

*1er Stage : Mortalité au Lac Saint Jean, Impact des
liens familiaux*

*2ème stage : Méthodologie d'imputation de données
d'enquête*

Viard Julie
Université de Rennes 1

Département de mathématiques, **Université de Québec à Montréal**
Division des Méthodes d'Enquêtes auprès des Entreprises, **Statistique Canada**

16 mai 2013

Sommaire

- 1 Mortalité au Lac Saint Jean, Impact des liens familiaux
- 2 Méthodologie d'imputation de données d'enquête
- 3 Les apports de l'expérience + Césure

*1er Stage : Stage de recherche au département de mathématique de
L'UQAM*

Mortalité au Lac Saint Jean, Impact des liens familiaux

- 1 Présentation de l'étude et notions utilisées
- 2 Traitement des données et prise en compte de la censure
- 3 Impact des liens parentaux sur la mortalité

Mortalité au Lac Saint Jean, Impact des liens familiaux

Présentation de l'étude et notions utilisées

Objectif de l'étude :

Existe-t-il un lien de corrélation entre l'âge du décès d'un individu et celui de ses parents ?

Calcul de probabilités de survie en fonction des antécédants parentaux

Construction de tables de mortalités dépendantes des conditions de décès des parents.

Définition INSEE : « *Une table de mortalité annuelle suit le cheminement d'une génération fictive de 100 000 nouveau-nés à qui l'on fait subir aux divers âges les conditions de mortalité observées sur les diverses générations réelles, durant l'année étudiée.* »

Mortalité au Lac Saint Jean, Impact des liens familiaux

Présentation de l'étude et notions utilisées

Provenance des données :



Le fichier BALSAC : Banque de données informatisée développée depuis 40 ans à l'Université de Québec à Chicoutimi. Construit à partir des actes de l'état civil du Québec.

Nos données : Base référençant l'intégralité des individus nés et/ou décédés au Saguenay entre 1838 et 1978.

Mortalité au Lac Saint Jean, Impact des liens familiaux

Présentation de l'étude et notions utilisées

Format des données :

ind	sexe	datn	datd	age_declare	pere	datn_pere	datd_pere	mere	datn_mere	datd_mere
303	2	5/12/1898	6/19/1899	513	5247515	8/29/1869	11/16/1957	5247516	1/13/1875	3/13/1947
304	1	5/2/1910	6/2/1910	501	5247515	8/29/1869	11/16/1957	5247516	1/13/1875	3/13/1947
305	2	10/4/1894	7/30/1895	510	5247515	8/29/1869	11/16/1957	5247516	1/13/1875	3/13/1947
306	2	2/23/1906	3/8/1906	314	5247515	8/29/1869	11/16/1957	5247516	1/13/1875	3/13/1947
307	1	5/2/1910	5/6/1910	303	5247515	8/29/1869	11/16/1957	5247516	1/13/1875	3/13/1947
308	1	4/26/1912	4/26/1912	999	5247515	8/29/1869	11/16/1957	5247516	1/13/1875	3/13/1947
309	2	11/19/1913	5/13/1914	506	5247515	8/29/1869	11/16/1957	5247516	1/13/1875	3/13/1947
310	2	7/26/1899	7/1/1918	19	5247515	8/29/1869	11/16/1957	5247516	1/13/1875	3/13/1947
311	2	4/2/1897	NA	NA	5247515	8/29/1869	11/16/1957	5247516	1/13/1875	3/13/1947
312	2	11/22/1917	5/19/1918	506	5247515	8/29/1869	11/16/1957	5247516	1/13/1875	3/13/1947

FIGURE: Echantillon de la base de données

Mortalité au Lac Saint Jean, Impact des liens familiaux

Présentation de l'étude et notions utilisées

Indicateurs de mortalité

- **La durée de vie** Soit T la variable aléatoire représentant la durée de vie d'un individu. On définit T_x comme la **durée de vie résiduelle** de l'individu sachant qu'il est vivant à l'âge x :

$$T_x = [T - x | T > x]$$

- **Le quotient de mortalité**

On définit le quotient de mortalité entre x et $x + t$ par :

$${}_tq_x = 1 - {}_tp_x = P(T_x \leq t) = P(T \leq x + t | T > x)$$

C'est la probabilité qu'un individu d'âge x a de décéder avant l'âge $x + t$.

Mortalité au Lac Saint Jean, Impact des liens familiaux

Présentation de l'étude et notions utilisées

- le **taux instantané de mortalité** :

Soit μ_{x+t} le **taux instantané de mortalité** des individus d'âge x à la date t . Il représente le taux de mortalité par unité de temps, ie qu'il mesure le risque qu'un individu d'âge x décède entre la date t et la date $t + \Delta t$. On le définit par :

$$\mu_{x+t} = \lim_{\Delta t \rightarrow 0^+} P[t < T_x \leq t + \Delta t | T_x > t] / \Delta t$$

Indicateur très souvent utilisé pour observer la mortalité, généralement $\Delta t = 1$

Mortalité au Lac Saint Jean, Impact des liens familiaux

Présentation de l'étude et notions utilisées

Le modèle de Lee & Carter

Modèle basé sur une **approche prospective**.

Deux paramètres liés à l'âge x et un troisième lié au temps t

$$\ln(\mu_{x+t}) = a_x + b_x k_t + \epsilon_{x,t}$$

- a_x s'interprète comme la valeur moyenne des $\ln(\mu_{x+t})$ par âge au cours du temps.
- k_t décrit l'évolution générale du taux instantané de mortalité dans le temps.
- b_x traduit la sensibilité de la mortalité instantanée à l'âge x par rapport à l'évolution générale k_t

Mortalité au Lac Saint Jean, Impact des liens familiaux

Présentation de l'étude et notions utilisées

Le modèle de Lee & Carter

Pour assurer identifiabilité du modèle :

$$\sum_x b_x = 1$$

$$\sum_t k_t = 0$$

Estimation des paramètres par la méthode des moindres carrés ordinaires.

Prédiction des taux instantanés de mortalité à venir grâce à l'extrapolation de la tendance temporelle K_t

Les $\epsilon_{x,t}$ sont des résidus supposés iid selon une loi $\mathcal{N}(0, \sigma^2)$

Critique du modèle : hypothèse d'homoscédasticité des erreurs.

Mortalité au Lac Saint Jean, Impact des liens familiaux

Présentation de l'étude et notions utilisées

Amélioration du modèle de Lee & Carter :

Alternative de Brouhns, Denuit et Vermunt, le **modèle log poisson** :

${}_tD_x$: nombre de décès d'individus d'âge x durant l'année t .

${}_tL_x$: nombre d'individus qui ont l'âge x à la date t (=exposition)

On a :

$${}_tD_x \sim \mathcal{P}({}_tL_x \cdot \mu_{x+t})$$

Avec :

$$\mu_{x+t} = \exp(a_x + b_x k_t)$$

Le modèle binomial :

$${}_tD_x \sim \mathcal{B}({}_tL_x, \mu_{x+t})$$

Mortalité au Lac Saint Jean, Impact des liens familiaux

Traitement des données et prise en compte de la censure

Traitement des données effectué :

- Problème de données manquantes
- Récupération d'un maximum d'information concernant les parents
- Définition des années d'observation
- Lissage des âges par la méthode des splines.

Mortalité au Lac Saint Jean, Impact des liens familiaux

Traitement des données et prise en compte de la censure

Données censurées : cas d'une **censure fixe à droite**, à partir d'une date fixe C (ici 1978) la durée de vie des individus encore en vie n'est pas observable.

Exposition : nombre d'individus en vie par âge et par année.

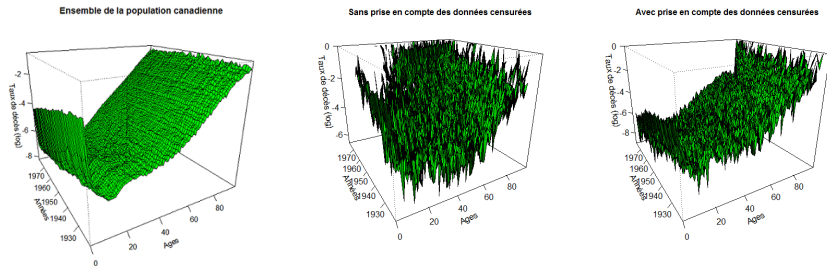


FIGURE: Comparaison des surfaces de mortalité avec prise en compte ou non des données censurées

Mortalité au Lac Saint Jean, Impact des liens familiaux

Traitement des données et prise en compte de la censure

Taux instantanés de mortalité : Paramètre a_x

Rappel modèle Lee & Carter :

$$\ln(\mu_{x,t}) = \alpha_x + \beta_x k_t + \epsilon_{x,t}$$

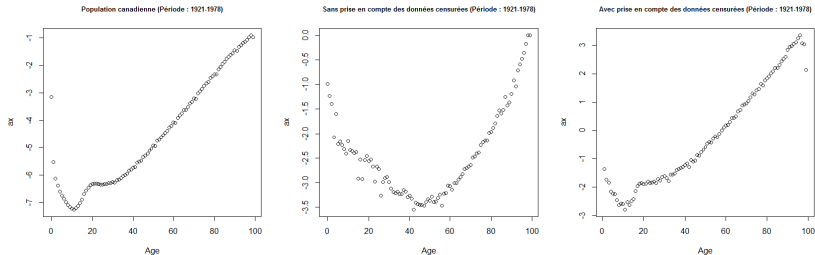


FIGURE: Evolution du paramètre a_x par rapport à l'âge des individus

Mortalité au Lac Saint Jean, Impact des liens familiaux

Traitement des données et prise en compte de la censure

Taux instantanés de mortalité : Paramètre b_x

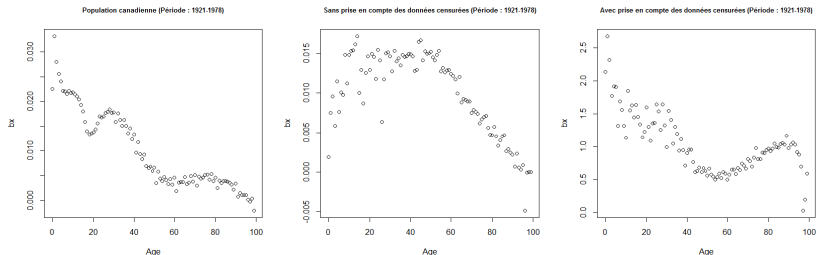


FIGURE: Evolution du paramètre b_x par rapport à l'âge des individus

Mortalité au Lac Saint Jean, Impact des liens familiaux

Traitement des données et prise en compte de la censure

Taux instantanés de mortalité : Paramètre k_t

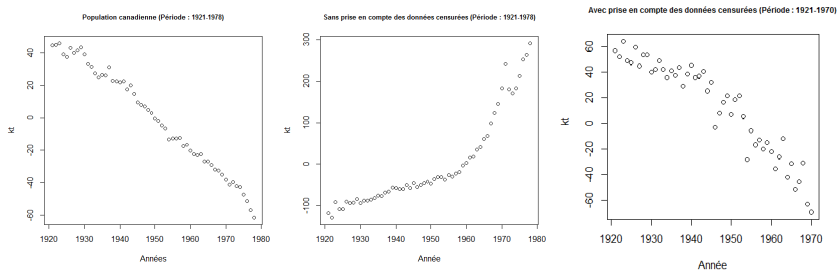


FIGURE: Evolution du paramètre k_t dans le temps

Mortalité au Lac Saint Jean, Impact des liens familiaux

Impact des liens parentaux sur la mortalité

Mortalité suivant l'âge de décès des parents

Comment tenir compte des données censurées concernant les parents ?

65% de date de décès des père manquantes et 62% des mères (après récupération des données)

Une solution pour réduire le problème de censure : travailler sur les périodes les plus anciennes.

Mortalité au Lac Saint Jean, Impact des liens familiaux

Impact des liens parentaux sur la mortalité

Mortalité suivant l'âge de décès des parents

Classement des individus selon quatre catégories suivant leurs antécédents parentaux

- Les deux parents sont décédés jeunes
- Le père est décédé jeune, la mère est décédée âgée
- Le père est décédé âgé, la mère est décédée jeune
- Les deux parents sont décédés âgés

Il nous faut alors définir un âge à partir duquel on considère le parent décédé jeune.

Mortalité au Lac Saint Jean, Impact des liens familiaux

Impact des liens parentaux sur la mortalité

Mortalité suivant l'âge de décès des parents

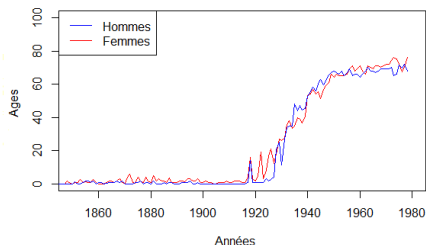


FIGURE: Mediane age au décès population du Saguenay

Mortalité au Lac Saint Jean, Impact des liens familiaux

Impact des liens parentaux sur la mortalité

Mortalité suivant l'âge de décès des parents

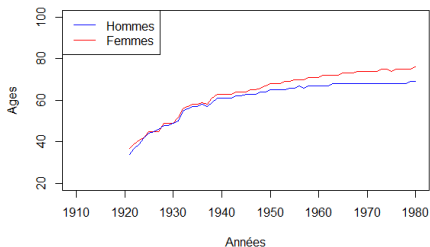


FIGURE: Mediane age au décès population nationale

Problème La très grande majorité des individus se retrouve dans la catégorie « les deux parents sont décédés agés »

Mortalité au Lac Saint Jean, Impact des liens familiaux

Impact des liens parentaux sur la mortalité

Mortalité suivant le nombre de parents en vie au début de l'année

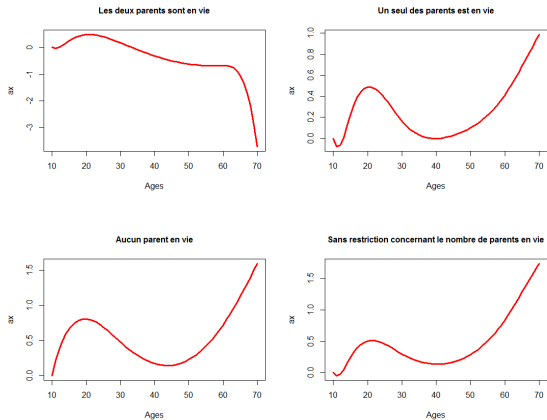


FIGURE: Evolution du paramètre a_x suivant le nombre de parents en vie

Mortalité au Lac Saint Jean, Impact des liens familiaux

Impact des liens parentaux sur la mortalité

Rapport entre taux instantanés de mortalité avec restriction concernant le nombre de parents en vie et taux instantanés de mortalité sans restriction

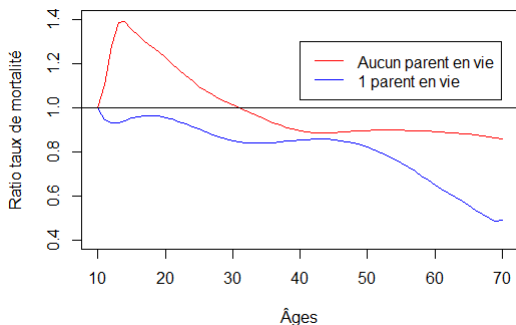


FIGURE: Ratio des taux de mortalité par âge

Mortalité au Lac Saint Jean, Impact des liens familiaux

Impact des liens parentaux sur la mortalité

Méthode construction table de mortalité :

- Modéliser taux de mortalité instantané pour chaque catégorie
- Extrapolation de la tendance temporelle (méthodologie de Box et Jenkins)
- Calculer probabilité de décès à partir des taux instantané de mortalité. Pour cela on utilise la relation suivante :

$${}_t p_x = \exp\left(-\int_0^t \mu_{x+u} du\right)$$

Mortalité au Lac Saint Jean, Impact des liens familiaux

CONCLUSION

- Les résultats de cette étude ne nous permettent malheureusement pas de conclure
- Si tel avait été le cas, on aurait construit différentes tables de mortalités suivant les antécédents parentaux.
- Construire différentes tables suivant le sexe des individus.
- Prendre en compte le sexe des parents lors de la segmentation de la population.