paramétrique de
Cox

Résumé des notions essentielles

Généralités

- Jusque là, nous connaissions analyser des données transversales (information obtenues à un moment donnée), sur une période de temps assez courte, souvent l'année civile.
- Dorénavant, sachons qu'il existe des données longitudinales qui sont des données relatives à l'évolution dans le temps au niveau individuel, observée périodiquement pendant un laps de temps donné permettant l'analyse des biographies.
- Importance de tenir compte du facteur **temps** dans l'analyse (*C'est la particularité de l'analyse des biographies*).

Généralités

- L'analyse de survie, c'est l'analyse des **données censurées**.
- L'objectif principal est d'étudier directement le délai de survenue d'un événement.
- La durée de survie s'exprime généralement par des courbes obtenues à l'aide d'un calcul probabiliste.
- L'analyse de survie prend en compte simultanément le nombre d'événements survenus pendant une période donnée, le moment où ces événements se produisent et les sujets pour lesquels l'événement ne s'est pas encore réalisé (**données censurées**).

Généralités

- Dans ce type d'analyse, il s'agit d'expliquer l'occurrence (apparition dans le temps ou dans l'espace) d'un événement au cours du temps (accès à l'emploi, accès à la propriété, accès au logement, sortie d'une cohorte, entrée dans les rapports sexuelles, etc.)
- Analyser **l'interaction** entre les événements (fumer la cigarette et avoir le cancer, le mariage et la naissance du premier enfant, etc.)
- Faire une analyse **causale** (ex. le contexte social influence le comportement des femmes en matière de fécondité, les divorces cause des abandons scolaires, etc.)
- Il faut se familiariser avec les dénominations suivantes: **l'analyse des transitions, des durées, de survie, causale.**

Éléments introductifs

C'est quoi l'Analyse causale?

- L'événement A est la cause l'événement B si :
- A et B Sont reliés de manière fonctionnelle;
- A précède B dans le temps;
- La variable explicative et la variable à expliquer doivent être situées dans le temps;
- La variable explicative précède le phénomène à étudier, n'oubliez pas on est dans le temps;
- Le principe de la priorité temporelle de la cause sur l'effet doit être respecté.

Eléments introductifs

Signification d'un taux de survie

- Dans un groupe donné de malades, le taux de survie à 10 ans est de 0,53 : veut dire qu'un malade appartenant à ce groupe a 53% de chances de survivre au bout de 10 ans à partir d'une date d'origine fixée avant.

Rôles de l'analyse des biographies

- L'analyse des biographies permet de traiter des données longitudinales censurées.
- Ces données peuvent notamment être obtenues dans des enquêtes comportant un volet rétrospectif, décrivant l'histoire résidentielle, familiale ou professionnelle de l'individu.
- On peut alors étudier l'éventuelle survenue d'un évènement dans le cycle de vie d'un individu en fonction de ses caractéristiques.

Exemples

- Mesure de variables sur des patients dans différentes conditions: assis, couché, après avoir pris le médicament....
- Mesure de variables sur des patients toutes les semaines données

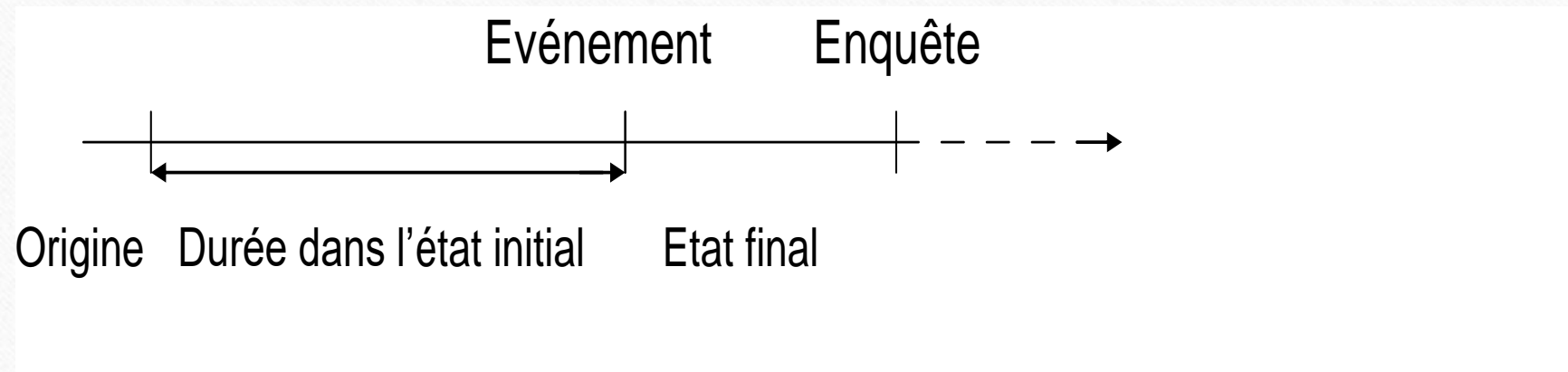
Notions d'événement et de troncatures

Un événement est la transition d'un état qualitatif à un autre.

- Exemple: L'enfance est la transition de l'état d'enfant à celui d'adulte
- NB: La transition peut se faire vers un seul ou plusieurs états.
- Exemple: La rupture du mariage peut se faire par le divorce ou par le veuvage

Notions d'événement et de troncatures

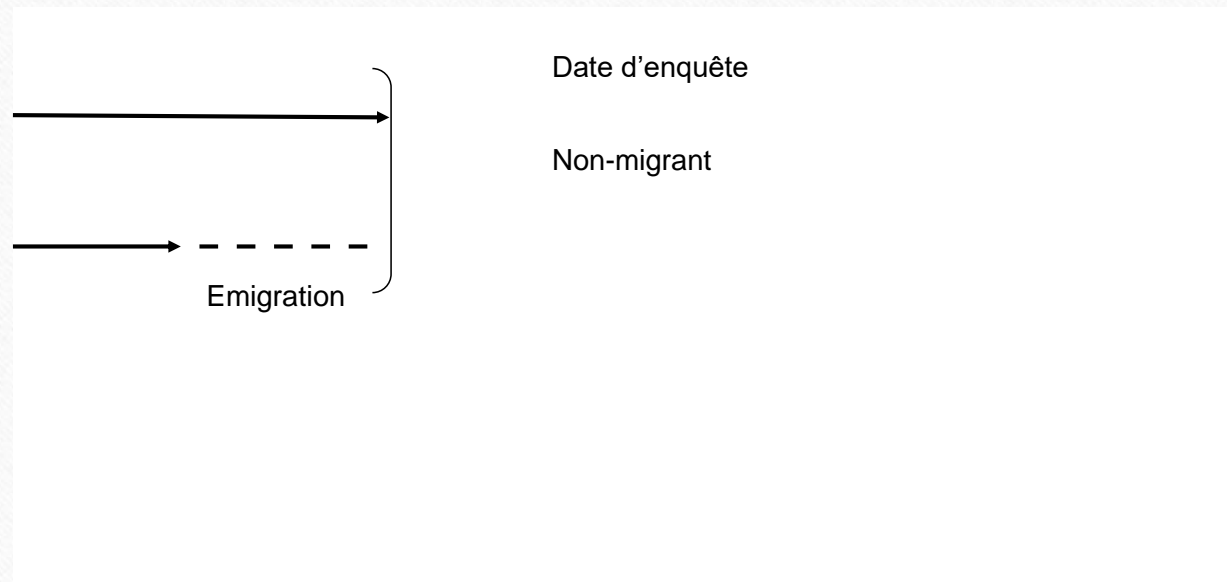
Durée avant l'événement



Notions d'événement et de troncatures

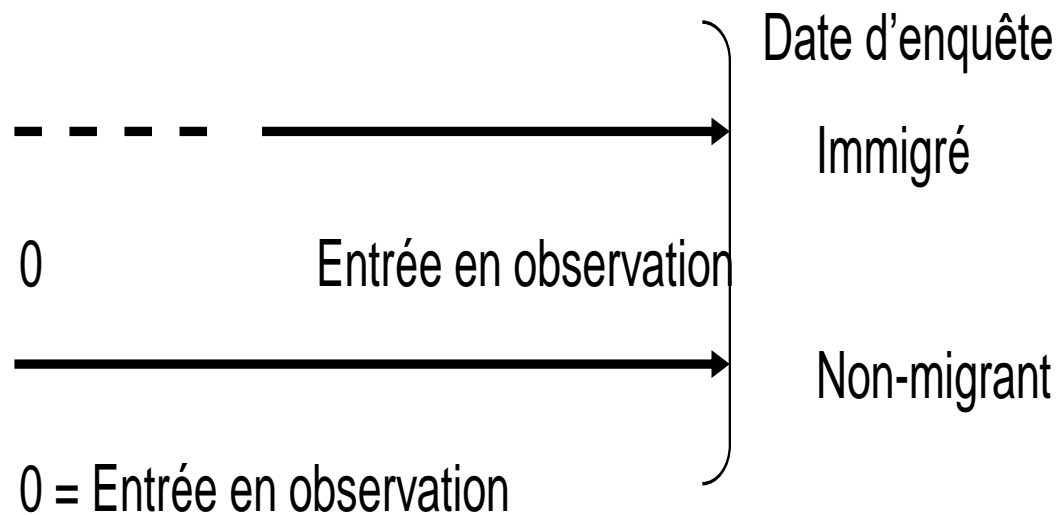
- **Censure:** lorsque l'évènement n'est pas survenu pour certaines personnes, on dira que l'information est censurée pour ces personnes.
- Il est difficile d'observer les individus sur toute leur vie ou même sur un intervalle de temps donné.
- **Troncature à droite:** l'observation de l'évènement est interrompue par la date de l'enquête ou par l'émigration.

Notions d'événement et de troncatures



Notions d'événement et de troncatures

Troncature à gauche : Entrée dans le champ d'observation (Immigration)



Notions d'événement

Etat initial	Etat final	Evénement	Origine	Durée
Vivant	Mort	Décès	Naissance	Age au décès
Célibataire	Marié	Mariage	Naissance	Age au mariage
Allaité	Non allaité	Sevrage	Naissance	Durée d'allaitement

Recueil et nature des données

Recueil des biographies

- Il est question de retracer les événements connus par une personne à partir d'un moment donné de sa vie jusqu'à la date de l'enquête;
- Ce type de collecte nécessite de faire appel à la mémoire à court et à moyen terme, de bien repérer dans le temps les événements vécus par les enquêtés
- **Outils de collecte**
- Questionnaire biographique et fiche **AGEVEN**

Recueil et nature des données
Fiche AGEVEN d'une enquête sur les logements

Voir fiche

Sources de données biographiques

- les Enquêtes rétrospectives,
- les Enquêtes « prospectives » (enquêtes à passages répétés, registres de population, observatoires de population).

Où trouver les données biographiques

- Instituts de recherches (ISSP Institut Supérieur des Sciences de la Population (UJKZ),
- INSS (Institut des Sciences des Sociétés)
- etc

En somme

- Variable à expliquer ou variable dépendante qui est la durée avant l'occurrence d'un événement précis (mariage, migration, décès, etc.), mesurée par rapport à une origine.
- On s'intéresse au risque que l'événement se produise en fonction de la durée écoulée depuis une origine
- La durée ou le risque dépend des variables explicatives fixes (qui ne varient pas comme le sexe, l'ethnie, etc.) et des variables en fonction du temps (état matrimonial, statut d'activité)

Limites

- Approche basée sur la mémoire de moyen ou long terme
- Risques d'oubli ou de la datation imprécise de certains événements
- L'analyse s'intéresse aux groupes d'individus
- Problème de représentativité de groupe pour certaines transitions (petite taille de l'échantillon)
- La succession des événements importe plus que leur datation précise

La méthode de Kaplan-Meier

- La méthode de Kaplan-Meier permet d'estimer les **fonctions de survie**.
- L'estimateur de Kaplan and Meier [1958] est un estimateur classique de la fonction de survie qui ne nécessite pas que les intervalles de temps soient réguliers.
- Cette technique est utilisée pour l'analyse de données de survie, qu'il s'agisse d'individus (recherche sur la leucémie par exemple), ou de produits (résistance au temps d'un outil de production par exemple).

La méthode de Kaplan-Meier

- On entendra parler beaucoup de **sortie d'observation ou d'étude**: l'on sort de l'étude si l'on meurt (les produits se cassent ou sont hors d'usage)
- **On sort aussi de l'étude** lorsqu'on est guérit, ou encore lorsque l'on perd la trace (déménagement par exemple),
- ou encore lorsque que l'étude est interrompue.

La méthode de Kaplan-Meier

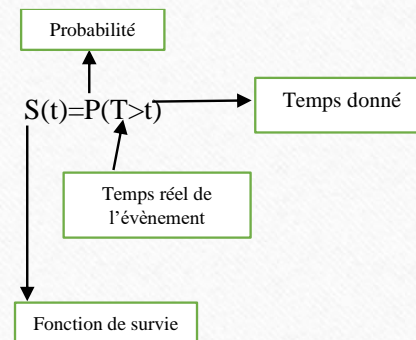
- Le premier type d'information est appelé « **données événement** », tandis que le second est appelé « **données censurées** ».

La méthode de Kaplan-Meier

- L'analyse de Kaplan-Meier **permet aussi** de comparer des populations, en s'appuyant sur leur courbe de survie.
- Par exemple, il peut être intéressant de comparer les temps de survie des hommes et des femmes face à une même maladie,
- ou de comparer les temps de casse pour un même produit fabriqué sur deux chaînes de production différentes.

La fonction de survie

- **La fonction de survie:** C'est la probabilité qu'un patient vive au moins un certain temps après le diagnostic de sa maladie (peut-être un cancer). Elle prédit le pronostic favorable ou défavorable et cela peut permettre au médecin de mieux adapter le traitement pour un patient.



Méthode d'analyse descriptive : estimateurs de Kaplan-Meier

- **Tables de séjour de Kaplan-Meier**
- Elles permettent de décrire les événements qui varient au cours du temps.
- **Estimateurs de Kaplan-Meier**
- Ils mesurent la probabilité d'occurrence de l'événement à un instant t_i pour les individus n'ayant pas jusque-là connu l'événement étudié.
- **Courbe de survie de Kaplan-Meier**
- Elle représente la distribution de la durée avant la réalisation de l'événement
- Elle s'interprète comme la proportion de « survivant » pour chaque durée de séjour dans un état donné.

Méthode d'analyse descriptive : estimateurs de Kaplan-Meier

Quotients instantanés $h(t_i) = \frac{d(t_i)}{N(t_i)}$

$d(t_i)$ = nombre d'événements observés en t_i ; $N(t_i)$ = effectif soumis au risque juste avant t_i ; $t_1, t_2, \dots, t_k, \dots, t_n$, âges auxquels les individus ont connu l'événement étudié

Proportion des survivants à chaque temps (âge) :

$$S(t_i) = \prod_{j \leq i} (1 - h(t_j))$$

Exemple d'une table de séjour de Kaplan-Meier

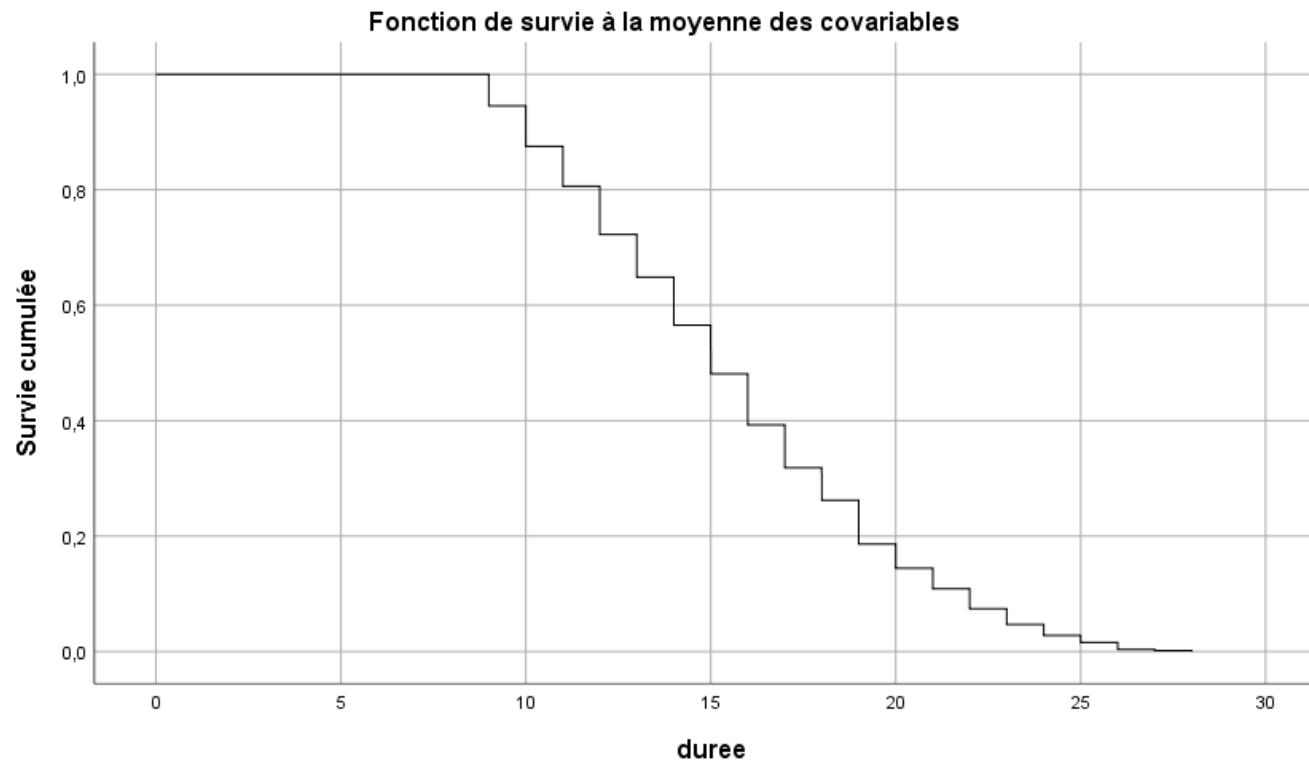
Intervalle de temps (t_i)	Nombre d'individus en début d'intervalle $N(t_i)$	Événements $d(t_i)$	Troncatures	Quotient s $h(t_i)$	Proportion des survivants $S(t_i)$
1 2	50	3	0	0,060	0,94
2 3	47	0	1	0,000	0,94
3 4	46	3	2	0,065	0,88
4 5	41	2	0	0,049	0,84
5 6	39	1	2	0,026	0,82

Exemple d'une table de séjour de Kaplan-Meier

$$h(t_1) = \frac{3}{50} = 0,06 \Rightarrow S(t_1) = 1 - 0,06 = 0,94$$

$$h(t_2) = \frac{0}{47} = 0 \Rightarrow S(t_2) = [1 - h(t_1)][1 - h(t_2)] = 0,94$$

Courbe de survie, Kaplan Meier



Le modèle semi paramétrique de cox

La méthode semi paramétrique (modèle de Cox) permet d'effectuer une modélisation et de mesurer l'impact de variables explicatives, fixes ou dépendantes du temps, sur le quotient instantané.

Méthode d'analyse explicative : régression semi-paramétrique de Cox

- La méthode de Cox permet de mesurer le risque relatif instantané de connaître l'événement étudié si on a une caractéristique z_i : **$h(t, z) = h_0(t)\exp(z\beta)$**
- $h_0(t)$ est le quotient instantané pour la catégorie de référence ($z = 0$)
- β représente les coefficients estimés de z_i :
- **$z\beta = z_1\beta_1 + z_2\beta_2 + \dots + z_n\beta_n$**
- L'exponentiel de chaque coefficient (**$z\beta$**) fournit le risque relatif de connaître l'événement par rapport à la catégorie de référence

$$\exp(\beta_1) = \frac{h(t; z_1)}{h(t; 0)}$$

Régression de cox: un exemple de resultat

Variables dans l'équation								
	B	Erreur standard	Wald	df	Sig.	Exp(B)	95,0% CI pour Exp(B) Plus bas Supérieur	
sexe	-,050	,050	,990	1	,320	,951	,861	1,050
quel est/était votre statut d'occupation dans ce logement à			140,360	2	,000			
quel est/était votre statut d'occupation dans ce logement à (1)	-2,463	,449	30,061	1	,000	,085	,035	,205
quel est/était votre statut d'occupation dans ce logement à (2)	-2,580	,244	111,428	1	,000	,076	,047	,122
groupe d'âges	1,076	,059	329,663	1	,000	2,933	2,612	3,295

EXEMPLE PRATIQUE

Premier départ du toit parental

Ce qu'il faut pour commencer

- Description des données biographiques
- Conceptualisation d'un événement
- Construction du fichier d'analyse
- Analyse des données

Description des données biographiques

- **Lieu:** à Bafia
- **Année de l'enquête :** 2010
- **Population cible :** Individus âgés de 15 à 34 ans en 2010
- **Strates retenues :** Hommes et femmes appartenant aux groupes d'âges 15-24 ans et 25-34 ans
- **Taille de l'échantillon :** 1579
- **Parmi les informations collectées:** L'itinéraire résidentiel de chaque enquêté (logement actuel, premier logement, statut dans le logement, etc.)

Conceptualisation de l'événement : premier départ du toit parental

- **Événement étudié :** Premier départ du toit parental depuis l'âge de 6 ans
- **Début de l'observation:** Première cohabitation avec les parents
- **Population soumise au risque:** Tous les individus ayant cohabité avec leurs parents
- **Durée d'observation:** âge au moment du premier départ du toit parental – âge au moment de la première cohabitation
- **Troncatures :** Ceux qui cohabitent toujours avec leurs parents au moment de l'enquête,
- **Pour ceux qui cohabitent encore avec leurs parents:**

Durée = date de l'enquête – date de la cohabitation avec les parents

Conceptualisation d'un événement

- Un fichiers de données est fourni nommé fusion. Celui-ci est issue de la fusion des bases(indiv.sav) et de (loge.sav).

Le premier comporte les données sur les caractéristiques sociodémographiques des enquêtés. Le deuxième est un ensemble de données événementielles sur les différentes périodes de logements occupés par les enquêtés.

Dans le premier fichier, chaque ligne correspond à un individu et chaque colonne à une variable tandis que dans le second fichier (fichier biographique), caque ligne correspond à un événement (transition) et chaque colonne une variable.

Conceptualisation d'un événement

NB: Pour faire l'analyse des biographies, on a besoin de deux éléments fondamentaux

Durée d'observation (avant l'événement s'il se réalise)

L'Indice de troncature qui prend la valeur 1 si l'événement se réalise et 0 sinon

Le début d'observation doit être un instant de référence commun à tous les individus

La date de fin d'une période est égale à la date de début de la période suivante

Application sur spss_25

• ---

Références bibliographiques

- Adeline Morisot (2015). Méthodes d'analyse de survie, valeurs manquantes et fractions attribuables temps dépendantes : application aux décès par cancer de la prostate. Médecine humaine et pathologie. Université Montpellier, 2015. Français.
- NGANAWARA Didier (2017-2019), Cours Master 2 en démographie, IFORD, Yaoundé II
- Kankoé SALLAH, Méthode actuarielle d'estimation des courbes de survie : principe, différences avec la méthode de Kaplan-Meier. DELBIM – Université de Lomé, 7p.
- Kaplan EL, Meier P. (1958). Nonparametric estimation from incomplete observations. J Am Stat Assoc, (53): p. 457-81
- <https://www.xlstat.com/fr/solutions/fonctionnalites/analyse-kaplan-meier>