

RAPPORT DE STAGE

MASTER 2 MASS

Analyse de la fécondité

A partir des données du système de suivi
démographique de la zone de Niakhar, Sénégal
par la méthode d'analyse statistique des biographies

Avril-Août 2012



Photo prise dans la zone de Niakhar, Sénégal, juin 2012

Maître de stage : Valérie Delaunay
Référent universitaire : Sébastien Oliveau

Ambre BUIATTI

Aix-Marseille Université - Institut de recherche pour le développement

Master 2 de Mathématiques Appliquées et Sciences Sociales,
Spécialité Analyse des populations

Rapport de stage

Analyse de la fécondité,
à partir des données du système de suivi démographique de la
zone de Niakhar, Sénégal,
par la méthode d'analyse statistique des biographies.

Rédigé par :

Ambre BUIATTI

Etudiante en Master 2 MASS, analyse des populations, Aix-Marseille Université
Stagiaire à l'Institut de Recherche pour le Développement (IRD), UMR 198 URMITE

Encadré par :

Valérie DELAUNAY
Démographe, chercheur à l'IRD
UMR 151 LPED et UMR 198 URMITE

&

Sébastien OLIVEAU
Maître de conférences en géographie, Aix-Marseille Université
UMR 7300 ESPACE

Année 2012

Avant-propos

Au cours de la deuxième et dernière année de formation les étudiants du master MASS (Mathématiques Appliquées et Sciences Sociales) doivent effectuer un stage de fin d'étude d'une durée de 4 à 6 mois. Ce stage est développé autour d'une préoccupation professionnelle et doit se réaliser dans une administration, institut, entreprise ou tout autre organisme professionnel. L'objectif de ce stage de fin d'étude est de permettre à l'étudiant de mettre en pratique les outils théoriques et méthodologiques acquis au cours de sa formation, d'identifier ses compétences et de conforter son objectif professionnel. Il est donc une application directe des connaissances et outils d'analyse acquis durant la formation à travers des réalisations pratiques. Il permet aussi de compléter les enseignements théoriques dispensés tout au long de la formation.

Le présent document constitue le rapport de stage que j'ai rédigé sur le thème « Analyse de la fécondité, à partir des données du système de suivi démographique par la méthode d'analyse statistique des biographies » sous la direction de Madame Delaunay Valérie, chercheur à l'IRD. Il est le résultat d'un stage de quatre mois effectué du 16 avril au 16 août 2012 au sein de l'Unité Mixte de Recherche 198 (UMR 198 URMITE) de l'Institut de Recherche pour le Développement (IRD), situé à Dakar au Sénégal.

Remerciements

Mes remerciements s'adressent tout d'abord conjointement aux deux personnes qui ont particulièrement marquées cette fin de cycle universitaire :

J'adresse ma profonde reconnaissance envers ma tutrice Valérie Delaunay qui a assuré les bonnes conditions de mon stage, et ce dès le début par son accueil chaleureux. J'ai apprécié son encadrement ainsi que ses conseils et encouragements qui m'ont permis d'accomplir ce travail de la meilleure façon qu'il soit. Elle a su rendre ce stage d'autant plus intéressant et agréable avec notamment la découverte de l'observatoire démographique de Niakhar et de son fonctionnement. C'était un réel plaisir d'avoir une tutrice pleine d'humanité, de gentillesse et de professionnalisme.

Je remercie également Marie-Christine Roubaud, responsable du master MASS, qui malgré ses nombreuses occupations, a toujours su trouver du temps pour ses étudiants et a su rendre joviale la vie de notre promotion MASS durant ces deux dernières années. Je suis reconnaissante envers son investissement dans ma recherche de stage. Sa présence lors de certaines difficultés rencontrées au cours de mon parcours universitaire a été un soutien appréciable.

Je tiens aussi à remercier chaleureusement l'équipe de l'UMR 198 de l'IRD, surtout l'équipe du bâtiment H3 du campus de Hann et en particulier mes collègues et amis, Laetitia Douillot pour son accueil, ses conseils et sa gentillesse, Nourou Sylla pour son aide et son soutien, Emilie Ndiaye et Prosper Ndiaye pour leur aide dans les procédures administratives et leur compagnie quotidienne, ainsi que toutes les autres personnes qui de près ou de loin ont su rendre le cadre de mon stage convivial.

Une pensée particulière pour Antoine Ndour pour sa gentillesse et son attention à mon égard.

Un remerciement s'adresse aussi aux enquêteurs de la zone de Niakhar qui m'ont fait découvrir le travail de recueil des données et aux villageois de Gadiak pour leur accueil souriant lors de mon bref passage sur le terrain d'étude.

Que Philippe Bocquier soit remercié pour son précieux travail qu'il réalise sur la méthode d'analyse des biographies mise en œuvre sous le logiciel Stata, son travail a servi de base au mien.

Merci à Sébastien Oliveau d'avoir accepté une première fois d'être mon référent universitaire, et surtout de l'avoir accepté une deuxième fois.

Merci à Julie Raud pour son aide et sa disponibilité concernant toute les démarches administratives.

Que ma famille aussi soit remerciée pour son aide dans la réalisation de mes séjours à l'étranger en cette deuxième année de master.

SOMMAIRE

Avant-propos.....	2
Remerciements	3
INTRODUCTION GENERALE	8
PREMIERE PARTIE	9
I _PRESENTATION DE L'ENTREPRISE.....	10
1.1 L'Institut de Recherche pour le Développement (IRD)	10
1.2 Le campus international UCAD-IRD de Hann et son UMR 198 URMITE	12
II _PRESENTATION DE LA MISSION.....	13
2.1 La mission et son but.....	13
2.2 Le déroulement de la mission	14
DEUXIEME PARTIE	16
Introduction.....	17
I. La méthode d'analyse biographique	18
1.1 Généralité.....	18
1.2 Estimateurs et modèles.....	19
1.3 Conceptualisation : analyse de la fécondité de Niakhar	22
II. La démarche d'analyse et les procédures	23
2.1 Mise en forme de la table « RESIDENCY.DTA ».....	26
2.2 Création de la variable de censure à droite	27
2.3 Création de la variable résidence	27
2.4 Création de la table fécondité.....	27
2.5 Création du calendrier par année et groupe d'âge	28
2.6 Création de la variable de censure et analyse de la fécondité.....	28
III. Analyse de la fécondité, zone de Niakhar, 1983-2011	30
3.1 Les indicateurs de la fécondité.....	31
3.2 Evolution générale de la fécondité.....	32
3.3 Fécondité et religion des femmes	37
3.4 Fécondité et instruction des femmes.....	39
3.5 Fécondité, niveau de vie et richesse agro-pastorale.....	41
Fécondité et niveau de vie, zone de Niakhar entre 1999 et 2003	42
Fécondité et niveau de richesse agro-pastorale, zone de Niakhar entre 1999 et 2003	45

3.6 Modélisation : les facteurs explicatifs de la fécondité	47
3.7 Fécondité et village	49
Conclusion	51
CONCLUSION du stage	52
BIBLIOGRAPHIE.....	53
ANNEXES.....	54
ANNEXES A : LES PROGRAMMES STATA	54
ANNEXE A.1: Programme « structure_standard.do »	54
ANNEXE A.2 : Programme « censure_OBE.do ».....	58
ANNEXE A.3 : Programme « création_résidence.do »	60
ANNEXE A.4 : Programme pour créer la table child.dta et pour merger la table residency et child « merge_fécondite.do »	62
ANNEXE A.5 : Programme pour la création du calendrier et des groupes d'âges "stsplit_calendrier_année.do"	66
ANNEXE A.6 : Programme pour l'analyse de la fécondité "analyse_fécondité.do"	69
ANNEXE A.7 : Programme pour la création de la variable vi_co_cu « création vi_vo_cu.do"	71
ANNEXE A.8 : Programme pour l'analyse des déterminants de la fécondité « analyse_fécondité.do"	73
ANNEXES B : LES FEUILLES DE CALCUL SOUS EXCEL (Feuilles-outilles)	84
ANNEXE B.1 : Matrice des incohérences.....	84
ANNEXE B.2 : Feuilles-outilles pour le calcul des indicateurs de la fécondité	85
ANNEXES C : LES SORTIES STATA.....	87
ANNEXE C.1 : Table de survie par groupe d'âge des femmes de ne pas avoir d'enfant sur l'ensemble de la période	87
ANNEXE C.2 : Table de survie par groupe d'âge des femmes d'avoir un enfant sur l'ensemble de la période.....	87
ANNEXE C.2 : Table de survie des femmes sans connaître l'évènement naissance, par période	88
ANNEXE C.3 : Personne-année, Nombre de naissance, Taux de fécondité (15-49 ans) par année seulement et par année et groupe d'âge quinquennal.....	89
ANNEXE C.4 : Personne-année, Nombre de naissance, Taux global de fécondité (15-49 ans) par période, période et groupe d'âge	94
ANNEXE C.5 : Age moyen des femmes à la 1ère naissance	95
ANNEXE C.6 : Evolution de la répartition des naissances par groupe d'âge des femmes selon la période de naissance, période 1 1984-1988 et période 6 2009-2011.	96
ANNEXE C.7 : Courbes de Kaplan-Meier, probabilité de survie des femmes n'ayant pas connu l'évènement naissance à chaque âge selon les six périodes.....	96

ANNEXE C.8 : Analyse fécondité et religion, test du log-rang et courbe de survie.....	97
ANNEXE C.9 : Analyse fécondité et instruction, test du log-rang.....	98
ANNEXE C.10 : Analyse fécondité et niveau de vie, test du log-rang.....	99
ANNEXE C.11 : Analyse fécondité et niveau de richesse agro-pastorale, test du log-rang.....	104
ANNEXE C.12 : Modèle de Cox	105
ANNEXE D : PRESENTATION DE L'IRD AU SENEGAL	107
ANNEXE E : PRESENTATION DU SYSTEME DE SUIVI DEMOGRAPHIQUE DE LA ZONE DE NIAKHAR ET DE SES DONNEES	108
Résumé	109

Table des illustrations

<i>Tableau 1 : Parcours de « vie » d'une femme dans la zone de Niakhar</i>	22
<i>Tableau 2 : Codes des évènements</i>	24
<i>Tableau 3 : Aperçu de la table residency.dta</i>	26
<i>Tableau 4 : Détails des informations fournies par la commande stset sous Stata</i>	28
<i>Figure 1 : Intensité des naissances par âge de la mère</i>	31
<i>Figure 2 : Estimation de la courbe de survie des femmes de ne pas connaître l'évènement « enfant ».</i>	31
<i>Figure 3 : Evolution du taux brut de natalité et du taux global de fécondité, zone de Niakhar, 1984-2011</i>	33
<i>Figure 4 – Evolution de l'indicateur synthétique de fécondité et de l'âge moyen à la maternité, zone de Niakhar, 1984-2011</i>	33
<i>Figure 5 : Evolution des taux de fécondité par groupe d'âge selon la période de naissance ,1984-2011, zone de Niakhar</i>	34
<i>Figure 6 : Evolution des taux de fécondité par groupe d'âge selon la période de naissance, période 1 1984-1988 et période 6 2009-2011</i>	35
<i>Tableau 5 – Taux de fécondité par groupe d'âges, indice synthétique de fécondité (ISF) et âge moyen à la maternité, selon les années d'observation</i>	36
<i>Figure 7 : Répartition des femmes selon leur religion, zone de Niakhar, 1983-2011</i>	37
<i>Tableau 6 : Analyse descriptive de la fécondité par religion des femmes de 1983 à 2011</i>	37
<i>Figure 8 : Taux de fécondité des femmes selon leur religion, zone de Niakhar, 1983-2011</i>	38
<i>Figure 9 : ISF des femmes musulmanes et ISF des femmes chrétiennes, Niakhar, 1984-2011</i>	38
<i>Figure 10 : Répartition des femmes selon leur niveau d'instruction, zone de Niakhar, 1983-2011</i>	39
<i>Tableau 7 : Analyse de la fécondité par niveau d'instruction des femmes, Niakhar, 1983-2011</i>	40
<i>Figure 11 : Indice Synthétique de Fécondité en fonction du niveau d'instruction des femmes, Niakhar, 1999-2003</i>	40
<i>Figure 12 : Taux de fécondité des femmes selon son niveau d'instruction, zone de Niakhar, 1984-2011.</i>	41
<i>Figure 13 : Répartition des femmes selon le niveau de vie, zone de Niakhar, 1999-2003.</i>	42
<i>Tableau 9 : Taux global de fécondité en fonction du niveau de vie des femmes, Niakhar, 1999-2003</i>	43
<i>Figure 14 : Indice Synthétique de Fécondité en fonction du niveau de vie des femmes, Niakhar, 1999-2003</i>	43
<i>Figure 15 : Taux de fécondité des femmes selon leur niveau de vie, zone de Niakhar, 1999-2003.</i>	44
<i>Figure 16 : Répartition des femmes selon le niveau de richesse agro-pastorale, zone de Niakhar, 1999-2003.</i>	45
<i>Tableau 11 : Taux global de fécondité en fonction du niveau de richesse agro-pastorale des femmes.</i>	45
<i>Figure 17 : Taux de fécondité selon le niveau de richesse agro-pastorale des femmes, zone de Niakhar, 1999-2003</i>	46
<i>Figure 18 : Indice Synthétique de Fécondité en fonction du niveau de richesse agro-pastorale des femmes, Niakhar, 1999-2003</i>	46
<i>Tableau 12 : Coefficients du modèle semi-paramétrique à risque proportionnel de Cox pour l'arrivée de l'évènement naissance dans la vie féconde des femmes (sans interaction).</i>	47
<i>Tableau 13 : Coefficients du modèle semi-paramétrique à risque proportionnel de Cox pour l'arrivée de l'évènement naissance dans la vie féconde des femmes (avec interaction).</i>	48
<i>Figure 19 : Zone de Niakhar, Population, 2002</i>	49
<i>Figure 20 : Indice Synthétique de Fécondité en fonction du village de résidence de la femme, Niakhar, 1984-2011</i>	50
<i>Figure 21 : Représentation cartographique : Indice Synthétique de Fécondité en fonction du village de résidence de la femme, Niakhar, 1984-2011</i>	50

INTRODUCTION GENERALE

Ce rapport présente le travail effectué lors de mon stage à l'Institut de Recherche pour le Développement (IRD) au sein de l'unité mixte de recherche 198 « URMITE » (Unité de recherche sur les maladies infectieuses et tropicales émergentes). Ce stage de fin d'étude d'une durée de quatre mois s'est déroulé du 16 avril au 16 août 2012. Il a permis de m'ouvrir au monde professionnel et d'y appliquer les compétences acquises tout au long de mon cursus universitaire.

Au sein de cette unité de recherche qui s'occupe des suivis démographiques et sanitaires de plusieurs zones du Sénégal mon travail était d'analyser les données du système de suivi démographique de la zone de Niakhar, zone rurale du Sénégal. Il consistait plus précisément à la production de programmes d'analyse réalisés sous le logiciel Stata et à la rédaction d'un rapport d'analyse démographique.

L'analyse s'est orientée sur le thème de la fécondité avec pour but de mettre à jour l'évolution de la fécondité dans la zone de Niakhar sur la période du suivi démographique, 1983-2011. Pour y parvenir il m'a fallu mettre en place des programmes spécifiques à l'analyse des données de suivi démographique en employant la méthode d'analyse statistique des biographies. Puis par le biais de procédures adéquates les indicateurs transversaux de la fécondité ont pu être obtenus tels les taux de fécondité et l'indice synthétique de fécondité.

Une première phase de familiarisation m'a permis de prendre connaissance des données du suivi démographiques de Niakhar, du contexte d'étude ainsi que de la méthode statistique d'analyse des biographies mis en œuvre sous le logiciel Stata. Cette étape m'a permis par la suite de mener du mieux possible la mission qui m'avait été confiée. Le travail le plus complexe a été celui de la production des programmes d'analyse. La démarche d'analyse occupe d'ailleurs une place importante dans le rapport d'analyse de la fécondité, le cheminement des procédures y est commenté. A travers ce travail j'ai pu aboutir à des résultats d'analyse et étudier l'évolution de la fécondité de 1983 à 2011 ainsi que différents déterminants de la fécondité.

Afin de restituer le travail effectué durant ce stage, le présent rapport est scindé en deux grandes parties. La première partie présente le contexte du stage dans son ensemble avec d'une part la présentation de l'IRD et de l'UMR 198 « URMITE », puis d'autre part la mission du stage et son déroulement. La deuxième partie se compose du rapport d'analyse de la fécondité. Dans cette section un point est consacré à la démarche d'analyse et l'autre point aux résultats de l'analyse de la fécondité.

PREMIERE PARTIE

Cette première partie présente le contexte du stage dans son ensemble avec une présentation de l'organisme d'accueil, de la mission du stage et de son déroulement.

I _PRESENTATION DE L'ENTREPRISE

1.1 L'Institut de Recherche pour le Développement (IRD)



L'IRD est un organisme de recherche français, il succède en 1998 à l'ORSTOM (Office de la Recherche Scientifique et Technique d'Outre-mer) créé en 1944. C'est un établissement public à caractère scientifique et technologique, placé sous la double tutelle des ministères chargés de la Recherche et de la Coopération. Il mène ses projets en partenariat¹ avec d'autres organismes nationaux et étrangers.

Cet institut a pour vocation de mener des activités de recherche, de formation et d'innovation dont l'objectif est de contribuer au développement social, économique et culturel des pays du Sud. Il privilégie l'interdisciplinarité et centre ses recherches, depuis plus de 65 ans, sur les relations entre l'homme et son environnement en Afrique, Méditerranée, Amérique latine, Asie et dans l'Outre-Mer tropical français. L'IRD traite de questions cruciales sur les maladies tropicales et de civilisation, les relations entre santé et environnement, les changements climatiques, les ressources en eau, la sécurité alimentaire, les écosystèmes tropicaux et méditerranéens, les risques naturels, la pauvreté, la vulnérabilité et les inégalités sociales, les migrations, l'évolution du marché du travail, etc.

L'institut s'appuie sur son réseau de représentations implantées à l'étranger pour organiser ses collaborations. Il déploie ses activités à l'international depuis son siège, à Marseille, et ses deux centres métropolitains de Montpellier et de Bondy. Il dispose de 5 implantations dans les régions et collectivités d'Outre-mer (Réunion, Guyane, Martinique, Nouvelle-Calédonie et Polynésie française) et de 23 implantations dans des pays situés essentiellement dans la zone intertropicale (cf. carte page suivante).

La gouvernance de l'institut est confiée à un président. Pour réaliser ses missions, il s'appuie sur trois directions générales déléguées :

- la Direction générale déléguée aux ressources

Elle regroupe les directions et missions opérationnelles de l'IRD (ressources humaines, finances, affaires juridiques, systèmes d'information, communication interne, qualité, santé et sécurité au travail) ;

- la Direction générale déléguée à la science

Elle coordonne l'activité de recherche dans les domaines de l'Environnement et des Ressources, de la Santé et des Sociétés.

¹ L'institut dispose d'un dispositif partenarial en France et à l'étranger. Les principaux partenaires de l'IRD sont :

- au plan international : Agence nationale de recherche sur le sida et les hépatites virales (ANRS), Centre de coopération internationale en recherche agronomique pour le développement (CIRAD), Centre de Recherche Halieutique Méditerranéenne et Tropicale (CRH), Centre national de la recherche scientifique (CNRS), etc.

- au plan sénégalais : Universités (UCAD, UGB, Thiès, Ziguinchor), Centre hospitalier national universitaire de Fann (CHNU), Institut Pasteur de Dakar (IPD), Centre de suivi écologie (CSE), Institut de Technologie alimentaire (ITA), Académie Nationale des Sciences et Techniques du Sénégal (ANSTS), Centre de Recherche sur les Politiques sociales (CREPOS), etc.

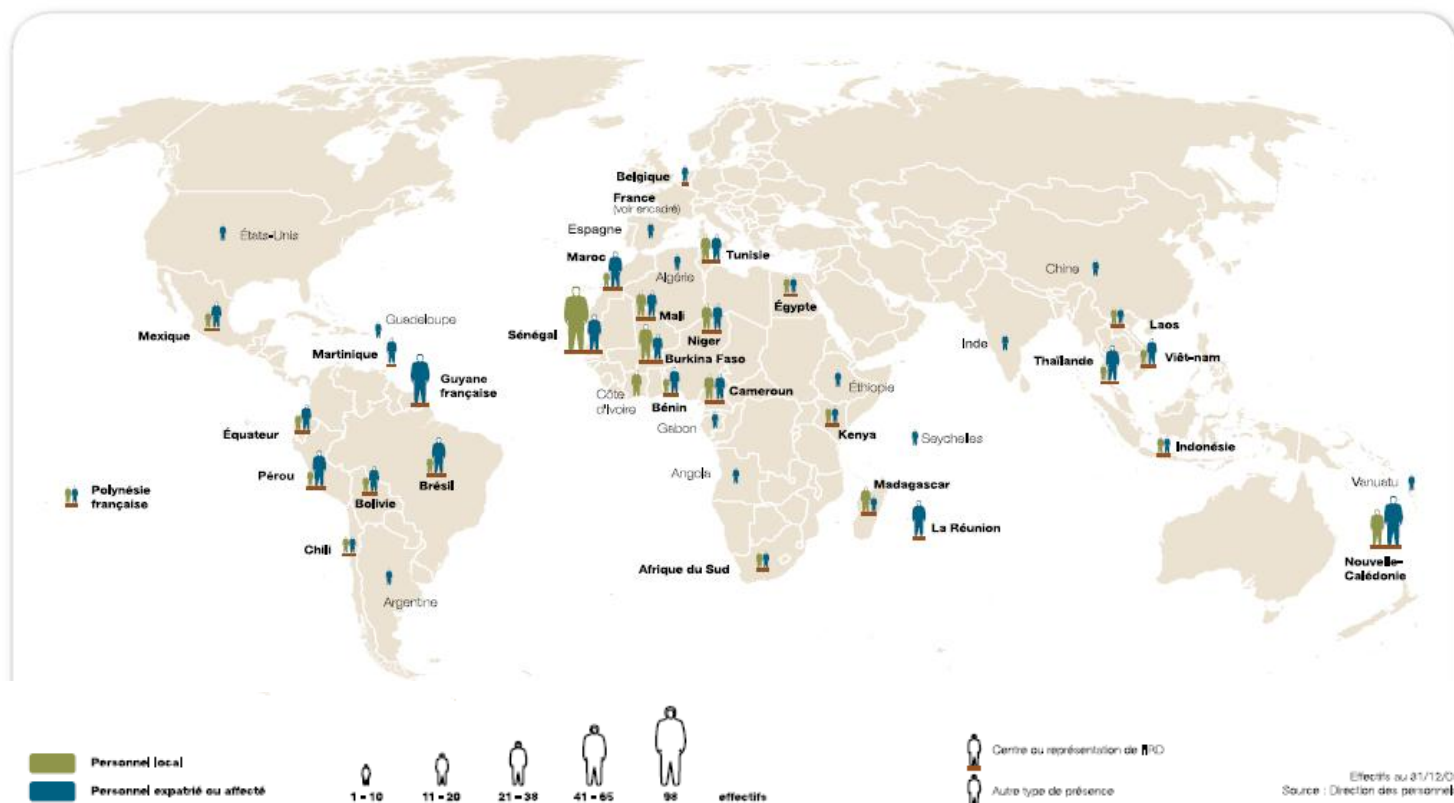
Pour plus d'information sur les partenaires de l'IRD : <http://www.ird.fr/les-partenariats/nos-principaux-partenaires>

Les thématiques de recherche relèvent de trois départements :

- le département environnement et ressources,
- le département sociétés,
- le département santé ;

- la Direction générale déléguée à l'Agence Inter-établissements de Recherche pour le Développement (AIRD)

Entre les trois directions générales déléguées et le Président, il existe aussi un cabinet et une mission de la géostratégie et du partenariat.



L'IRD dans le monde (Rapport de la Représentation au Sénégal, avril 2011, p.5)

Ce centre de recherche emploie plus de 2 200 agents dont plus de 800 chercheurs, 1 000 ingénieurs et techniciens et plus de 300 personnels locaux. L'accès aux filières d'emplois de l'IRD pour des postes de chercheurs, ingénieurs et techniciens se font par différentes voies : concours, sans concours (sur certaines conditions), voie contractuelle, volontariat civil².

Sur le plan financier l'IRD bénéficie principalement des budgets récurrents des tutelles ainsi que des aides financières de la part de ses partenaires. La communication³, quant à elle, occupe une place importante. En externe, les résultats des recherches menées à l'IRD sont publiés sous forme d'articles dans des revues scientifiques, mais également sous forme d'ouvrages, de cartes, de communications dans des colloques, etc. En interne l'IRD publie aussi avec notamment un rapport d'activité annuel.

² Les postes proposés sont accessibles sur le site de l'IRD www.ird.fr à la rubrique « nous-rejoindre », avec les explications des démarches à faire pour candidater en fonction du poste.

³ Photographies, films et vidéos, supports papiers et en ligne sont différentes ressources de l'IRD accessibles sur le site dans la rubrique « médiathèque ».

Ayant effectué mon stage au sein de l'IRD au Sénégal, je vous présente celui-ci avec plus de détails.

❖ **La représentation de l'IRD au Sénégal** est actuellement la plus importante implantation à l'étranger en termes de programmes, d'effectifs et de budget. Elle étend sa compétence sur quatre autres pays d'Afrique de l'Ouest : les îles du Cap Vert, la Gambie, la Guinée Bissau et la Mauritanie. Les agents de l'IRD sont accueillis en expatriation au Sénégal et en mission de courte durée dans les pays sous compétence. Elle compte un effectif de 38 expatriés et de 94 CDI locaux (juin 2011). Son siège en centre-ville abrite les services administratifs, la communication et l'Agence comptable secondaire. Le dispositif scientifique est situé sur des campus en co-gestions, interdisciplinaires : le campus UCAD-IRD de Hann (celui dans lequel j'ai effectué mon stage, il est présenté dans la sous-partie suivante), le centre commun ISRA/IRD de Bel-Air, le centre de Mbour en cours d'évolution avec l'installation de Aims-Sénégal, sans oublier la station de Niakhar qui constitue le plus ancien système de surveillance sanitaire et démographique d'Afrique et les stations de Dielmo et Ndiop.

L'IRD au Sénégal s'appuie sur des équipes de recherche rattachées à 17 UMR/UMI. Les activités de recherches se regroupent essentiellement autour de 6 pôles relevant de la Direction générale déléguée à la Science (DGDS) :

- Le pôle des sciences de l'univers,
- Le pôle de modélisation et outils d'observation
- Le pôle « Biologie des sols tropicaux »
- Le pôle « Biologie et écologie aquatique »
- Le pôle « Paludisme, VIH-Sida, maladies associées et maladies émergentes »
- Le pôle « Lutte contre la pauvreté »

(Cf. Schéma Annexe D).

Il existe la présence d'un ensemble de laboratoires très spécialisés dont celui qui m'a accueilli en stage : l'UMR 198 URMITE, présentée dans la suite de ce rapport.

1.2 Le campus international UCAD-IRD de Hann et son UMR 198 URMITE



C'est principalement sur le **campus international UCAD-IRD de Hann** que j'ai effectué les quatre mois de mon stage. Ce campus abrite 20 bâtiments construits par l'IRD, dont le bâtiment H3 qui accueille une partie de l'UMR 198. Ces bâtiments hébergent des services communs de l'administration, des laboratoires et des bureaux de plusieurs implantations secondaires d'unité de recherche. Ils hébergent aussi des chambres de passages pour l'accueil de stagiaires venant de l'étranger, comme ce fut le cas pour moi.

Les équipes de recherche implantées sur ce campus sont des équipes mixtes de recherches UCAD-IRD. Elles sont rattachées à deux écoles doctorales de l'UCAD : Sciences de la vie, de la santé et de l'environnement (Sev), Etudes sur l'homme et la société (Ethos)⁴.

Les équipes appartiennent aux unités suivantes :

- l'UMR 216, mère et enfant face aux infections tropicales (ED-Sev),
- l'UMR 225, développement, institutions et mondialisation (ED-Ethos),

⁴ site de l'IRD : <http://senegal.ird.fr>, rubrique des implantations principales.

- l'UMR 196 centre population et développement (ED-Ethos),
- l'UMR 151, laboratoire population environnement, développement (ED-Ethos),
- l'UMR 198, unité de recherche sur les maladies infectieuses et tropicales et émergentes (ED-Sev),
- l'UMR 208, patrimoines locaux (ED-Ethos).

L'UMR 198 URMITE, l'unité de recherche sur les maladies infectieuses et tropicales émergentes est une unité mixte : CNRS, Inserm, IRD, Aix-Marseille Université. Elle fait partie du département « Santé » de l'IRD. L'objet principal de ses recherches est la découverte et l'identification moléculaire de pathogènes émergents, l'étude des insectes vecteurs et la recherche thérapeutique. L'UMR gère en collaboration avec l'UCAD, la LSHTM (London School of Hygiene & Tropical Medicine) et le ministère de la Santé et de la Prévention du Sénégal le système de surveillance démographique et sanitaire des districts de Fatick, Bambey et Mbour. Son programme repose en partie sur le suivi étroit, continu et de longue durée de populations sénégalaises dans les principaux sites d'études que sont : Dielmo, Ndiop, Niakhar, Mlomp et Bandafassi. La connaissance de la population et de son évolution par le biais des analyses démographiques est donc essentielle pour permettre ces recherches de santé, qui portent principalement sur l'étude du paludisme endémique.

Son directeur est Didier Raoult, ses responsables sont Jean-François Trape, directeur de recherche et Cheikh Sokhna, chargé de recherche. L'équipe de l'UMR 198 se compose de 38 chercheurs, 26 ingénieurs et techniciens. Son implantation principale est située à la Faculté de médecine La Timone à Marseille. Son implantation secondaire est située à l'IRD au Sénégal, accueillie principalement sur le Campus international UCAD-IRD de Hann à Dakar. Cette unité dispose en outre de la plate-forme de Niakhar ainsi que des stations de Dielmo et Ndiop et les observatoires de santé et de population de Mlomp et Bandafassi.



Bureau des stagiaires et doctorants, bâtiment H3, Campus Ucad-IRD de Hann, Dakar, juin 2012



Station de Niakhar, juin 2012

II_PRESENTATION DE LA MISSION

2.1 La mission et son but

La mission de stage avait pour but initial d'analyser les données du Système de Suivi Démographique de Niakhar au Sénégal, avec notamment l'analyse de la fécondité, de la nuptialité et de la migration par la mise à jour des tendances, l'analyse des variations saisonnières et l'analyse de

déterminants. Elle consistait aussi à la production de programmes d'analyse sous Stata et à la rédaction d'un rapport d'analyse. Au cours du stage la démarche d'analyse et la vérification des données se sont révélées plus longues que prévu et ont entraîné la redéfinition de la mission. L'analyse s'est alors orientée sur la fécondité, son évolution et ses déterminants.

Comme il l'a été mentionné précédemment l'UMR 198 URMITE au Sénégal supervise les systèmes de suivi démographique (SSD) de Niakhar, Mlomp et Bandafassi. C'est dans le cadre de cette surveillance démographique que ma mission de stage a été confiée.

Cette mission est utile pour connaître l'évolution de la fécondité de la zone rurale de Niakhar entre 1983 et 2011 et de repérer certains déterminants sociaux économiques pouvant influencer le niveau de fécondité. Ce travail pourra permettre une certaine planification et un approfondissement des recherches en lien avec la baisse de la fécondité étudiée.

La connaissance de la population et de son évolution disponible par le biais des analyses démographiques est essentielle pour permettre les recherches en santé.

Les intérêts de cette unité de me confier cette mission résidaient principalement à avoir des analyses sur le suivi démographique du site de Niakhar. De plus le travail de ce stage est utile sur le plan de la démarche d'analyse, celle-ci pourra servir d'exemple pour d'autres analyses de suivi démographique réalisées avec la méthode d'analyse des biographies.

De mon côté ce stage a été enrichissant professionnellement et personnellement. Il est une première expérience professionnelle dans la mise en pratique de mes connaissances et compétences acquises lors du cursus universitaire. Il m'a permis l'apprentissage de nouvelles méthodes statistiques et la poursuite de l'utilisation du logiciel Stata. De plus ce stage m'a permis de me former sur l'identité numérique, les recherches d'articles scientifiques et le logiciel de bibliographie Zotero par le biais d'une formation intitulée « L'internet en Sciences humaines et sociales : repère, méthodes, et enjeux ». Ce stage a aussi été formateur sur le plan relationnel avec l'intégration au sein de l'URMITE. Travailler hors du monde universitaire m'a permis d'acquérir une plus grande autonomie dans le travail ainsi qu'un sens d'adaptation et d'organisation plus grand. Un des principaux apports dans ce stage a été la connaissance des systèmes de suivi démographique (SSD) avec en particulier la découverte du fonctionnement du suivi démographique de la zone de Niakhar. J'ai eu l'opportunité de me rendre sur le terrain d'étude et d'accompagner les enquêteurs faire le suivi démographique. Cette ouverture dans le monde de la recherche me fait réfléchir quant à mon avenir professionnel (poste de volontaire international, thèse).

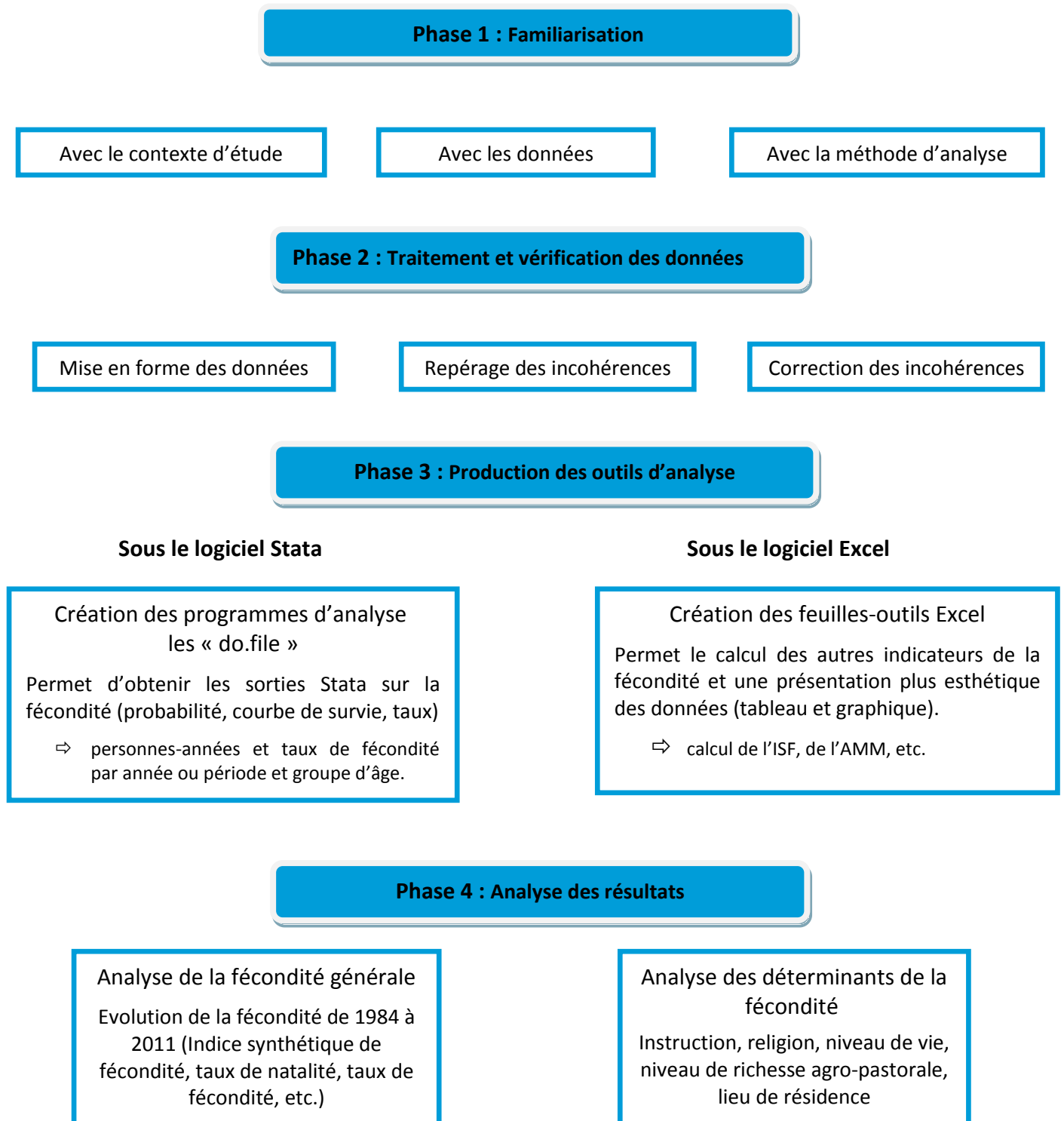
2.2 Le déroulement de la mission

La mission du stage a commencé par la phase de familiarisation. J'ai dû me familiariser avec le contexte de l'étude, les données à analyser et la méthode d'analyse.

Une fois cette phase de familiarisation effectuée, j'ai commencé le traitement des données. Cette étape a principalement consisté à la mise en forme des données dans des tables sous Stata et à la vérification des données. La vérification des données s'est révélée utile puisque une anomalie dans la base de données a été détectée, certaines dates n'étaient pas correctes à cause d'une erreur de formatage de ces dernières au moment de la saisie des données. Cette anomalie inversait le mois et le jour de certaines dates. Il a donc fallu passer un temps sur le repérage des incohérences dans les dates et sur la correction de ces dates dans la base de données. Je suis ensuite passé à la production des programmes permettant d'obtenir les principaux indicateurs de la fécondité par le biais de la méthode d'analyse des biographies. Après avoir obtenu les sorties de l'analyse de la fécondité sous Stata, j'ai créé des fiches outils sous Excel pour permettre le calcul d'autres indicateurs de fécondité.

Les étapes précédentes m'ont permis d'arriver au but voulu, c'est-à-dire à l'analyse des résultats avec l'analyse de l'évolution de la fécondité de 1984 à 2011 et l'analyse des déterminants de la fécondité.

Les principales étapes du déroulement de la mission de stage :



Les résultats de ce travail sont présentés dans la partie suivante.

DEUXIEME PARTIE

RAPPORT D'ANALYSE

Cette deuxième partie est le résultat du travail effectué dans le cadre du stage.

ANALYSE DE LA FECONDITE

A PARTIR DES DONNEES DU SYSTEME DE SUIVI DEMOGRAPHIQUE

ZONE DE NIAKHAR, SENEGAL

par la méthode d'analyse statistique des biographies, sous le logiciel STATA

(Event History Analysis Approach)

Introduction

Ce rapport a pour objectif principal de proposer une analyse de la fécondité de la zone de Niakhar, zone rurale du Sénégal, entre 1983 et 2011. Les données analysées sont issues du système de suivi démographique (SSD) de l'observatoire de Niakhar. Comme toutes données des SSD, ces données sont complexes et demandent des analyses spécifiques utilisant des procédures qui lui sont propres. Une grande partie du rapport est consacré à la méthode d'analyse, celle-ci repose sur l'analyse démographique des biographies (de l'anglais « Event History Analysis Approach ») et est mise en œuvre sous le logiciel Stata. Les procédures produites y sont présentées et commentées afin de valoriser l'exploitation des données issues des SSD. Les statistiques mathématiques sur lesquelles la méthode d'analyse repose ne seront pas traitées dans ce travail. D'autres ouvrages, notamment ceux de Daniel Courgeau et Eva Lelièvre, fournissent les principes théoriques utiles à la base de cette technique d'analyse. La démarche d'analyse s'inspire de celle proposée par Philippe Bocquier⁵ sur l'analyse de la mortalité et de la migration. Cette technique est réalisée dans le but de standardiser la méthode d'analyse et de permettre de produire des indicateurs comparables d'une époque à l'autre entre les différents sites de suivi démographique appartenant au réseau Indepth⁶. La démarche d'analyse est appliquée ici aux données du suivi démographiques de Niakhar, mais peut servir de guide pour d'autres analyses de la fécondité, et plus particulièrement pour les analyses transversales. La première partie expose les principes de l'analyse des biographies, ses particularités et son rôle dans l'analyse des SSD. La deuxième partie est consacrée à la présentation des procédures utiles pour permettre l'analyse transversale de la fécondité. Elle comprend entre autre l'exploitation des données et leur mise en forme, avec un point important consacré à la vérification de la qualité des données (repérage des incohérences et/ou anomalies). La dernière partie présente les résultats de l'analyse de la fécondité de la zone de Niakhar entre 1983 et 2011 et de ses déterminants.

⁵ Philippe Bocquier est professeur au Centre de recherche en démographie et Sociétés (Centre DEMO) de l'Université Catholique de Louvain, Belgique. Il a effectué et effectuée encore des travaux sur l'analyse des biographies.

⁶ Indepth : International Network Demographic Evaluation of populations and Their Health, est un réseau mondial regroupant les membres de différents sites des systèmes de suivi démographique. INDEPTH vise à développer des échanges d'ordre méthodologique et scientifique et de faciliter les approches comparatives d'analyse des systèmes de suivi démographique (www.indepth-network.org).

I. La méthode d'analyse biographique

Différents noms sont employés pour désigner la méthode d'analyse démographique des biographies, on parle d'analyse des données biographiques ou d'analyse biographique, d'analyse des biographies et des transitions, d'analyse de survie, d'analyse de durée de séjour ou encore d'analyse d'histoires de vie. En anglais cette analyse porte le nom d'Event History Analysis (EHA).

1.1 Généralité

L'expression « **analyse des biographies** » désigne une méthode d'analyse des données longitudinales individuelles datées. Elle est utilisée dans l'étude des phénomènes pouvant être représentés comme des suites d'événements. Son originalité réside sur le fait qu'elle analyse des relations dans le temps entre les différents événements de la vie. Elle consiste à prendre en considération le temps qui s'écoule entre un instant de référence commun à tous les individus analysés et la date de l'événement observé (ou date de transition) ou la date de sortie de l'observation. Cette méthode nécessite une conceptualisation rigoureuse de la question étudiée, et demande de s'assurer d'une bonne qualité des données.

Les possibilités d'analyses des données biographiques reposent sur les méthodes statistiques traitant des variables de « durée de vie », au sens durée de vie jusqu'à l'évènement étudié. Certaines de ces méthodes sont purement descriptives (estimateur de Kaplan-Meier, estimateur de Aalen), d'autres sont explicatives comme le modèle à risques proportionnels de Cox.

En démographie, l'application la plus simple de cette analyse se fait sur l'analyse de la mortalité, évènement non renouvelable, où l'analyse permet d'estimer la durée de vie des individus. Cette méthode d'analyse se répand sur les autres phénomènes démographiques dont notamment la fécondité, la migration et la nuptialité.

Dans son ouvrage « L'analyse des enquêtes biographiques à l'aide du logiciel STATA » datant de 1996, Philippe Bocquier propose une méthodologie des données biographiques. Il travaille actuellement sur cette méthodologie pour la mettre à jour. Il explique que du point de vue méthodologique les événements sont des changements d'état. L'analyse de ces changements d'état repose sur la modélisation du risque, qui est la vitesse qui régit ces changements d'état.

Cette analyse se fait sur le logiciel Stata. Ce dernier est bien adapté à l'analyse des données biographiques car il permet de traiter certains aspects ne figurant pas dans d'autres logiciels. De plus il constitue pour les sciences sociales une alternative très performante sur micro-ordinateur (Lelièvre et Bringé, 1998, p. 23). Il possède également l'avantage d'être un système ouvert, où l'utilisateur peut incorporer et mettre à disposition ses propres fonctions.

Grâce aux méthodes d'analyse statistique des biographies il est possible non seulement de retracer l'évolution d'un phénomène, mais aussi de tenter d'expliquer cette évolution et d'en dégager les facteurs démographiques, économiques et sociaux.

En bref :

L'analyse de survie c'est l'observation au niveau individuel de l'arrivée d'un ou plusieurs phénomènes, mesurée par une durée. Les bases de l'analyse reposent sur les concepts de population soumise au risque et de troncatures.

Exemple sur les données de Niakhar : les durées analysées correspondent à la durée de vie des résidents présent dans la zone jusqu'à l'évènement étudié (par exemple l'« émigration » ou l'« immigration », ou encore jusqu'à l'évènement « accouchement » tel il en est le cas pour l'analyse de la fécondité).

Notons que la migration et la fécondité sont des évènements démographiques renouvelables et demandent la mise en place d'une **analyse par étapes successives** afin de prendre en compte les différents épisodes de résidence des individus.

Il existe différentes méthodes pour analyser les durées, les méthodes appliquées dans le cadre de ce travail sont présentées ci-dessous.

1.2 Estimateurs et modèles

On peut estimer la courbe de survie grâce à la **méthode de Kaplan-Meier (KM)**. Cette méthode est un des outils les plus efficaces de l'analyse exploratoire des biographies. Il permet de tenir compte des données tronquées à droite. Cette méthode non-paramétrique d'estimation de la fonction de survie (fonction de séjour) prend en compte les différents temps de survenu des évènements. C'est une estimation de proche en proche. La fonction de survie est réactualisée à chaque temps de survenue de l'évènement (décès, naissances). Cet estimateur consiste à calculer une probabilité de connaître l'évènement dans chaque intervalle de temps, et de construire ainsi une table d'extinction ou de transition (ou la courbe équivalente) qui s'interprète simplement comme la proportion de « survivants » dans un état donné à chaque durée de séjour.

Les courbes de Kaplan-Meier représentent la distribution de la durée avant la réalisation d'un évènement. Puisqu'elles ont une signification probabiliste, on peut y associer un **intervalle de confiance** qui tiendra compte des effectifs soumis au risque à chaque durée. L'intervalle de confiance à 95 %, dit de Greenwood, est donné sous forme chiffrée dans la table de survie et est représenté graphiquement par la commande suivante :

```
sts graph, ci yline(.5) tmax(80) title("Evènement Naissance - Femme 1983-2011")
```

Chaque courbe (ou son équivalent chiffré) décrit le comportement hypothétique d'une cohorte d'individus qui auraient connu les mêmes conditions de vie. Pour que cette courbe corresponde effectivement au comportement d'une cohorte, il faudrait pouvoir suivre les individus jusqu'à leur décès (lorsqu'ils sortent effectivement de la population soumise au risque) et que cette cohorte soit homogène selon toutes les caractéristiques dont pourrait dépendre la réalisation de l'évènement. Cela veut dire que le calcul suppose que la seule hétérogénéité dans une cohorte est introduite par l'âge auquel chaque individu connaît l'évènement. (Bocquier, 2012, document à paraître).

* **La méthode Nelson-Aalen**, est aussi une méthode descriptive. L'estimateur de Nelson-Aalen consiste à calculer les quotients instantanés pour chaque durée et à en faire la somme cumulée plutôt que de faire le produit des proportions de survivants à chaque durée comme on le fait pour construire les courbes de Kaplan-Meier. Le niveau de chacune des courbes des quotients cumulés ainsi produites n'est pas interprétable en lui-même, mais on pourra **comparer les pentes des courbes entre elles : au moment t, la pente est une estimation de l'intensité du**

phénomène étudié. Concrètement il s'agit de calculer l'intensité cumulée (la somme cumulée des quotients instantanés) pour chaque type d'évènement, et d'en tracer la courbe. L'estimateur d'Aalen est multivarié et mesure l'intensité relative de chaque type d'évènement. De ce fait il n'est pas soumis à l'hypothèse d'indépendance entre chacun des types d'évènements.

On peut aussi obtenir **les quotients instantanés (hazard rates)**, ces derniers servent de base de calcul à l'ensemble des courbes et tableaux des étapes précédentes. Ces quotients sont calculés à chaque intervalle de temps et peuvent être représentés par une courbe. Il faut spécifier dans la commande `sts graph` l'option `hazard` :

```
sts graph, by(perio) hazard tmax(50)
```

C'est cette procédure que l'on applique pour obtenir les courbes représentant l'intensité des naissances par groupe d'âge des femmes.

La commande originale `sts graph` produit des courbes de Nelson-Aalen avec l'option `cumhaz`, mais pas pour des risques concurrents. Il faudra donc d'abord produire des tables de quotients cumulés pour chacun des risques concurrents avec `sts list` pour ensuite produire un graphique dans un tableur.

Un certain nombre de tests non paramétriques sont disponibles pour évaluer la significativité de la différence entre courbes. La commande `sts test` mesure la différence globale entre les courbes sélectionnées qui, avec le test de Mantel, appelé encore *logrank test*, est significative ou non au seuil de 10 %, 5%, etc. Ce test fait l'hypothèse de proportionnalité des quotients entre les catégories de la variable testée (par exemple le niveau d'instruction).

* **Le test du log-rank** a pour but de voir si les différences de survie entre échantillons sont significatives.

Les hypothèses du test sont les suivantes :

- ⎧ H0 : les fonctions de survies sont identiques pour les différents échantillons
- ⎧ Ha : les fonctions de survies sont différentes pour les différents échantillons

Sous H0, les différents échantillons étudiés sont considérés comme étant issus de la même population. De ce fait, on les regroupe en un seul échantillon sur lequel on estime les effectifs attendus. Ces derniers seront comparés aux effectifs observés. On vérifie que les conditions d'application du test soient vérifiées notamment celle de l'indépendance entre la différence des fonctions de survie des différents groupes et le temps. Les courbes de survie ne doivent donc pas se croiser. Dans le cas contraire un tel test n'aurait pas de sens dû aux problèmes de compensation.

Sous l'hypothèse H0 d'identité des distributions, pour chaque instant de changement d'état (naissance d'un enfant dans notre cas), on calcule la proportion de naissance et les valeurs attendues telles que : $E_1 + \dots + E_n = D_1 + \dots + D_n = M$

Avec,

E_i : le nombre de naissance attendues au cours du temps sous l'hypothèse nulle dans le groupe i ;

D_i : le nombre total des naissances observées au cours du temps dans le groupe i

M : le nombre total de naissances, tous échantillons confondus.

On calcule ensuite la statistique de test qui suit asymptotiquement une loi du Khi-deux à un degré de liberté sous l'hypothèse H0. On choisit le risque de première espèce α (10%, 5% et 1%). On calcule ensuite la p-valeur qui correspond au calcul d'une probabilité sous H0.

Stata calcule toutes ces statistiques automatiquement, il suffit d'interpréter les résultats obtenus. C'est **l'étape de décision**, elle consiste à voir si la p-valeur est inférieure ou égale au risque de première espèce. Si c'est le cas on peut rejeter H_0 au risque d'erreur α (de se tromper dans notre décision) et donc considérer que la survie diffère dans les différents groupes. Sinon on ne peut pas rejeter H_0 au risque α et on considère que les différences observées sont dues aux fluctuations d'échantillonnage.

La technique des tables de transition et les diverses techniques associées (tests non paramétriques, courbes de Nelson-Aalen) apportent une description très riche de l'événement. La dimension temporelle y est pleinement intégrée : on se rapproche du principe de priorité temporelle à la base de la relation de causalité. C'est donc un pas qualitatif important pour l'analyse. (Bocquier, 2012, document à paraître).

Une fois cette étape descriptive bien maîtrisée, on peut introduire les facteurs explicatifs dans l'analyse de l'événement par l'application du modèle de Cox.

* **Le modèle de Cox** est un modèle de régression multivarié où l'on prend en compte des données censurées ainsi que la variabilité inter-individuel en termes de survie (délai de survie, caractéristiques, etc.). Il s'applique à toutes situations où l'on étudie le délai de survenue d'un événement, comme par exemple la naissance d'un enfant, la réponse à un traitement, le décès, etc. Il permet d'expliquer la survie étudiée en fonction de l'instant t et de plusieurs autres variables explicatives de façon simultanée, tels que des caractéristiques intrinsèques aux sujets (âge, sexe, taille, etc.) des facteurs de risque, des facteurs pronostiques, des traitements pris par les individus, etc. Ceci permet à l'inverse du test du log-rank exact, de contrôler les effets cachés de certaines variables, après construction d'un groupe de référence à partir duquel toutes les comparaisons seront faites. Le modèle est dit semi-paramétrique⁷. Cette approche fut proposée par Cox pour établir une relation paramétrique entre les facteurs de risque de décès et la distribution des durées de survie, sans donner une forme particulière aux fonctions de survie. Il repose sur deux hypothèses qui assurent sa validité : ce modèle est log-linéaire et les risques relatifs⁸ de ce modèle ne dépendent pas du temps, on dit qu'ils sont constants sur tout intervalle de temps. Ainsi le modèle de Cox, comme tout autre modèle possédant cette dernière caractéristique, est dit à risques proportionnels.

Le principe de ce modèle est donc de combiner analyse de survie et régression, avec une hypothèse importante celle de la proportionnalité.

Stata permet l'application de ce modèle par la procédure *stcox*, par exemple nous pouvons avoir le modèle suivant où la variable à expliquer est le niveau de fécondité et les variables explicatives sont le niveau d'instruction (edu1 edu2 edu3 eduA), le niveau de vie des femmes (nv1 nv2) et le niveau de richesse agro-pastorale (rap1 rap2).

```
stcox edu1 edu2 edu3 eduA nv1 nv2 rap1 rap2
```

Les coefficients obtenus pourront ainsi montrer l'influence de ces variables sur le niveau de fécondité des femmes. Ainsi dans le modèle consacré à l'événement naissance dans la vie de la femme, une variable qui augmente le risque d'avoir un enfant multiplie le risque de base par une quantité supérieure à 1 et, inversement, pour une variable qui diminue le risque, tout ça par rapport à la catégorie de référence.

⁷ Comme pour le test du log-rank il existe une approche paramétrique, que nous n'élaborerons pas dans ce projet, où les distributions des durées de survies suivent une loi connue.

⁸ **Risques relatifs** : rapports des risques instantanés pour les individus de caractéristiques différentes.

1.3 Conceptualisation : analyse de la fécondité de Niakhar

La **population soumise au risque** est constituée, dans un premier temps, des hommes et des femmes de tout âge considérés comme résidents dans la zone de Niakhar à une période donnée, entre 1983 et 2011. A partir de cette population, il sera possible d'obtenir les taux brut de natalité. Dans un second temps la population soumise au risque sera restreinte aux femmes dans le but d'obtenir les autres indicateurs de fécondité.

La **date du début de l'observation** est celle du 1^{er} janvier 1983 et la date de fin d'observation est la date du dernier passage de 2012. L'analyse porte seulement sur les données allant jusqu'à fin 2011, le **choix de la date de censure à droite** est le 1^{er} janvier 2012.

L'**événement étudié** est l'évènement « naissance d'un enfant vivant » pour les femmes résidentes dans la zone au moment de la naissance de l'enfant. Cet évènement est un évènement renouvelable, il faudra le préciser dans l'analyse grâce à l'option *id()* de la commande *stset* de Stata.

Afin d'analyser la durée des séjours de résidence des individus dans la zone de Niakhar, il faut considérer les entrées et sorties de la zone de ces individus.

Les **issues d'entrée** en observation sont le recensement initial (de 1983), la naissance, l'immigration (externe) et l'entrée (immigration interne).

Les **issues de sorties** d'observation sont le décès, l'émigration (externe), la sortie (émigration interne) et la sortie d'observation par date de fin d'observation de l'enquête. Dans le dernier cas, les individus sont considérés comme présent dans la zone, on leur attribue le code d'évènement 9 (PZ). D'autres évènements sont enregistrés pour certains individus comme par exemple pour les individus en visite mais ils ont été exclus de l'analyse puisque cette dernière porte sur les personnes résidentes dans la zone et non de passage.

Un des points essentiels à définir est l'évènement étudié et la durée de survie.

On s'intéresse à la probabilité pour qu'un individu (une femme) connaisse l'évènement « enfant » pendant sa durée de vie (résidence) dans la zone entre la période d'observation. Les estimateurs de la fonction de durée obtenus par la méthode de Kaplan-Meier correspondent à la probabilité de survie qu'a une femme de ne pas connaître l'évènement « enfant ».

L'**instant initial** ($t=0$) est commun à tous les individus et doit être défini sans ambiguïté (Lelièvre et Bringé, 1998, p.6). Dans cette étude l'instant initial est la date de naissance de l'individu. A partir de cette date sera calculée la durée de présence de l'individu dans la zone, son âge à l'entrée et son âge à la sortie. Rappelons que la première entrée dans l'observation se fait soit par Recensement initial, soit par Naissance, soit par Immigration. Prenons l'exemple du parcours de « vie » de la femme 21 (Identifiant individuel = 21), née le 28 septembre 1967 (DoB) et entrée en observation par le recensement initial (code ENU) le 24 février 1983 (Tableau 1).

Tableau 1 : Parcours de « vie » d'une femme dans la zone de Niakhar

Individual Id	Location Id	DoB	Sex	datebeg	EventDate	Event Code	residence	ChildId	Child Sex	calendar _year	age group	foll_Eve ntCode	delivery	_st	_d	_origin	_t	_t0
21	1	28/09/1967	2	28/09/1967	24/02/1983	ENU	0					01-janv	0	0				
21	1	28/09/1967	2	24/02/1983	31/12/1983	01-janv	1			1983	15	01-janv	0	1	0 2.443e+11	16,2	15,4	
21	1	28/09/1967	2	31/12/1983	01/01/1985	01-janv	1			1984	15	OMG	0	1	0 2.443e+11	17,2	16,2	
21	1	28/09/1967	2	01/01/1985	15/05/1985	OMG	1			1985	15	IMG	0	1	0 2.443e+11	17,6	17,2	
21	1	28/09/1967	2	15/05/1985	03/04/1987	IMG	0					NAG	0	0				
21	1	28/09/1967	2	03/04/1987	28/09/1987	NAG	1			1987	15	01-janv	0	1	0 2.443e+11	20,0	19,5	
21	1	28/09/1967	2	28/09/1987	31/12/1987	01-janv	1			1987	20	CHD	0	1	0 2.443e+11	20,2	20,0	
21	1	28/09/1967	2	31/12/1987	28/03/1988	CHD	1	60412	1	1988	20	01-janv	1	1	1 2.443e+11	20,5	20,2	
21	1	28/09/1967	2	28/03/1988	01/01/1989	01-janv	1			1988	20	CHD	0	1	0 2.443e+11	21,2	20,5	
21	1	28/09/1967	2	01/01/1989	29/08/1989	CHD	1	63037	1	1989	20	01-janv	1	1	1 2.443e+11	21,9	21,2	
21	1	28/09/1967	2	29/08/1989	31/12/1989	01-janv	1			1989	20	01-janv	0	1	0 2.443e+11	22,2	21,9	
21	1	28/09/1967	2	31/12/1989	31/12/1990	01-janv	1			1990	20	OMG	0	1	0 2.443e+11	23,2	22,2	
21	1	28/09/1967	2	31/12/1990	19/07/1991	OMG	1			1991	20	OBE	0	1	0 2.443e+11	23,8	23,2	
21	1	28/09/1967	2	19/07/1991	01/01/2012	OBE	0						0	0				

Cette femme a connu un premier séjour dans la zone de Niakhar de ses 15 à 17 ans. Ensuite elle a émigré hors de la zone en 1985 pour revenir en 1987 à l'âge de 19 ans. Son deuxième séjour a duré plus de 4 années. C'est à l'âge de 24 ans qu'elle quitte la zone d'observation avec une seconde émigration en 1991, cette fois-ci c'est une migration sans retour enregistrée jusqu'à présent. On peut voir que pendant son deuxième séjour dans la zone cette femme a eu deux enfants, un en 1988 et un en 1989.

La **durée étudiée**, durée d'exposition au risque, est la durée de vie (de résidence) jusqu'à ce que l'évènement étudié arrive, ou dans le cas contraire jusqu'à la fin de la date d'observation de l'enquête. Elle est calculée depuis l'instant initial jusqu'au moment de la transition ou de la troncature. En cas de transition, la durée d'exposition correspond au cumul de toutes les durées d'exposition au risque, c'est-à-dire les périodes (épisodes) de résidence dans la zone entre la date d'entrée et la date de sortie de la zone.

Le **temps** est la date de l'évènement, il faut une horloge commune.

Ici les dates sont données en heures, jours, mois, années afin d'avoir le plus de précisions possibles et d'éviter la simultanéité entre évènements qui peuvent arriver le même jour.

Avant de pouvoir effectuer des analyses sur les durées, il faut s'assurer de la qualité et de la structure des données. C'est une étape primordiale pour permettre l'analyse de données biographiques. **Il est fondamental d'aboutir à un recueil cohérent des dates de transition d'une étape à l'autre de la vie.** Il est donc indispensable, au moment de la collecte, de situer les événements dans chacune des dimensions les uns par rapport aux autres (par exemple, l'immigration de la femme par rapport à la naissance de son enfant). Pour cela, le juste recueil de la succession des événements importe.

Pour une présentation du système de suivi démographique de la zone de Niakhar et de ses données se référer à l'annexe E.

II. La démarche d'analyse et les procédures

Cette partie présente les procédures réalisées sous Stata et sous Excel permettant le calcul des indicateurs transversaux de la fécondité à partir des données de système de suivi démographique. Rappelons que l'analyse quantitative des biographies se fait par le biais de techniques d'analyse complexes mais efficaces. Elle permet le calcul des taux démographiques de base, comme le taux brut de natalité et les taux de fécondité dans le cas de l'analyse de la fécondité. De plus elle permet la prise en compte du temps dans l'explication causale.

Plusieurs étapes essentielles sont mises en place afin de permettre l'analyse des données du système de suivi sur les durées de séjours des individus de la zone de Niakhar jusqu'à l'évènement étudié. Pour analyser la fécondité et créer les indicateurs transversaux six grandes étapes sont distinguées. La première étape, étape générale pour analyser les données du suivi démographique, est de **mettre en forme la table standard**. Il faut s'assurer que la structure de la base corresponde à une structure permettant l'analyse de durée. La structure minimum des données requise dans la table standard est composée des variables suivantes :

- *Identifiant de l'individu,*
- *sexe,*
- *évènement d'entrée (dans la zone d'étude),*
- *évènement de sortie (de la zone d'étude),*
- *date de l'évènement d'entrée,*
- *date de l'évènement de sortie*

Et

- *Identifiant de la mère (si création de la table des naissances à partir des entrées par naissances dans la zone).*

Toutes autres informations, par exemple le lieu d'habitation (identifiant village, concession, cuisine), la religion, le niveau d'instruction, etc. peuvent aussi être nécessaires dans le but d'y faire des analyses explicatives.

Les variables de base des données de Niakhar sont :

id, datnais, sexe, daten, datex, entry, exit et idmere

avec

id = identifiant de l'individu

datnais = date de naissance

sexe = sexe de l'individu

daten = date d'entrée

datex = date de sortie

entry = évènement d'entrée

exit = évènement de sortie

idmere = identifiant de la mère

Ces variables sont renommées afin de garder la nomination employée dans les programmes de Philippe Bocquier sur le modèle standard (Tableau 2).

Tableau 2 : Codes des évènements

Code initial (Niakhar)	Recodage (standard Indepth ; sur le modèle de P. Bocquier)	Labels (standard)	Définition
11	1	ENU	Recensement Initial
12	2	BTH	Naissance (d'un individu d'une femme résidente)
13	3	IMG	Immigration
62	6	ENT	Entrée (Chgmt-Ad-Arr)
61	5	EXT	Sortie(Chgmt-Ad-Dep)
71	4	OMG	Emigration
51	7	DTH	Décès
91	9	OBE	Présent dans la Zone
8		CHD	Enfant*

* *L'évènement ENFANT (CHD) est ajouté dans la table des parcours de vie des femmes (différent de accouchement = DLV, code 10 dans les programmes de P. Bocquier). Il correspond à la naissance d'un enfant qu'a une femme au cours de sa période (ou ses épisodes) de vie dans la zone d'étude de Niakhar (code 12 dans la base de Niakhar).*

Une fois cette table standard créée l'analyse des durées de chaque séjours des individus pourra être réalisée grâce à la commande **stset** de Stata. Mais avant il est indispensable de vérifier la qualité des données, essentielle pour avoir des résultats significatifs.

Actuellement plusieurs incohérences sont présentes dans les données dues en grande partie à une anomalie dans la base de données (les jours et les mois de certaines dates ont été inversés) ainsi qu'à d'autres erreurs qui proviennent probablement de la saisie des données. Cette anomalie sur les dates a rendu l'analyse plus complexe et plus longue puisqu'un temps de travail a été consacré à la correction de ses incohérences dans l'espoir d'avoir des données les plus justes possible. Ce n'est qu'après cette étape de vérification et de correction des données qu'il est judicieux de mettre en place la procédure permettant le calcul des durées de vie des individus à l'entrée dans la zone puis à la sortie afin de connaître la durée qu'ils ont effectué dans et hors zone. Ensuite il reste à effectuer les coupes transversales, ici au 1^{er} janvier de chaque année, (et les coupes par groupe d'âge) pour obtenir le calendrier par année (et groupe d'âge) et ainsi permettre le calcul des indicateurs transversaux de la fécondité par année (et groupe d'âge).

La démarche d'analyse est expliquée plus clairement dans l'encadré qui suit.

PROCEDURE A SUIVRE POUR ANALYSER LA FECONDITE (par la méthode d'analyse des biographies sous le logiciel Stata) :

1 – Exécuter le fichier `structure_standard.do`

=> crée la table `residency.dta` et la matrice des incohérences.

Ce programme permet la conversion du fichier "RESIDENCY EPISODE FILE" (fichier initial avec les événements de début et de fin en épisode) nommé "yourepisodefile.dta" en fichier "RESIDENCY EVENT FILE" (fichier de base avec une observation par événement). Cette nouvelle structure de la table est plus appropriée pour l'analyse des biographies, P. Bocquier parle de fichier au format EHA (Event History Analysis).

=> **On passe d'un fichier avec des épisodes à un fichier par événements (1 événement par ligne).**

2 – Exécuter le fichier `censure_OBE.do`

=> ajoute l'observation de fin pour chaque individu à la date de censure choisie (événement nommé OBE).

=> **Ce programme permet de créer la variable de censure à droite (à la date de fin de l'observation).**

3 – Exécuter le fichier `création_résidence.do`

=> crée la variable "residence" afin de distinguer les périodes de séjours des individus dans la zone et hors de la zone (1 si présence et 0 sinon).

4 – Exécuter le fichier `merge_fécondité.do`

=> crée la table `resid_fec.dta`, i.e. la table avec les événements "naissance" en plus des autres événements dans le parcours des femmes présentes dans la zone d'étude.

=> **Ajout de l'évènement naissance (enfant) dans le parcours des femmes**

5 – Exécuter le fichier `stplit_calendrier_année.do`

=> **crée le calendrier par année** (coupe chaque séjour en années), et par groupe d'âge.

6 – Exécuter le fichier `analyse_fécondité.do`

=> **Analyse de la durée** de vie sans enfant des femmes et de la **distribution des naissances** selon l'âge des femmes à la naissance de l'enfant. Sorties des taux brut de natalité, taux de fécondité par années et autres indicateurs.

!!! Avant de commencer il est utile d'ajouter de la mémoire

```
set mem 1000000
```

et d'avoir attribuer le code 91 aux individus présents dans la zone au dernier passage, en attente du prochain, (ici pour ces individus datex=1jan2015) dans la table initiale. !!!

2.1 Mise en forme de la table « RESIDENCY.DTA »

La table standard « residency.dta » est créée en exécutant le programme du fichier « structure_standard.do » (ANNEXE A.1).

Cette table comprend une observation par ligne (1 évènement = 1 ligne), et non plus un épisode avec entrée et sortie sur la même ligne comme c'est le cas dans la table initiale.

Pour créer cette table standard, il nous faut renommer et créer certaines variables. Les variables renommées sont « EventCode » (code de l'évènement), et « EventDate » (date de l'évènement) ; les variables créées sont « datebeg » (date de l'évènement précédant), « foll_event » (évènement suivant) et « file » (file=1 si évènement de sortie ; file=2 si évènement d'entrée).

On obtient la table standard (Tableau 3) avec les variables de bases *IndividualId* (identifiant de l'individu), *DoB* (date de naissance de l'individu), *Sex* (sexe de l'individu), *EventDate* (date de l'évènement), *EventCode* (code de l'évènement), *file* (1 si sortie, 2 si entrée), *datebeg* (date de début, c'est-à-dire date de l'évènement précédant), *foll_event* (évènement suivant), *same_event* (pour repérer si deux même évènement se suivent), etc.

Tableau 3 : Aperçu de la table residency.dta

	IndividualId	LocationId	DoB	Sex	datebeg	EventDate	EventCode	IdMere	file	EventCode_~1
1	2	1	03jan1926 12:00:00	2	03jan1926 12:00:00	24feb1983 12:00:00	ENU	-1	2	11
2	2	1	03jan1926 12:00:00	2	24feb1983 12:00:00	18jun1987 12:00:00	DTH	-1	1	51
3	3	1	08feb1968 12:00:00	2	08feb1968 12:00:00	24feb1983 12:00:00	ENU	2	2	11
4	3	1	08feb1968 12:00:00	2	24feb1983 12:00:00	15jun1984 12:00:00	OMG	2	1	71
5	3	1	08feb1968 12:00:00	2	15jun1984 12:00:00	28mar1986 12:00:00	IMG	2	2	13
6	3	1	08feb1968 12:00:00	2	28mar1986 12:00:00	15feb1989 12:00:00	OMG	2	1	71
7	5	1	13oct1963 12:00:00	2	13oct1963 12:00:00	24feb1983 12:00:00	ENU	2	2	11
8	5	1	13oct1963 12:00:00	2	24feb1983 12:00:00	15apr1983 12:00:00	OMG	2	1	71
9	7	1	21feb1958 12:00:00	1	21feb1958 12:00:00	24feb1983 12:00:00	ENU	-1	2	11
10	7	1	21feb1958 12:00:00	1	24feb1983 12:00:00	06jan1989 12:00:00	DTH	-1	1	51
11	9	1	23apr1928 12:00:00	2	23apr1928 12:00:00	24feb1983 12:00:00	ENU	-1	2	11
12	9	1	23apr1928 12:00:00	2	24feb1983 12:00:00	13nov2003 12:00:00	DTH	-1	1	51
13	11	1	24aug1964 12:00:00	1	24aug1964 12:00:00	24feb1983 12:00:00	ENU	1051220	2	11
14	11	1	24aug1964 12:00:00	1	24feb1983 12:00:00	01jan2015 12:00:00	OBE	1051220	1	91
15	12	1	11jan1962 12:00:00	2	11jan1962 12:00:00	24feb1983 12:00:00	ENU	9	2	11
16	12	1	11jan1962 12:00:00	2	24feb1983 12:00:00	15jun1984 12:00:00	OMG	9	1	71
17	12	1	11jan1962 12:00:00	2	15jun1984 12:00:00	09oct1986 12:00:00	IMG	9	2	13
18	12	1	11jan1962 12:00:00	2	09oct1986 12:00:00	15jun1989 12:00:00	OMG	9	1	71
19	12	1	11jan1962 12:00:00	2	15jun1989 12:00:00	07aug1992 12:00:00	IMG	9	2	13
20	12	1	11jan1962 12:00:00	2	07aug1992 12:00:00	15jan1994 12:00:00	OMG	9	1	71
21	12	1	11jan1962 12:00:00	2	15jan1994 12:00:00	01aug1994 12:00:00	IMG	9	2	13
22	12	1	11jan1962 12:00:00	2	01aug1994 12:00:00	01jan2015 12:00:00	OBE	9	1	91
23	17	1	20oct1906 12:00:00	2	20oct1906 12:00:00	24feb1983 12:00:00	ENU	-1	2	11
24	17	1	20oct1906 12:00:00	2	24feb1983 12:00:00	25dec2003 12:00:00	DTH	-1	1	51

Une fois cette structure établie, un travail d'exploration des données est nécessaire dans le but de repérer les incohérences des données et de les résoudre au maximum. Les incohérences sont souvent dûes à des problèmes de dates. Par exemple il arrive que les dates entre évènements

courants et précédants soient similaires ce qui entraîne des problèmes dans l'analyse. Ainsi si la date de l'évènement naissance et la même que la date de l'évènement décès (c'est-à-dire que le décès a eu lieu le jour de la naissance) cela entraîne des problèmes pour l'analyse. Le choix a été fait d'ajouter 6 heures à la date de décès, afin de considérer tout de même ces deux évènements.

Après avoir résolu, dans la mesure du possible, une première partie des incohérences, nous pouvons créer la matrice des incohérences afin de repérer d'autres incohérences dans les données. Il est essentiel de faire cette matrice pour calculer le pourcentage d'erreur des données et connaître le biais de l'analyse. Il est recommandé de copier dans une feuille de calcul Excel la matrice obtenue et de calculer le nombre de données avec incohérences (Annexe B.1).

2.2 Création de la variable de censure à droite

Cette étape consiste à créer la date de censure à droite. Le programme se trouve dans le fichier « censure_OBE.do » (Annexe A.2).

2.3 Création de la variable résidence

Il est question ici de créer la variable residence dans la table standard. La variable résidence est très importante puisque c'est elle qui définit si la personne est résidente ou non dans la zone d'observation. Le programme qui permet de créer cette variable se trouve dans le fichier « residence.do » (Annexe A.3).

Elle indique si l'individu est entré dans la zone ou sorti de la zone d'étude (s'il est présent, sorti ou entrée dans la zone), residence=1 si l'individu est sorti (décès, émigration, sortie) et residence=0 si l'individu est entrée (RI, Nais, Im, Entrée).

2.4 Création de la table fécondité

Jusqu'à présent dans la table standard ne figure que les évènements d'entrées et de sorties de la zone ainsi que leur dates. L'évènement « enfant » des femmes n'apparaît pas, il doit donc être ajouté en tant qu'évènement dans la vie des femmes. Il faut travailler préalablement sur le fichier des naissances afin de reconstituer les différentes naissances qu'a pue connaître une même femme dans la zone d'étude et ainsi obtenir la table resid_fec.dta. Le programme détaillé se trouve dans le fichier « création_fécondité.do » (Annexe A.4).

Le principe est de fusionner ce fichier des naissances avec le fichier contenant les autres évènements, afin d'obtenir dans un fichier unique l'ensemble des évènements vécus (dans la zone de niakhar) par l'individu, classés dans le temps.

Ce travail informatique nécessite des procédures spécifiques de fusion des données dans le temps. Rappelons qu'elles ont été testées et mises au point par Philippe Bocquier en 1996 sur le logiciel Stata (Bocquier, 1996).

2.5 Création du calendrier par année et groupe d'âge

Cette cinquième étape consiste à effectuer des coupes transversales avec la commande `stsplit` à date fixe et à âge fixe. Elle est nécessaire par la suite pour pouvoir reconstituer les indicateurs transversaux (Annexe A.5).

2.6 Création de la variable de censure et analyse de la fécondité

Il faut calculer un indice de censure correspondant au risque étudié (ici la censure est faite sur l'évènement « enfant » (ou `delivery`), `delivery=1` si l'évènement est arrivé, sinon `delivery=0`).

Ensuite nous pouvons commencer l'analyse de survie sur la fécondité sous Stata (Annexe A.6).

Stata impose une série de commande pour l'analyse des biographies (avec le préfixe `st` pour *survival time analysis*) qui ne peuvent être exécutées qu'après avoir déclaré le fichier comme approprié pour ce type d'analyse. La procédure `stset` de Stata est le point de départ de l'analyse. Cette commande sert à identifier les variables de durée d'observation et de censure afin de connaître la durée du (ou des) séjour(s) de l'individu entre l'entrée et la sortie de la zone en considérant les censures.

Dans le cas de l'analyse de la fécondité des femmes résidentes, la procédure est la suivante :

```
stset EventDate if residence==1 & Sex==2, id(IndividualId) failure(delivery==1) ///
      origin(time DoB) time0(datebeg) exit(time .) scale(31557600000)
```

L'option `id(IndividualId)` est indispensable pour indiquer que chaque individu peut avoir plusieurs observations (périodes) reliées entre elles par un identifiant (`IndividualId`). Le temps de début de chaque période est également indiquée par l'option `time0(datebeg)`. La durée ici est exprimée en année depuis la naissance.

Cette commande fournit un certain nombre d'informations (Tableau 4).

Tableau 4 : Détails des informations fournies par la commande `stset` sous Stata

id	IndividualId	Variable identifiant les observations relatives à un même individu.
failure event	delivery == 1	Condition définissant la censure à droite.
Obs. time interval	(datebeg, EventDate)	Variable définissant le temps de début et de fin de période. La fin de la période inclut pour le calcul de la population au risque l'unité de temps indiquée en <code>EventDate</code> . Cette unité est exclue au début de la période suivante en <code>datebeg</code> .
exit on or before	failure	La censure se produit à l'évènement.
t for analysis	time/31557600000 (cad en année)	Les résultats seront présentés dans l'unité de temps de <code>EventDate</code> , divisé par 365,25 (dans notre exemple, en année). Les calculs se font avec l'unité de temps original.
if exp	residence==1 & Sex==2	Les observations du fichier pour lesquelles <code>residence==1 & Sex==2</code> sont soumises à la commande <code>stset</code> , les autres non. Ne pas confondre avec l'option <code>if(exp)</code> qui permet notamment de traiter les sorties temporaires d'observation.
total obs.	1314012	Nombre de lignes d'observations dans le fichier.
ignored at outset due to -if <exp>-	687344	Nombre de lignes d'observations exclues avant calculs en raison de <code>if exp</code> .
Obs. remaining	626079	Nombre de lignes d'observations restantes pour l'analyse.
Subjects	40181	Nombre d'individus représentés pour l'analyse.

failures in multiple failure-per-subject data	36327	Nombre d'événements connus par les individus représentés pour l'analyse. Chaque individu peut connaître plusieurs événements.
total analysis time at risk from t = 0	437728,6	Nombre d'unités de temps d'analyse vécues par les individus représentés pour l'analyse.
earliest observed entry t	0	Temps de première entrée en observation.
last observed exit t	111,3945	Temps de dernière sortie d'observation (remarquez que l'unité de temps est celle de l'analyse, ici l'année).

La procédure *stset* permet de créer une série de variables :

=> **_st** indique si l'observation est prise en compte (code 1) ou non (code 0) pour l'analyse
=> **_t0** indique le temps de début de période (âge de l'individu à l'instant d'entrée)
=> **_t** indique le temps de fin de période (âge de l'individu à l'instant de sortie).
La date d'origine ici est la date de naissance de l'individu.
=> **_d** indice de censure (0 ou 1)
(Variable d'interruption, elle permet de savoir si l'individu connaît ou non l'évènement durant l'observation (ici 1 si décès et 0 sinon)).

Après avoir effectué la procédure *stset*, il est possible d'obtenir la table de survie par la commande *stslst*, les courbes de survie par la commande *sts graph*, ainsi que les taux de fécondité par la commande *stptime*.

Il existe plusieurs courbes de survie comme nous avons pu le présenter en première partie, la plus courante est la courbe de Kaplan-Meier. Celle-ci s'obtient par la commande simple *sts graph* et présente la proportion de survivants. D'autres courbes peuvent être obtenues grâce à des options dans cette commande. Il est possible d'avoir la courbe de Kernel qui présente des séries de quotients. Par exemple on peut exécuter la commande suivante pour obtenir les courbes de Kernel de l'évènement étudié selon le sexe des individus.

```
sts graph, by(Sex) tmax(85) hazard ci kernel(rectangle) width(1 1)
```

Il y a aussi l'option « *graalen* », elle permet de tracer la courbe de cumul des quotients pour chacun des risques (partie 1.2).

RAPPEL : en comparant les courbes il faut se rappeler que les niveaux ne sont pas comparables entre les graphiques, mais seulement à l'intérieur d'un même graphique, entre les courbes tracées pour les évènements concurrents.

Une fois ces étapes réalisées, on peut exporter les données obtenues sous Stata dans Excel afin de calculer les autres indicateurs de la fécondité comme l'ISF et l'AMM et d'obtenir des graphiques facilement. Pour cela des feuilles de calcul sous Excel ont été formatées (Annexes B).

III. Analyse de la fécondité, zone de Niakhar, 1983-2011

Le terme « fécondité » fait référence au nombre de naissances vivantes des femmes en âge de procréer.



Photos prises dans la zone de Niakhar lors du suivi démographique, juin 2012

De janvier 1983 à décembre 2011, d'après les données étudiées, la population de Niakhar comptait **78 046** individus dont **40 858** femmes et **37 088** hommes (100 individus pour qui cette information n'est pas indiquée). Parmi ces 78 046 individus, 106 530 observations sont enregistrées (une observation = un événement d'entrée ou de sortie), dont **38165** entrées par naissance.

Les données analysées dans la mesure de la fécondité (hormis le taux de natalité qui porte sur l'analyse des hommes et des femmes) sont les données des femmes entre 1983 et 2011. La population soumise au risque de connaître l'évènement « naissance » est constituée des femmes présentes dans la zone de Niakhar, soit un échantillon de **40181** femmes, qui a connu **36327** naissances. Dû à des anomalies⁹ et incohérences relevées dans les données, 571 observations ont été exclues de l'analyse (la date de début était supérieure à la date de l'évènement suivant $\text{datebeg} > \text{EventDate}$). Le biais reste cependant très faible, moins de 1%, et n'aura pas d'influence sur la mesure de la fécondité. (Annexe B.1 : matrice des incohérences).

La fécondité des femmes est étudiée entre leur 15^{ème} année et 49^{ème} année, bien qu'il arrive pour certaines femmes d'avoir des enfants à un âge plus jeune ou plus vieux (cf. page suivante).

Les femmes de 15-49 ans sont au nombre de **19 785**, pour lesquelles il a été enregistré **36228** naissances dans la zone. On remarque que 99 naissances sont de femmes de moins de 15 ans ou de plus de 49 ans.

- En 1983, on compte 3192 femmes résidentes dans la zone d'étude en âge de procréer (15-49 ans), ces dernières ont connu 735 naissances (soit un taux de fécondité de 230 enfants pour mille femmes)¹⁰.

⁹ Pour certaines dates le jour et le mois se sont inversés.

Cette anomalie peut entraîner des incohérences dans le déroulement des événements.

¹⁰ Les données de 1983 sont données à titre indicatif, elles sont sujettes à un biais d'erreur plus important (début du suivi démographique).

- En 1984, 4977 femmes en âge de procréer (15-49 ans) sont enregistrées comme résidentes dans la zone d'étude, ces dernières ont connu 1182 naissances (soit un taux de fécondité de 237 enfants pour mille femmes).
- En 2010, l'effectif des femmes de 15 à 49 ans et l'effectif des naissances atteignaient respectivement les 8793 femmes et les 1631 naissances (soit un taux de 185 naissances pour 1000 femmes).
- En 2011, le nombre de femmes de 15-49 ans s'élevait à 9034 pour 1493 naissances (soit un taux de fécondité de 165 enfants pour 1000 femmes).

On observe une hausse de la population féminine entre la période d'observation avec un effectif qui a presque doublé entre 1984 et 2011. Par contre l'on remarque une baisse du taux de fécondité avec un taux global de fécondité passant de 237 à 165 enfants pour 1000 femmes. Le taux global sur l'ensemble de la période s'élève à 202 naissances pour 1000 femmes.

En ce qui concerne les naissances d'enfants de femmes ne faisant pas partie de l'intervalle 15-49 ans, seul 0,04 % des femmes ont eu des enfants avant 15 ans et plus de 13 % après 49 ans (Annexe C.1 : table de survie). Sur les graphiques il est difficile de percevoir ces petits effectifs, on observe effectivement une grande majorité des naissances chez les femmes entre 15 et 49 ans (Figure 1 et 2).

Figure 1 : Intensité des naissances par âge de la mère

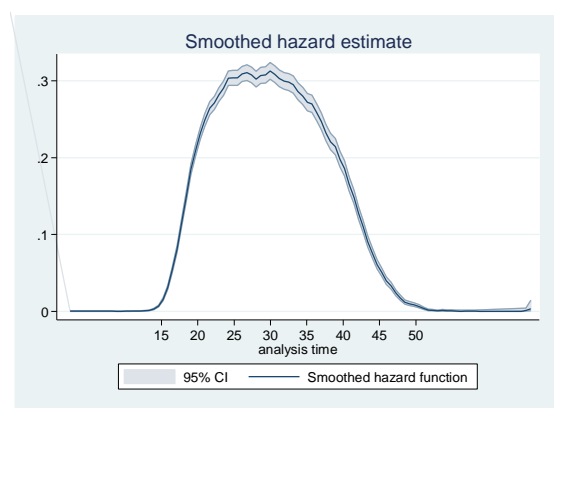
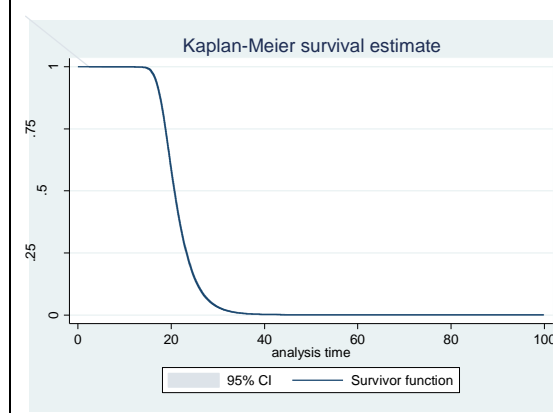


Figure 2 : Estimation de la courbe de survie des femmes de ne pas connaître l'évènement « enfant ».



Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

3.1 Les indicateurs de la fécondité

Dans le cadre de cette analyse de la fécondité par année et groupe d'âge quinquennaux, les indicateurs utilisés sont des **indicateurs transversaux**. Les indicateurs longitudinaux (descendance finale, ...) ne seront pas étudiés ici.

Les indicateurs utilisés sont les suivants :

➤ **Taux brut de natalité**

Il se calcule en rapportant le nombre de naissances enregistrées à l'effectif moyen de la population sur une même période, tous âges et sexes confondus.

Cet indicateur est peu satisfaisant d'un point de vue analytique puisqu'il prend au dénominateur de nombreuses personnes qui ne sont pas concernées par la natalité (hommes, enfants, personnes âgées).

➤ **Taux de fécondité**

Le taux global de fécondité est le rapport entre les naissances et l'effectif moyen des femmes en âge de procréer sur une même période.

Ces deux indicateurs, taux brut et taux global, devraient présenter les mêmes évolutions si la structure par âge et sexe de la population ne connaît pas de changement brutal au cours du temps.

Les taux de fécondité par âge, sont obtenus par le rapport entre le nombre de naissances de mère d'âge x à $x+n$ à la population de femmes d'âge x à $x+n$ soumises au risque, allant du temps t au temps $t+a$.

Ici nous utilisons des groupes d'âge quinquennaux, n est égal à 5 ; et des périodes d'observation par année puis par période regroupant plusieurs années (intervalle de 5 années entre 1984 et 2008, et pour la dernière période un intervalle de 3 années, 2009 2010 et 2011).

➤ **Indice Synthétique de Fécondité**

L'indice synthétique de fécondité (ISF) est obtenu par la somme des taux de fécondité. Il représente le nombre moyen d'enfants qui seraient nés vivants d'une femme (ou d'un groupe de femmes) pendant sa vie si elle vivait ses années de procréation en se conformant aux taux de fécondité par âge d'une année donnée. Cet indice exprime en un chiffre unique la fécondité des femmes à une époque donnée. Autrement dit, il représente le nombre d'enfants qu'aurait une femme, pourvu que les taux de fécondité d'une année donnée puissent s'appliquer à elle tout au long de sa vie reproductive (voir ci-dessous le mode de calcul de l'ISF).

L'ISF est un des indicateurs principaux calculé dans les analyses transversales de la fécondité.

➤ **Age Moyen à la Maternité**

L'âge moyen à la maternité (AMM) correspond à la moyenne pondérée des âges de reproduction par les taux de fécondité par âge correspondant.

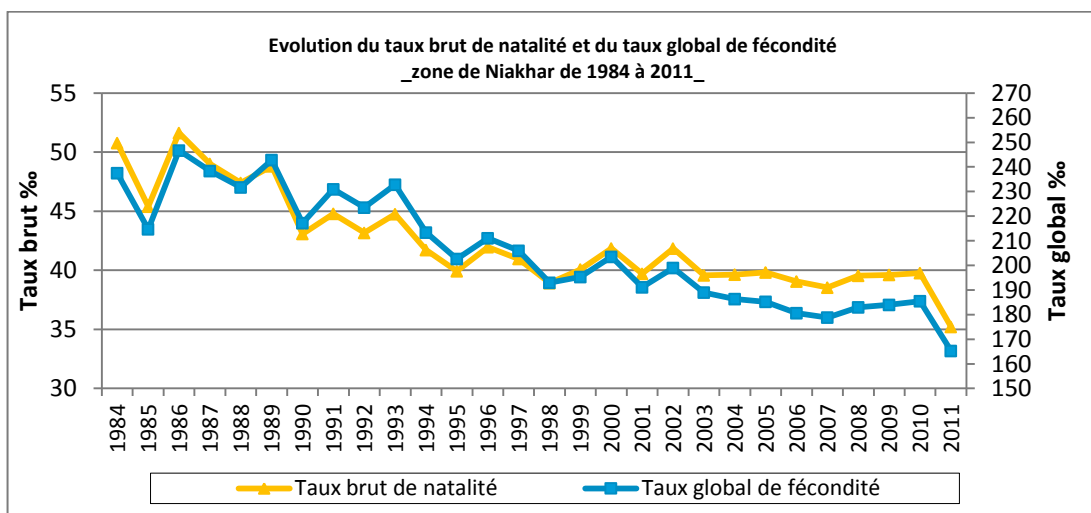
Il existe d'autres indicateurs d'âge moyen comme par exemple celui de l'âge moyen des mères à la naissance de leur enfant mais nous préférons l'AMM car cet indicateur est moins sensible aux changements de structure par âge.

3.2 Evolution générale de la fécondité

La fécondité générale dans la zone de Niakhar entre 1984 et 2011 est élevée, l'on observe un indicateur synthétique moyen de l'ordre de 7 enfants par femme sur l'ensemble de la période. Cependant la tendance est à la baisse, avec un indice passant de 8 enfants par femme en 1984, à 6 enfants par femme en 2011 (Figure 4 et Tableau 5). On peut retrouver cette baisse de l'ISF au travers du taux brut de natalité et du taux global de fécondité qui présentent tous deux une tendance à la baisse, à une échelle différente. On observe un déclin marqué dans les années 1990 (Figure 3).

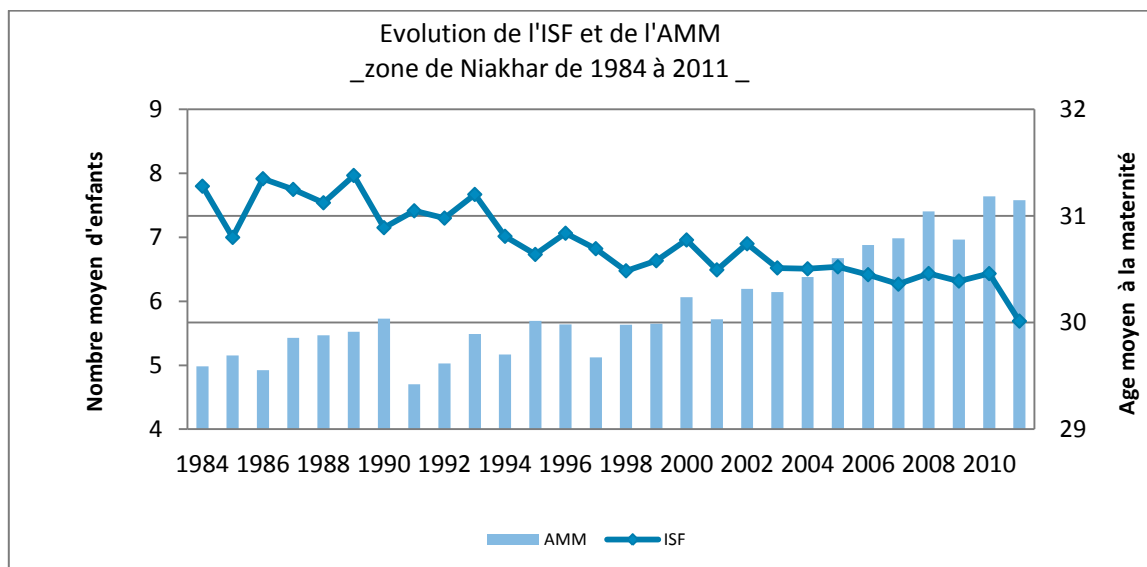
Le niveau des taux correspond à une population en transition de la fécondité (Delaunay, 2006). Effectivement la fécondité dans la zone de Niakhar, et dans le Sénégal en général, est dans une phase de transition depuis les années 1990 avec des comportements de fécondité à la baisse. Cependant cette transition évolue lentement dans la zone de Niakhar avec une fécondité qui demeure toujours élevée. La fécondité reste le moteur de la croissance démographique. (Delaunay, Adjamagbo, Lalou, 2006). Si on s'intéresse au sexe de l'enfant, on observe qu'il y a légèrement plus de naissances masculines que féminines. Il y a 51% de naissance de garçon, et 49 % de naissance de fille, soit à la naissance un rapport de masculinité de 104 garçons pour 100 filles. La mortalité des garçons étant globalement plus élevée que celle des filles, le sex ratio diminue avec l'âge et les femmes deviennent majoritaires comme c'est le cas dans la plupart des pays.

Figure 3 : Evolution du taux brut de natalité et du taux global de fécondité, zone de Niakhar, 1984-2011



Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

Figure 4 – Evolution de l'indicateur synthétique de fécondité et de l'âge moyen à la maternité, zone de Niakhar, 1984-2011



Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

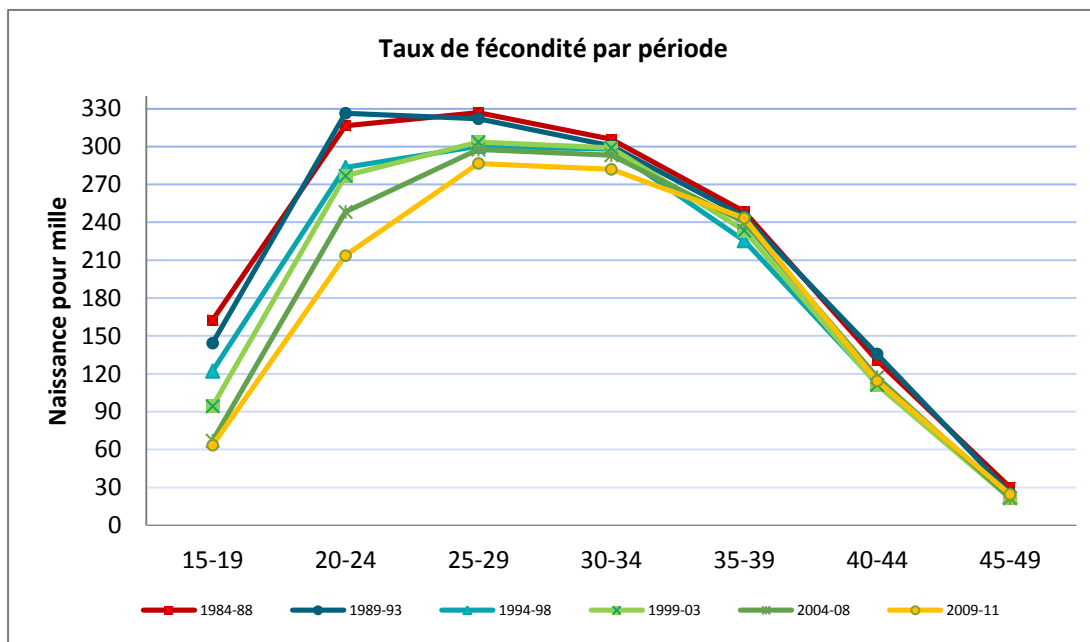
On constate la présence d'un effet de calendrier sur les naissances dans la vie des femmes.

En effet, au cours des trente dernières années, on assiste à une baisse des taux de fécondité des femmes dans la vingtaine (Figure 5). Le taux de fécondité chez les mères âgées de 15 à 19 ans est passé de 194 enfants pour 1000 femmes à un taux de 58 enfants pour 1000 femmes entre 1984 et 2011 (Tableau 5). L'âge moyen à la maternité présente une tendance à la hausse (Figure 4), dépassant les 31 ans en 2010-2011¹¹. Cette augmentation de l'âge moyen à la maternité témoigne de la baisse de la fécondité aux jeunes âges et du maintien d'une forte fécondité aux âges élevés.

Les Sénégalaises âgées de 25 à 29 ans demeurent les plus fécondes au début des années 1980, alors que dans les années 2000 elles sont suivies de plus en plus près par les femmes âgées de 30 à 34 ans, voir même dépassées par ces dernières. Dans le groupe des femmes de 25 à 29 ans, 283 sur mille ont donné naissance à un enfant en 2010, contre 287 sur mille pour les femmes âgées de 30 à 34 ans.

L'explication la plus plausible de cette baisse de la fécondité des femmes des classes d'âges les plus jeunes est celle du changement du calendrier de l'âge au mariage. Des études menées auprès de cette population montrent que ces changements sont liés au fait que les femmes se marient plus tard et que la fécondité avant le mariage est socialement réprouvée. (Delaunay, 1994 et Delaunay, Adjmagbo, Lalou, 2006). Cette baisse se traduit dans les graphiques par un décalage des courbes des périodes les plus récentes par rapport aux courbes des périodes les plus anciennes, principalement aux jeunes âges des femmes (Figure 5). L'équivalent chiffré de ces courbes est obtenu dans l'annexe C.4.

Figure 5 : Evolution des taux de fécondité par groupe d'âge selon la période de naissance ,1984-2011, zone de Niakhar

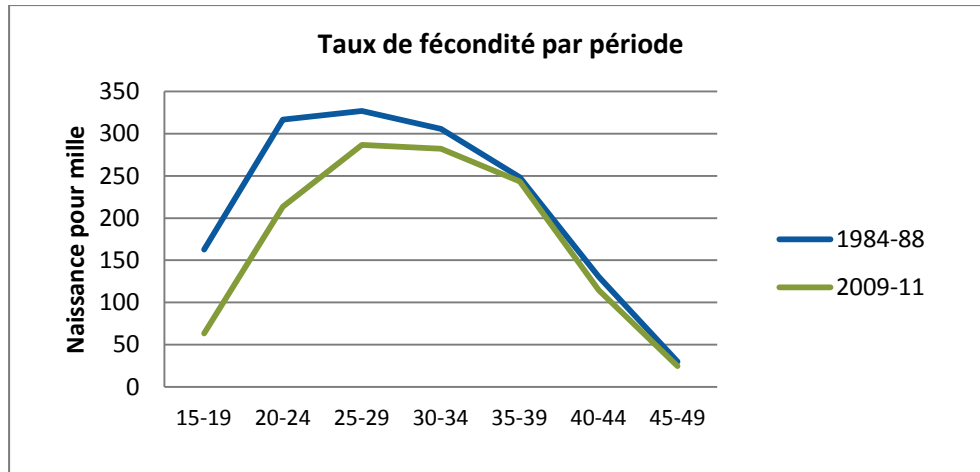


Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

¹¹ Selon les sorties Stata, la moyenne de l'âge des femmes au moment des naissances est de 28,4 ans alors qu'en faisant la méthode du calcul de l'AMM, on obtient un âge moyen à la maternité de 30 ans. (la méthode de calcul est différente ce qui explique cet écart de 2 ans). Comme il l'a été mentionné dans la partie 1.2 le choix de l'indicateur de l'âge moyen est celui de l'AMM.

Afin de mieux visualiser cette différence entre les taux, nous pouvons regarder le graphique suivant qui représente les taux de fécondité des femmes par groupe d'âge des deux périodes extrêmes de l'analyse, la période 1 1984-1988 et la période 6 2009-2011 (Figure 6).

Figure 6 : Evolution des taux de fécondité par groupe d'âge selon la période de naissance, période 1 1984-1988 et période 6 2009-2011



Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

Nous pouvons aussi visualiser l'allure de cette évolution directement sous Stata (Annexe C.6) par la commande suivante :

```
sts graph, by(perio) tmax(50) tmin(15) hazard kernel(rectangle)
```

Les courbes de Kaplan-Meier (Annexe C.7) représentent la probabilité de survie des femmes de ne pas connaître l'évènement naissance à chaque âge selon les six périodes (périodes des naissances). En comparant les périodes de naissances, on voit apparaître un écart entre les courbes. Au fil du temps, la fécondité évolue, on remarque bel est bien que les femmes ont de moins en moins d'enfants.

Tableau 5 – Taux de fécondité par groupe d'âges, indice synthétique de fécondité (ISF) et âge moyen à la maternité, selon les années d'observation

Année	ISF	Taux de fécondité %							Age Moyen à la maternité
		15-19	20-24	25-29	30-34	35-39	40-44	45-49	
1984	7,80	193,80	306,54	323,04	301,71	281,86	130,63	22,07	29,59
1985	7,00	154,90	279,89	314,96	274,98	221,55	127,59	25,62	29,69
1986	7,92	167,25	333,46	349,00	337,47	246,52	110,08	39,27	29,55
1987	7,75	167,04	332,17	313,14	294,53	257,91	152,71	32,11	29,86
1988	7,54	128,78	331,23	332,98	316,92	235,25	131,59	31,09	29,88
1989	7,97	167,67	319,35	358,02	297,98	251,02	172,60	26,61	29,92
1990	7,15	134,20	313,05	285,85	277,47	255,96	136,17	27,33	30,04
1991	7,41	154,79	327,29	308,19	325,95	242,51	100,18	23,80	29,42
1992	7,30	134,96	340,73	315,39	291,24	222,72	127,29	27,55	29,62
1993	7,67	131,67	332,89	343,08	309,02	252,65	142,83	21,82	29,89
1994	7,01	136,35	297,39	292,72	326,20	225,81	88,00	36,52	29,70
1995	6,73	131,08	262,28	288,96	304,68	209,05	122,40	27,93	30,02
1996	7,06	113,55	289,39	335,81	290,14	245,85	114,92	22,94	29,98
1997	6,82	126,65	296,13	295,72	294,40	223,30	111,21	16,89	29,67
1998	6,48	105,84	273,09	288,77	271,61	222,52	122,90	10,45	29,98
1999	6,63	105,01	258,16	314,12	302,85	222,33	116,18	8,04	29,99
2000	6,96	84,24	287,43	332,52	302,83	255,36	111,82	17,33	30,24
2001	6,49	100,51	277,87	283,78	279,41	233,16	103,05	20,27	30,03
2002	6,90	95,83	280,62	308,25	324,23	213,71	130,36	26,62	30,32
2003	6,52	86,64	278,79	284,65	285,86	242,78	94,37	31,08	30,29
2004	6,51	72,10	262,48	313,70	290,35	227,77	117,21	17,79	30,43
2005	6,54	73,98	253,09	290,64	310,22	258,57	98,80	22,30	30,60
2006	6,42	65,09	252,26	297,86	297,87	222,65	121,79	25,52	30,73
2007	6,27	66,24	251,27	269,78	280,86	243,85	125,18	16,21	30,79
2008	6,43	59,26	225,81	316,35	288,91	246,83	124,06	25,38	31,04
2009	6,32	69,18	227,12	303,14	297,38	229,54	110,14	26,64	30,78
2010	6,43	63,33	232,46	282,74	287,17	266,91	129,17	24,67	31,18
2011	5,69	58,00	183,17	274,72	262,35	233,51	103,93	22,11	31,15

Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

Les calculs de l'ISF et de l'AMM sont effectués sous Excel, ces fiches-outils de calcul Excel sont présentées en Annexe B. 2.

Dans les parties suivantes il est étudié le lien entre fécondité et religion, fécondité et niveau d'instruction de la femme, fécondité et niveau de vie, fécondité et niveau de richesse agro-pastorale et fécondité et village. Des tests de significativité (tests du log-rank) seront réalisés afin de savoir s'il y a un effet significatif ou non de la variable explicative étudiée sur la variable à expliquer, ici la fécondité.

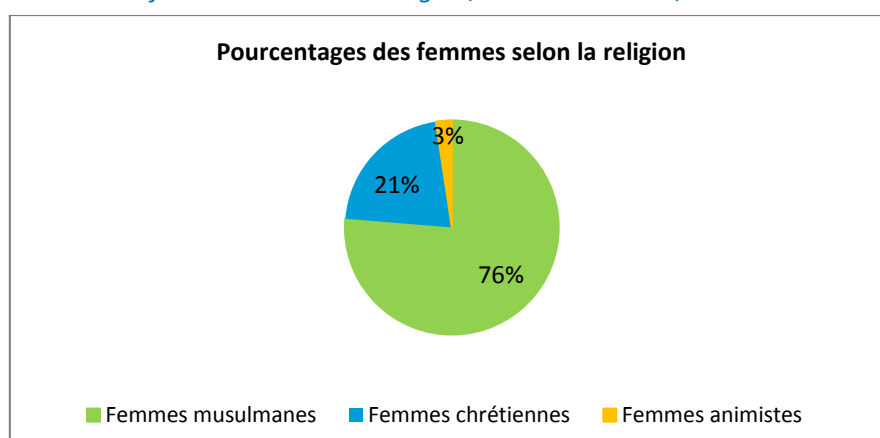
3.3 Fécondité et religion des femmes

Il est question dans cette sous partie d'étudier la fécondité des femmes selon leur religion afin de savoir si la religion peut avoir un effet significatif sur le niveau de fécondité.

En ce qui concerne la religion des femmes, pour 67 % d'entre elles ont connaît l'information, elles sont soit musulmanes, soit chrétiennes, soit animistes. Cependant pour les 33 % restant l'information n'est pas transmise (elles sont 13 % à ne pas savoir leur religion et 20 % dont l'information est manquante).

La plupart des femmes de la région de Fatick sont musulmanes, elles le sont à 76 % soit plus des trois quart des femmes, alors qu'elles sont 26 % à être chrétienne et seulement 3 % à être animiste (Figure 7).

Figure 7 : Répartition des femmes selon leur religion, zone de Niakhar, 1983-2011



Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

Les indicateurs de la fécondité des femmes sont tous relativement proches qu'elles soient musulmanes, chrétiennes ou animistes (Tableau 6), cependant on remarque que d'après l'analyse descriptive l'ISF des femmes animistes est légèrement plus élevé que celui des autres femmes.

Tableau 6 : Analyse descriptive de la fécondité par religion des femmes de 1983 à 2011

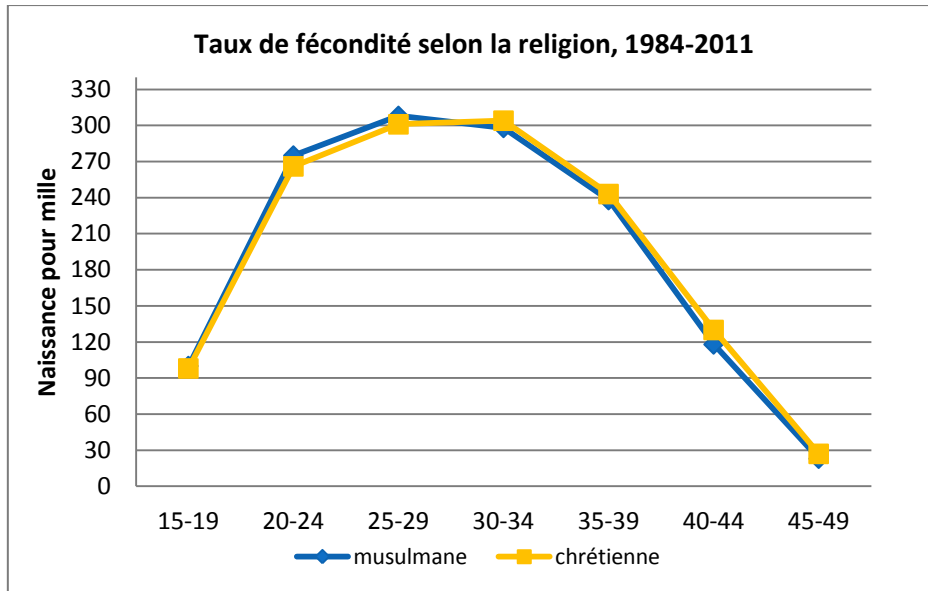
	Femmes musulmanes	Femmes chrétiennes	Femmes animistes
Effectifs des femmes	20544	5735	643
Effectifs des naissances	26273	7195	740
Age moyen à la première naissance (dans la zone)	21,5 ans	21,6 ans	21,2 ans
Age moyen des naissances (Stata)	28,6 ans	28,7 ans	29,1 ans
AMM	30,14 ans	30,40 ans	29,69 ans
ISF sur la période 84-2011	6,94	6,98	7,22

Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

Le niveau de fécondité des femmes animistes varie d'une année à l'autre du fait qu'il y a peu de femmes animistes dans la zone d'étude. L'ISF pour ces femmes varie fortement avec un indice minimum de 5 enfants par femme et un indice maximum de presque 11 enfants par femme. Elles ne seront donc pas prises en compte dans l'analyse qui suit.

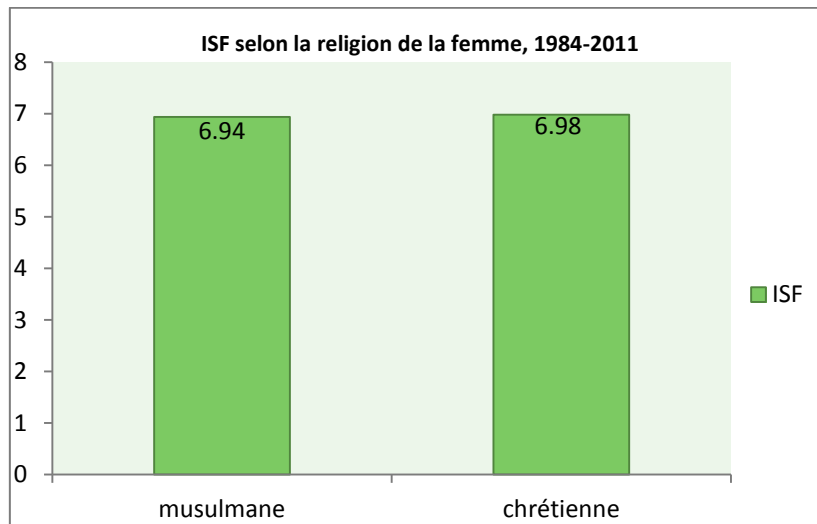
On remarque peu de différence entre les courbes représentant les taux de fécondité des femmes chrétiennes et les taux de fécondité des femmes musulmanes (Figure 8). En ce qui concerne l'ISF on peut en tirer les mêmes conclusions, l'ISF des femmes musulmanes comme des femmes chrétiennes, est en moyenne de 7 enfants par femme (Figure 9).

Figure 8 : Taux de fécondité des femmes selon leur religion, zone de Niakhar, 1983-2011



Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

Figure 9 : ISF des femmes musulmanes et ISF des femmes chrétiennes, Niakhar, 1984-2011



Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

D'après les courbes d'estimation de Kaplan-Meier obtenues sous Stata pour ces deux catégories, les femmes musulmanes et les femmes chrétiennes, on remarque bel et bien que les deux courbes ont la même tendance, on ne peut d'ailleurs pas les distinguer tellement elles sont superposées entre elles (Annexe C.8, courbe de survie). Faire un test du log-rank n'est donc pas judicieux, à titre indicatif celui-ci montre un effet non significatif de la variable religion avec les modalités « musulmanes » et « chrétiennes » sur la fécondité des femmes (Annexe C.8).

Pour conclure, la religion de la femme n'a pas d'effet significatif sur la fécondité. Celles-ci ont leurs enfants aux mêmes âges, avec des taux de fécondité semblables et un ISF moyen de 7 enfants. La variable religion n'est donc pas une variable que l'on pourra prendre en compte dans un modèle de Cox.

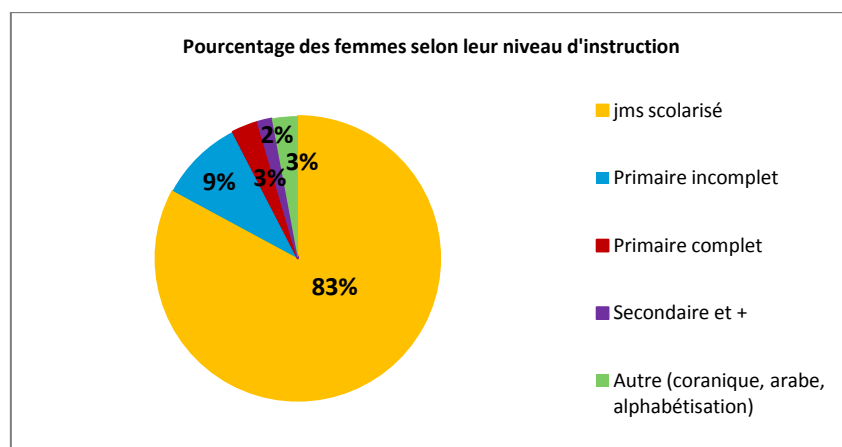
Continuons l'analyse de la fécondité, cette fois-ci en étudiant le lien entre fécondité et niveau d'instruction.

3.4 Fécondité et instruction des femmes

L'étude de la fécondité des femmes selon leur niveau d'instruction va permettre de savoir si le niveau d'instruction peut avoir un effet significatif sur le niveau de fécondité. Avant de mettre en lien ces deux variables, présentons la situation de l'instruction des femmes dans la zone de Niakhar.

Les femmes les plus instruites, c'est-à-dire avec un niveau secondaire et plus, sont les moins nombreuses avec un pourcentage de seulement 2 % (Figure 10). A l'inverse les femmes jamais scolarisées sont les plus nombreuses, elles représentent 83 % des femmes de la zone de Niakhar entre 1983 et 2011. Ces statistiques tiennent compte seulement des données où l'information est enregistrées. Notons que pour 0,2 % des femmes on ne connaît pas leur niveau d'instruction et que pour 50 % des femmes l'information est manquante.

Figure 10 : Répartition des femmes selon leur niveau d'instruction, zone de Niakhar, 1983-2011



Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

Les femmes jamais scolarisées sont celles ayant le plus d'enfants, avec un taux de fécondité global de 208 enfants pour 1000 femmes. A l'inverse les femmes les plus instruites, avec un niveau secondaire et plus, sont celles ayant le moins d'enfant avec un taux global de 118 enfants pour 1000 femmes. Cette différence du niveau de fécondité est flagrante lorsque l'on regarde l'ISF, les femmes les moins instruites ont en moyenne 7 enfants alors que les plus instruites seulement 4 enfants, soit 3 enfants de moins (Tableau 7 et Figure 11). De plus, les femmes les plus instruites ont leur première naissance dans la zone à un âge plus avancé que celui des autres femmes, avec une moyenne d'âge à la première naissance de 27 ans, contrairement aux autres femmes qui ont leurs premiers enfants entre 21 et 23 ans.

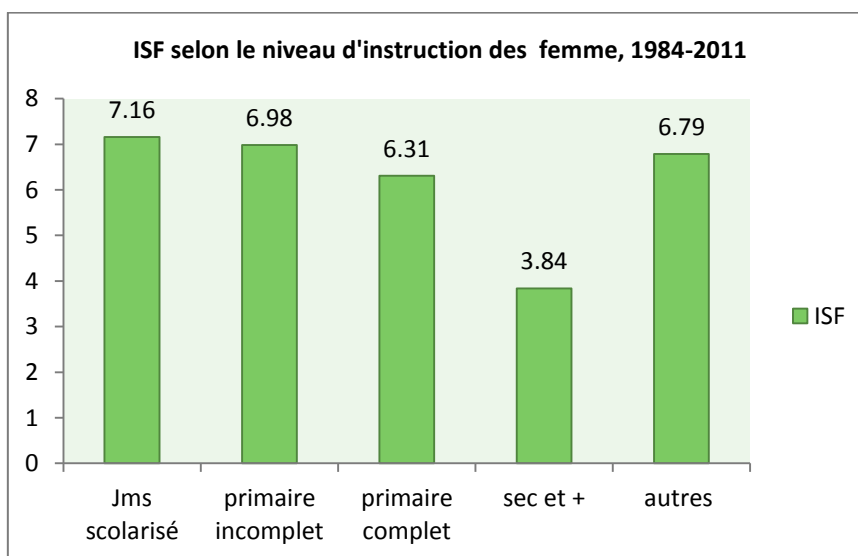
Les femmes ayant un niveau d'instruction classé dans la catégorie Autre (coranique, arabe, alphabétisation), ont un taux de fécondité compris entre celui des femmes n'ayant jamais été scolarisées et celui des femmes avec un niveau primaire complet.

Tableau 7 : Analyse de la fécondité par niveau d'instruction des femmes, Niakhar, 1983-2011

	Femmes jamais scolarisées	Femmes niveau primaire incomplet	Femmes niveau primaire complet	Femmes niveau secondaire et +	Femmes Autre (coranique, etc.)
Effectifs des femmes	17131	1974	619	329	605
Effectifs des naissances	28939	2498	1204	357	1155
Age moyen à la première naissance (dans la zone)	21,18 ans	22,89 ans	22,32 ans	25,69 ans	21,55 ans
Age moyen des naissances (Stata)	28,89 ans	26,65 ans	27,81 ans	27,53 ans	27,38 ans
AMM	29,97 ans	30,79	30,53 ans	29,53 ans	29,97 ans
ISF sur la période 84-2011	7,16	6,98	6,31	3,84	6,79

Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

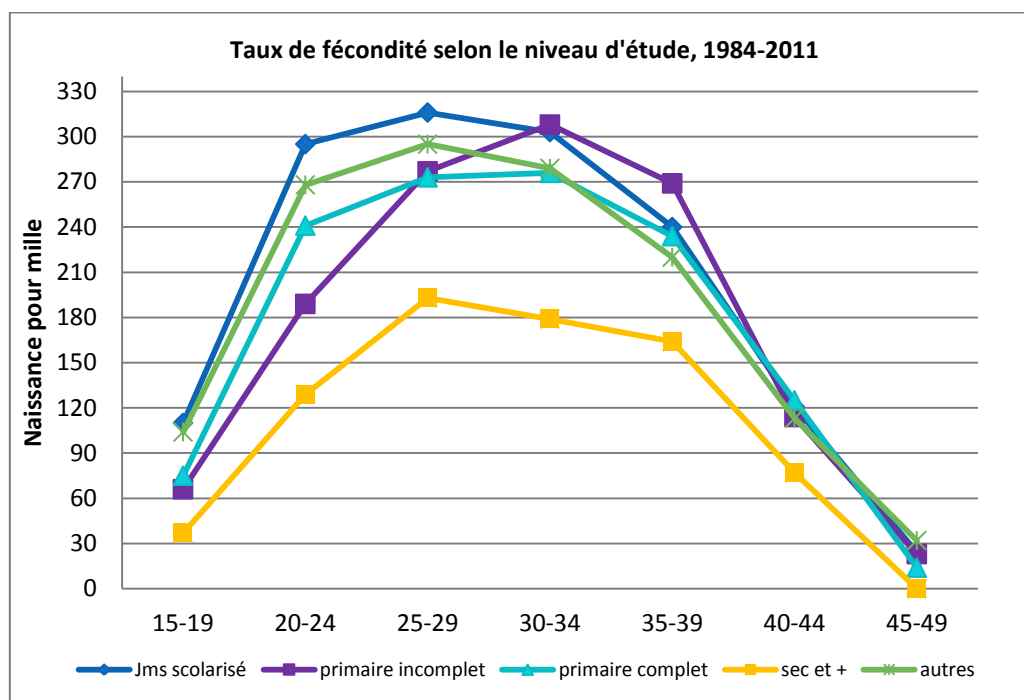
Figure 11 : Indice Synthétique de Fécondité en fonction du niveau d'instruction des femmes, Niakhar, 1999-2003



Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

D'après les résultats obtenus, le niveau d'instruction a un effet sur le niveau de fécondité. Effectivement plus une femme est instruite au moins elle a d'enfants et inversement les femmes les moins instruites ont plus d'enfants. On peut observer cette tendance sur le graphique suivant, la courbe du niveau de fécondité des femmes les plus instruites est la moins élevée alors que celle des femmes n'ayant jamais été scolarisé est la plus élevée (Figure 12). Cette tendance est repérable aussi dans le graphique représentant les courbes de survie estimées par la méthode de Kaplan-Meier (Annexe C.9). De plus on remarque un décalage de calendrier, les femmes les plus instruites ont tendance à avoir leur premier enfant à un âge plus avancé, en moyenne vers 25-26 ans.

Figure 12 : Taux de fécondité des femmes selon son niveau d'instruction, zone de Niakhar, 1984-2011.



Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

Vérifions la significativité de cette variable du niveau d'instruction sur la fécondité des femmes par l'application du test du log-rank. Le test est significatif, on peut conclure que le niveau d'instruction a un effet significatif sur le niveau de fécondité de la femme au cours de sa vie féconde au seuil de 1% de se tromper (Annexe C.9). Plus une femme est instruite, moins elle aura de chance d'avoir beaucoup d'enfant et inversement au moins une femme est instruite au plus elle aura de chance d'avoir beaucoup d'enfant.

3.5 Fécondité, niveau de vie et richesse agro-pastorale

Dans le cadre de cette étude, nous reprenons les indicateurs créés par Théodore HOUNGBEGNON et Atikpo Mawunya ADESU dans leur mémoire de fin de cycle « Construction d'indicateurs de niveau de vie des ménages et relation avec la migration de travail dans une zone rurale du Sénégal : Niakhar »¹². Grâce à leur travail un indicateur sur le niveau de vie des ménages des résidents de Niakhar, nommé ISNV (Indice synthétique du niveau de vie) et un Indicateur Synthétique de Richesse Agro-Pastorale (ISRAP) sont disponibles.

L'ISNV et l'ISNV2 (indice normalisé) permettent de mesurer le niveau de vie des ménages ;

L'ISRAP et l'ISRAP2 (indice normalisé) permettent, quant à eux, de mesurer le niveau de richesse agro-pastorale.

¹² Pour plus d'information se référer à leur travail « Construction d'indicateurs de niveau de vie des ménages et relation avec la migration de travail dans une zone du Sénégal : Niakhar », Théodore HOUNGBEGNON et Atikpo Mawunya ADESU, 2012, Mémoire de fin de cycle-ITS, IRD UMR 198.

Le niveau de vie et le niveau de RAP des ménages sont observés en 2003, nous étudierons son effet sur la fécondité durant la période qui comprend l'année 2003 en supposant que le niveau de ces deux indicateurs des années entourant 2003 soit du même ordre.

Comme précédemment, un test du log-rank est effectué afin de savoir si le niveau de vie et le niveau de richesse agro-pastorale des ménages ont un effet significatif sur le niveau de fécondité.

Pour pouvoir faire cette analyse du niveau de fécondité des femmes et des deux indicateurs synthétiques des ménages, il a fallu avant tout attribuer à tous les membres d'un même ménage son niveau de vie et de RAP correspondant. Le programme permettant la création de la variable `vi_co_cu` (la variable identifiant du ménage) sous Stata est présenté en annexe A.7.

Fécondité et niveau de vie, zone de Niakhar entre 1999 et 2003

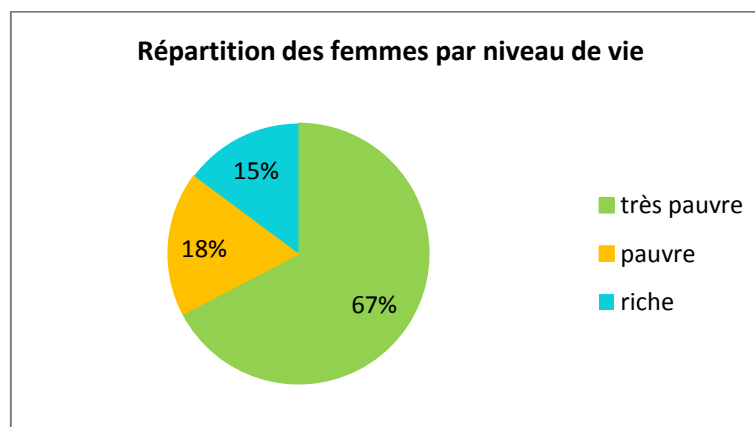
Il s'agit de savoir ici quel effet le niveau de vie des ménages dans lequel la femme vit a-t-il sur le niveau de fécondité de la femme.

Nous testerons la significativité du niveau de vie selon une classification en 3 classes de niveau (cette analyse est aussi réalisée sur la classe pauvre et la classe non pauvre et aussi en fonction de 5 classes de niveau de vie (Annexe C.10)). Pour répartir les individus dans ces différentes classes nous avons repris la classification ascendante hiérarchique réalisée par T. H. et A. M. A. à partir de l'ISNV créé par leur soin. Le seuil de pauvreté estimé à partir de leur classification est de 0,2809. Sur la base de ce seuil, la création des classes a pu être réalisée sous Stata.

L'étude de l'effet du niveau de vie des ménages sur la fécondité entre 1999 et 2003 porte sur un échantillon de 18136 femmes de tout âge. Parmi ces femmes 5432 enfants sont nés. L'effectif des personnes_années est de 27586 pour les femmes de 15 à 49 ans, le nombre de naissance de 5421 (soit 11 naissances de moins que les naissances totales, c'est-à-dire que ces 11 naissances sont de femmes hors la tranche d'âge 15-49 ans). Le taux global de fécondité (femme 15-49 ans) est de 197 naissances pour 1000 femmes.

Sur les 18136 femmes étudiées, plus de la majorité des femmes sont considérées comme vivant dans des ménages « très pauvre » (Figure 13). Effectivement 67% (12214) font parties de la classe très pauvre, 18% (3235) de la classe pauvre et 14% (2687) de la classe « riche ».

Figure 13 : Répartition des femmes selon le niveau de vie, zone de Niakhar, 1999-2003.



Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

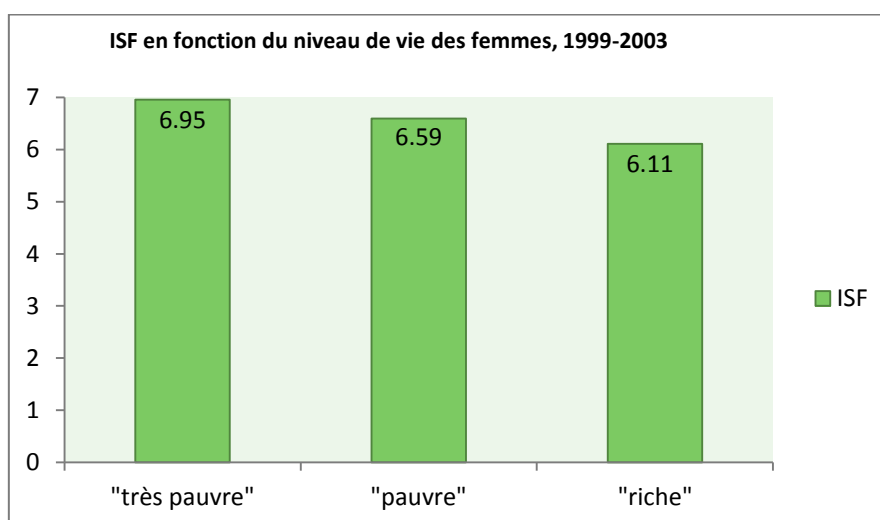
La tendance est la même que lorsqu'on a une répartition en 2 classes (Annexe C.10), plus la femme est pauvre, plus elle a un taux de fécondité élevé et plus elle est « riche » moins le taux de fécondité est élevé (Tableau 9). Le taux de fécondité pour les femmes considérées comme « très pauvre » est de 201 enfants pour 1000 femmes alors qu'il est de 179 pour les femmes considérées comme « riches ». Cet écart du taux de fécondité se traduit sur l'ISF, les femmes les plus riches ont en moyenne un enfant de moins que les femmes les plus pauvres (Figure 14).

Tableau 9 : Taux global de fécondité en fonction du niveau de vie des femmes, Niakhar, 1999-2003

Niveau de vie des femmes	Personne-année (période)	Naissance	Taux global de fécondité ‰
Groupe 1 « très pauvre »	18327	3684	201
Groupe 2 « pauvre »	5071	987	195
Groupe 3 « riche »	4188	750	179
Total	27586	5421	197

Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

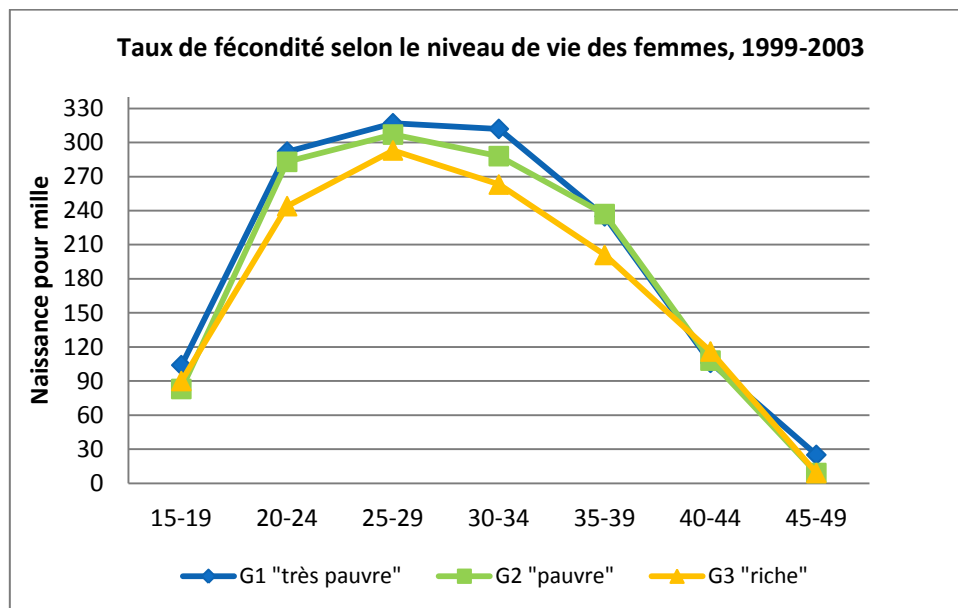
Figure 14 : Indice Synthétique de Fécondité en fonction du niveau de vie des femmes, Niakhar, 1999-2003



Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

Les taux de fécondité sont moins élevés pour les femmes les plus riches (Figure 15). On remarque de légères différences aussi entre le niveau de fécondité des femmes « très pauvres » et des femmes « pauvres ». Les femmes « pauvres » ont dans l'ensemble moins d'enfants que les femmes considérées comme « très pauvre ». Notons que dans la classification en 5 classes (Annexe C.10), on remarque que les femmes les plus riches ont leur premier enfant plus tard que les autres femmes. Il y a donc, en plus d'avoir un effet sur le nombre d'enfant, un effet sur le calendrier des naissances de la femme.

Figure 15 : Taux de fécondité des femmes selon leur niveau de vie, zone de Niakhar, 1999-2003.



Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

L'effet de la variable niveau de vie est significatif au seuil de 1 % sur le niveau de fécondité des femmes. Dans l'ensemble les femmes appartenant à la classe des ménages les plus « riches » ont moins de « risque » d'avoir des enfants que les femmes appartenant à la classe des ménages pauvres (Annexe C.10). Cette conclusion reste la même, que ce soit une répartition en 2, 3 ou 5 classes de niveau de vie.

En conclusion, l'effet du niveau de vie sur la fécondité des femmes, que ça soit sur l'année 2003 ou sur la période 1999-2003 et quel que soit la classification, est significatif. Le test fait pour les données de 2003 (Annexe C.10) a cependant un seuil d'erreur plus important, 10% et non plus 1% comme c'est le cas pour l'analyse faite entre 1999-2003.

Le niveau de vie joue un rôle sur le niveau de fécondité des femmes, les femmes avec un meilleur niveau de vie ont dans l'ensemble moins d'enfants. Mais est-ce le niveau de vie qui joue un rôle sur le niveau de fécondité des femmes, ou inversement ? La réponse n'est pas fournie ici mais il serait intéressant de l'approfondir.

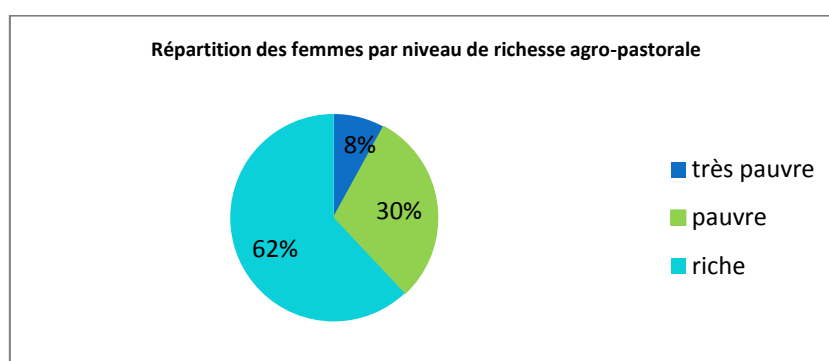
A présent passons à l'analyse de l'effet du niveau de richesse agro-pastorale sur la fécondité des femmes. Rappelons que la zone de Niakhar est une zone rurale où le travail de la terre est la principale activité des ménages.

Fécondité et niveau de richesse agro-pastorale, zone de Niakhar entre 1999 et 2003

L'étude de l'effet du niveau de richesse agro-pastorale des ménages sur la fécondité se fait seulement pour la période 1998-2003, la population étudiée reste donc la même que pour celle de l'étude sur le niveau de vie (personnes_années = 27586 ; naissance = 5421 et taux de fécondité = 197 ‰).

A l'inverse du niveau de vie, la plupart des femmes ont un niveau de RAP élevé. Il concerne 62% d'entre elles alors que seulement 8% des femmes sont considérées comme très pauvres en ce qui concerne la RAP (Figure 16).

Figure 16 : Répartition des femmes selon le niveau de richesse agro-pastorale, zone de Niakhar, 1999-2003.



Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

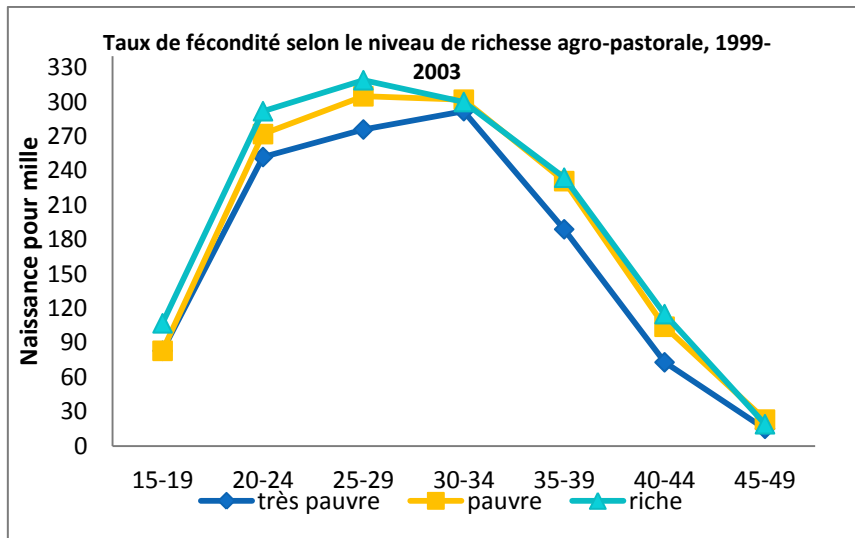
Le taux de fécondité global des femmes considérées comme très pauvres est de 176 enfants pour 1000 femmes alors que pour les femmes riches il est de 206 enfants pour 1000 femmes, soit 30 points de différences (Tableau 11). Quelques soit l'âge de la femme cette tendance reste la même, les taux sont moins élevés pour les femmes avec un niveau de richesse agro-pastorale lui aussi moins élevé (Figure 17). Plus une femme a un niveau de RAP élevé, plus son niveau de fécondité est élevé. A l'inverse plus une femme a un niveau bas de RAP, moins elle a d'enfants. Effectivement les femmes les plus riches en matière de richesse agro-pastorale ont en moyenne 7 enfants, soit un enfant de plus que les femmes les plus pauvres (Figure 18). Ces résultats laissent penser que les familles dotées d'une plus grande richesse agro-pastorale ont plus d'enfants car il y a un besoin de main d'œuvre plus important.

Tableau 11 : Taux global de fécondité en fonction du niveau de richesse agro-pastorale des femmes.

Niveau pastorale des femmes	Personne-année (période)	Naissance	Taux global de fécondité ‰
1 « très pauvre »	2215	390	176
2 « pauvre »	8393	1542	184
3 « riche »	16978	3489	206
Total	27586	5421	197

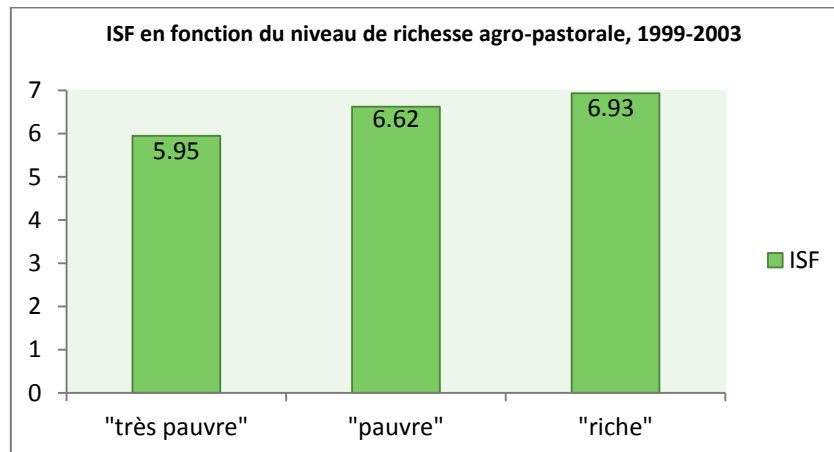
Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

Figure 17 : Taux de fécondité selon le niveau de richesse agro-pastorale des femmes, zone de Niakhar, 1999-2003



Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

Figure 18 : Indice Synthétique de Fécondité en fonction du niveau de richesse agro-pastorale des femmes, Niakhar, 1999-2003



Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

D'après les résultats obtenus par le test du log-rank il y a un effet significatif du niveau de RAP sur le niveau de fécondité des femmes au cours de leur période de fécondité, 15-49 ans, au seuil d'erreur de 1% (Annexe C.11).

3.6 Modélisation : les facteurs explicatifs de la fécondité

Le modèle de Cox permet ici d'étudier le niveau de fécondité des femmes en fonction de la durée de la vie féconde des femmes et de plusieurs variables explicatives de façon simultanée, ici les variables explicatives retenues sont le niveau d'instruction de la femme, le niveau de vie et le niveau agro-pastorale du ménage dans lequel la femme vit. Les résultats obtenus sont disponibles en annexe C.12.

Le risque relatif d'avoir des enfants faiblit avec l'instruction, il est statistiquement significatif quel que soit le niveau d'instruction avec un seuil d'erreur variant entre 1 % et 10 % (Tableau 12). On remarque que les femmes les plus instruites (edu3) ont un risque plus faible d'avoir des enfants (0,59). C'est d'ailleurs pour cette caractéristique que le risque est le plus faible comparé à toutes les autres caractéristiques étudiées ici. Le risque relatif d'avoir des enfants faiblit lui aussi avec le niveau de vie, mais moins qu'avec le niveau d'instruction, il est significatif statistiquement qu'avec le niveau de vie le plus élevé (nv2). Seul le niveau de richesse agro-pastorale n'affaiblit pas le niveau de fécondité, effectivement plus le niveau est haut plus le risque relatif d'avoir des enfants est élevé.

On remarque que le niveau d'instruction influe plus sur le niveau de fécondité que les autres variables. Les écarts entre les risques relatifs des différentes modalités par rapport à celle de référence sont plus élevés contrairement aux écarts des risques des autres variables.

Tableau 12 : Coefficients du modèle semi-paramétrique à risque proportionnel de Cox pour l'arrivée de l'évènement naissance dans la vie féconde des femmes (sans interaction).

Variables #	Catégorie	Période 1999-2003
Niveau d'instruction (edu0)	edu1	0,81***
	edu2	0,84**
	edu3	0,59***
	eduA	0,87*
Niveau de vie (nv0)	nv1	0,95
	nv2	0,91**
Niveau de richesse agro-pastorale (rap0)	rap1	1,08
	rap2	1,15***

Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

La modalité de référence figure entre parenthèses.

***Significatif au seuil de 1 % ;

**Significatif au seuil de 5 % ;

*Significatif au seuil de 10 %.

Dans le modèle de Cox suivant, l'interaction entre niveau d'instruction et niveau de vie est prise en compte afin d'étudier l'effet de ces deux variables réunies et ainsi éviter les quelconques possibilités d'effets entre les deux variables elles-mêmes. On remarque que les femmes ayant à la fois un niveau élevé d'instruction et de vie sont celles ayant un risque relatif d'avoir une plus faible fécondité (Tableau 13). Cependant ce risque n'est pas significatif statistiquement.

Tableau 13 : Coefficients du modèle semi-paramétrique à risque proportionnel de Cox pour l'arrivée de l'évènement naissance dans la vie féconde des femmes (avec interaction).

Variabes #	Catégorie	Période 1999-2003
Niveau d'instruction (edu0)	edu1	0,86**
	edu2	0,89
	edu3	0,68**
	eduA	0,86**
Niveau de vie (nv0)	nv1	0,96
	nv2	0,98
Niveau d'instruction et niveau de vie (edu0 et nv0)	edu1nv1	0,89
	edu1nv2	0,73**
	edu2nv1	1,06
	edu2nv2	0,72*
	edu3nv1	0,84
	edu3nv2	0,67

Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

La modalité de référence figure entre parenthèses.

***Significatif au seuil de 1 % ;

**Significatif au seuil de 5 % ;

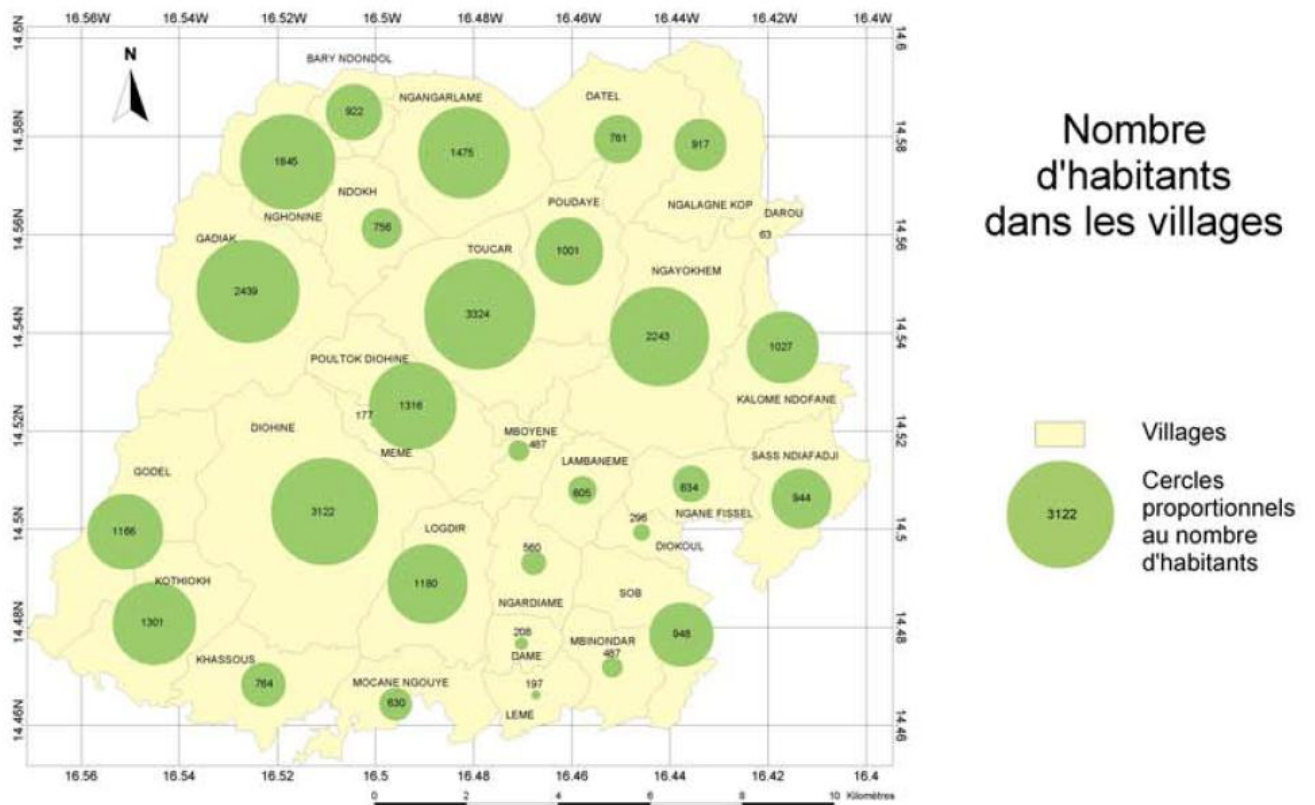
*Significatif au seuil de 10 %.

Malgré que les résultats ne soient pas tous statistiquement significatifs, on a observé précédemment que les variables « niveau d'instruction », « niveau de vie » et « niveau de RAP » ont des effets sur le niveau de fécondité des femmes en âge de procréer. Plus le niveau d'instruction de la femme est élevé, moins son nombre d'enfants l'est. Il en est de même avec le niveau de vie. Par contre à l'inverse, plus le niveau de RAP est élevé au plus la femme a de chance d'avoir un nombre d'enfants élevé.

3.7 Fécondité et village

L'étude de la fécondité selon le village va permettre de savoir s'il y a une différence entre les niveaux de fécondité selon le village de résidence de la femme durant la période d'étude 1984-2011. Les villages de Toucar et Dohine comprennent les effectifs les plus importants de femme, respectivement 11% et 10% des femmes résident dans ces villages. Alors que Darou et Diokoul sont les villages où les femmes sont les moins nombreuses, ils comprennent moins de 1% de l'effectif des femmes. Cette répartition des femmes en fonction des villages correspond à la répartition de la population générale de 2002 (Figure 19).

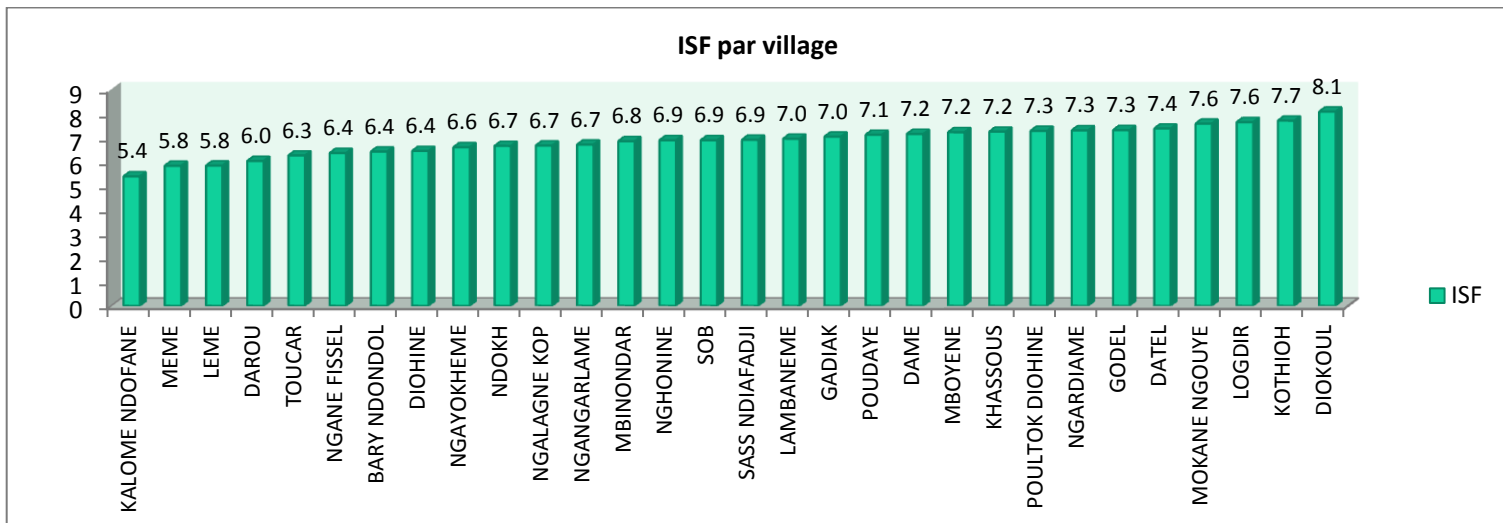
Figure 19 : Zone de Niakhar, Population, 2002



Zone de Niakhar, Population – Caractères généraux, Données IRD de 2002.

L'ISF par village varie entre 5 et 8 enfants par femme. Le village où les femmes ont l'indice le plus élevé est Diokoul, celui où il est le moins élevé est Kalome N dofane (Figure 20 et Figure 21). Il serait intéressant d'approfondir cette étude afin de savoir s'il y a une explication à cette différence (distance avec les écoles les plus proches et les postes de santé, travail des terres plus important, etc.).

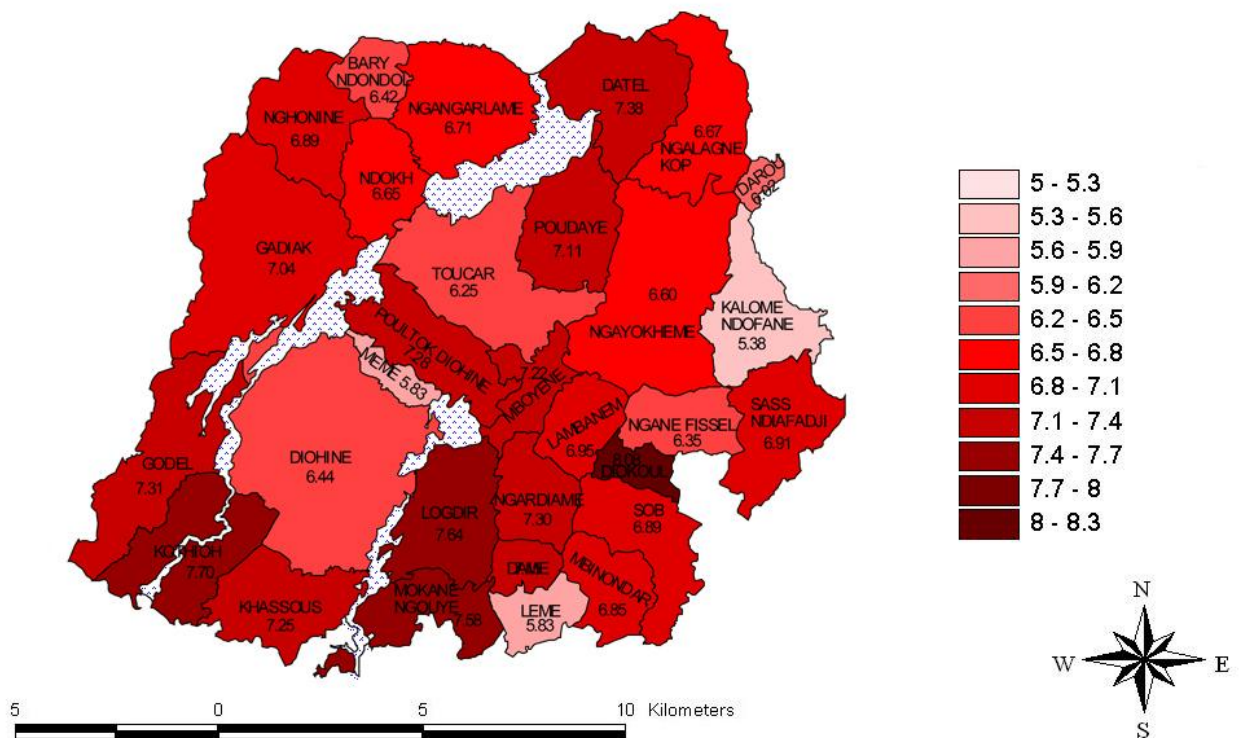
Figure 20 : Indice Synthétique de Fécondité en fonction du village de résidence de la femme, Niakhar, 1984-2011



Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

Figure 21 : Représentation cartographique : Indice Synthétique de Fécondité en fonction du village de résidence de la femme, Niakhar, 1984-2011

Indice synthétique de fécondité



Carte réalisée par Gilles Chauvancy, UMR 198 – URMITE - IRD de Dakar

Conclusion

L'analyse des biographies est appelée à jouer un rôle de plus en plus important en sciences sociales. Elle permet de prendre en considération les interactions entre évènements et permet également de tenir compte des réciprocitys entre évènements. Dans notre cas elle a surtout servi à faire une analyse transversale de la fécondité. Cette analyse a été réalisée par l'application de procédures spécifiques dans le but d'obtenir les taux de fécondité par année ou période et groupe d'âge. La démarche d'analyse est complexe et demande une mise en forme des données précise. Pour y parvenir plusieurs grandes étapes sont essentielles. Il faut avant tout mettre en forme les données pour ensuite exécuter les procédures spécifiques d'analyse permettant d'obtenir les résultats souhaités sur les temps de durée, les probabilités de survie et les taux correspondant à l'évènement étudié, ici la fécondité.

Par l'application de la méthode d'analyse des biographies nous avons obtenu les indicateurs transversaux de la fécondité de la zone de Niakhar de 1983 à 2011. Le nombre d'enfants par femme est en baisse depuis les années 1980, passant de 8 enfants par femme en 1984 à 6 enfants par femme en 2011. Le nombre de femmes en âge de procréer (15-49 ans), quant à lui, s'est accru, passant de 5343 femmes de 15-49 ans en 1984 à 9672 femmes de 15-49 ans en 2011. Bien que les femmes aient en moyenne moins d'enfants que leurs mères, le nombre absolu de naissances continue d'augmenter à cause de l'accroissement du nombre total de femme en âge de procréer, passant de 1182 naissances en 1984 à 1493 naissances en 2011. En ce qui concerne les déterminants de la fécondité, le niveau d'instruction de la femme, ainsi que le niveau de vie et le niveau de richesse agro-pastorale du ménage dans lequel la femme vit, ont une influence sur son niveau de fécondité. De manière générale, le niveau de fécondité des femmes baisse plus son niveau d'instruction est élevé. Il en est de même pour le niveau de vie, plus il est élevé plus le niveau de fécondité de la femme a tendance à diminuer. Par contre plus le niveau de richesse agro-pastorale augmente, plus le niveau de fécondité de la femme augmente aussi. La variable religion quant à elle n'a pas vraiment d'effet sur le niveau de fécondité des femmes.

Ces résultats ne sont que le début de l'analyse de la fécondité et demanderaient à être plus approfondis pour en savoir davantage sur les explications de la variabilité de la fécondité.

CONCLUSION du stage

A travers ce travail l'objectif était, d'une part, d'obtenir les principaux résultats sur la fécondité de la zone rurale de Niakhar au Sénégal entre 1983 et 2011 et, d'autre part, d'obtenir un processus d'analyse permettant l'analyse des données issues du système de suivi démographique de Niakhar.

Entre la compréhension de la méthode statistique utilisée et la prise en main du logiciel Stata cette mission s'est révélée délicate et a demandé un temps précieux à la réalisation des programmes d'analyses. Ce travail a été essentiel pour effectuer l'analyse de la fécondité et obtenir les indicateurs de la fécondité. L'analyse de la fécondité réalisée ici avec la méthode des biographies semble pertinente, effectivement les résultats obtenus sont semblables aux résultats obtenus dans le cadre d'études antérieures réalisées par le biais d'autres méthodes d'analyses.

Le travail effectué dans le cadre de ce stage a été utile sur plusieurs points : - Réalisation de programmes d'analyse ; - Détection d'incohérences dans les données et correction de ces incohérences dans la base de données ; - Principaux résultats sur l'évolution de la fécondité et de ses déterminants.

D'un point de vue plus personnel, ce stage a été intéressant et le réaliser au sein de l'IRD de Dakar a été très enrichissant. J'ai travaillé sur une méthode statistique complexe délicate à mettre en œuvre qui dépassait le cadre des enseignements théoriques reçus. De plus, pouvoir mettre en application des analyses sur des données de suivi démographique dans un cadre professionnel est vraiment instructif et rend l'analyse plus concrète et plus intéressante. Ce stage, réalisé à l'étranger dans un environnement dépaysant, a été d'autant plus formateur en me demandant un sens d'adaptabilité et d'intégration plus important.

Par la réalisation de ce stage dans le monde de la recherche, j'en suis venue à me poser des questions quant à mon avenir professionnel. Au jour d'aujourd'hui je me demande s'il ne serait pas intéressant de continuer dans cette voie la par le biais d'un poste de VI (Volontaire International) au sein de l'IRD dans cette même unité de recherche ou pourquoi pas de m'orienter sur une thèse.

BIBLIOGRAPHIE

Daniel Courgeau and Eva Lelièvre, *L'approche biographique en démographie*, (Revue française sociologie, 1989) pp. 55-74.

Donatien Beguy, Philippe Bocquier and Eliya Msiyaphazi Zulu, *Circular migration patterns and determinants in Nairobi slum settlements*, (Demographic Research, volume 23, article 20, sept 2010) pp. 549-586.

Eva Lelièvre and Arnaud Bringé, *Manuel pratique pour l'analyse statistique des biographies, Présentation des modèles de durée et utilisation des logiciels SAS, TDA, STATA*, Editions de l'Institut National d'Etudes Démographiques (Paris: PUF/DIFFUSION, 1998).

Georges de Noni, Jacques Panfili, etc., *La représentation au Sénégal, Institut de Recherche pour le Développement*, (Rapport de l'IRD, avril 2011).

Géraldine Duthé, *La transition sanitaire en milieu rural sénégalais. Evolution de la mortalité à Mlomp depuis 1985 et influence du paludisme chimiorésistant*. (Thèse de démographie, 2006).

Gilles Pison, *Les Changements démographiques au Sénégal* (INED, 1997).

Valérie Delaunay, *L'entrée en vie féconde, expression démographique des mutations socio-économiques d'un milieu rural sénégalais* (Les études du Ceped n°7, 1994).

Valérie Delaunay (Coord.), *La situation démographique et épidémiologique dans la zone de Niakhar au Sénégal 1984-1996* (Laboratoire population et santé ORSTOM, 1998).

Valérie Delaunay, Agnès Adjamagbo and Richard Lalou, *Questionner la transition de la fécondité en milieu rural africain : les apports d'une démarche longitudinale et institutionnelle* (Cahiers québécois de démographie 35, n° 1, 2006).

Valérie Delaunay, Adama Marra and Pierre Levi, *Analyser la fécondité à partir de données de Systèmes de Suivi Démographique, Application au site de Niakhar, Sénégal* (Les collections du CEPED, 2006).

Philippe Bocquier, *L'analyse des enquêtes biographiques à l'aide du logiciel STATA* (Paris : Ceped, 1996).

Philippe Bocquier and Philippe Antoine, *L'apport de l'analyse des biographies aux sciences sociales*.

Adresse url :

http://horizon.documentation.ird.fr/exl-doc/pleins_textes/pleins_textes_7/b_fdi_03_03/41672.pdf

Théodore Hounbégnon and Atikpo Mawunya Adesu, *Construction d'indicateurs de niveau de vie des ménages et relation avec la migration de travail dans une zone du Sénégal : Niakhar* (Mémoire de fin de cycle-ITS, IRD UMR 198, 2012).

ANNEXES

ANNEXES A : LES PROGRAMMES STATA

ANNEXE A.1: Programme « structure_standard.do »

```

*****
** CREATION DE LA TABLE STANDARD POUR L'ANALYSE DES DUREES DE SEJOURS : **
** "RESIDENCY.DTA" **
*****

clear
set mem 1000000
*enregistre les tables dans le dossier "TABLES"*
cd "D:\user\Desktop\stage IRD\TABLES"
*ouverture table de données resid_corr.dta (données de début 1983 à mi 2012)*
use "D:\user\Desktop\stage IRD\Tables_Bases_Niakhar\resid_corr.dta", clear
*nommer la table yourepisodefile.dta (structure de la table initiale = en épisode)
save yourepisodefile.dta, replace
*****
/* TEST stset => permet d'avoir un aperçu du nombre de données avec incohérences
dans les dates ou autres problèmes (nombre d'observations supprimées de l'analyse)
test sur les données allant du 1erjanvier1983 au 31décembre2011, évènement étudié =
décès */
stset datex if datnais~., id(id) failure(exit==51) origin(time datnais)
scale(365.25) enter(time mdy(1,1,1983)) exit(time mdy(12,31,2011)) time0(daten)
/*
on remarque qu'il y a des INCOHERENCES pour certaines observations
=> datex[_n-1]>daten
=> daten>datex
en tenir compte dans la suite de l'analyse
et corriger les incohérences, si possible
*/
*****
*RENOMMER LES VARIABLES (sur le modèle standard Indepth proposé par P.B)*
*****
use yourepisodefile.dta, replace
*sélection des variables
keep id vil datnais entry daten exit datex sexe idmere
*renomme les noms des variables (entry exit, etc. en EventCode EventDate,..)
rename id IndividualId
rename vil LocationId
rename datnais DoB
rename idmere IdMere
rename sexe Sex
rename exit EndEventCode
rename datex EndEventDate
rename entry StartEventCode

```

```

rename daten StartEventDate
save yourepisodefile, replace

*****
*      MISE EN FORME DE LA TABLE (1 ligne = événement)      *
*****

***Crée un fichier avec seulement les événements de fin
use yourepisodefile.dta
keep IndividualId LocationId DoB Sex EndEventCode EndEventDate IdMere ///
gen file=1
rename EndEventCode EventCode
rename EndEventDate EventDate
sort IndividualId EventDate
save endeventfile, replace
***Crée un fichier avec seulement les événements de début
use yourepisodefile.dta
keep IndividualId LocationId DoB Sex StartEventCode StartEventDate IdMere ///
gen file=2
rename StartEventCode EventCode
rename StartEventDate EventDate
sort IndividualId EventDate
save starteventfile, replace
****Assemble les 2 fichiers (Evènements de début et de fin) pour avoir le format
EHA
use starteventfile, clear
append using endeventfile
sort IndividualId EventDate EventCode
save residency, replace
*****

*****
*Vérification des codes des événements*
*****

tab EventCode,m
/*Se débarrasser des lignes dont les codes sont absurdes
   par exemple supprimer si EventCode==. ou ==-1 ou ==0 */

*****
*      RENOMMER LES CODES DES EVENEMENTS      *
*****

*Création de la variable EventCode_Initial pour garder une copie des anciens codes
cap drop EventCode_Initial
gen EventCode_Initial=EventCode
*Attribuer les codes standards aux événements (modèle de P.B.)
replace EventCode=1 if EventCode==11
replace EventCode=2 if EventCode==12
replace EventCode=3 if EventCode==13
replace EventCode=4 if EventCode==71
replace EventCode=5 if EventCode==61

```



```

replace EventCode=6 if EventCode==62
replace EventCode=7 if EventCode==51
replace EventCode=9 if EventCode==91
*****
*Création des labels
*****
lab var EventCode "Event occurred"
label define eventlab 1 "ENU" 2 "BTH" 3 "IMG" 4 "OMG" 5 "EXT" 6 "ENT" ///
7 "DTH" 8 "CHD" 9 "OBE" 18 "OBS" 19 "OBL" 20 "1Jan" 21 "NewAgeGroup", modify
lab val EventCode eventlab
*****
* Transformer le format des variables "dates" en "double": %tc
* Si EventDate et DoB sont des variables sous le format "int": %td ou %d
format EventDate DoB %td
local formatdate: format EventDate
if ("`formatdate'"=="%d" | "`formatdate'"=="%td") {
    replace EventDate=cofd(EventDate)
}
local formatdate: format DoB
if ("`formatdate'"=="%d" | "`formatdate'"=="%td") {
    replace DoB=cofd(DoB)
}
format DoB %tc
format EventDate %tc
*****
* RECTIFICATION DES DATES
*****
* AJOUTE un demi jours (12h00) à toute les dates (suggestions de Sam Clark et Kobus
Herbst)
replace DoB=DoB + (12*60*60*1000)
replace EventDate=EventDate + (12*60*60*1000)

***IDENTIFIE LES INDIVIDUS QUI SONT MORTS LE MEME JOURS QU'ILS SONT NES
capture drop same_dob_dod
gen same_dob_dod=1 if DoB==EventDate & EventCode==7
browse if same_dob_dod==1
* AJOUTE 6 heures à la date de décès
replace EventDate=EventDate + (6*60*60*1000) if same_dob_dod==1
drop same_dob_dod

***IDENTIFIE LES INDIVIDUS QUI ONT EMIGRES LE MEME JOURS QU'ILS SONT NES
capture drop same_dob_omg
gen same_dob_omg=1 if DoB==EventDate & EventCode==4
browse if same_dob_omg==1
* AJOUTE 6 heures à la date d'émigration
replace EventDate=EventDate + (6*60*60*1000) if same_dob_omg==1
drop same_dob_omg

***IDENTIFIE SORTIE ET ENTREE LE MEME JOURS

```

```

capture drop same_exit_entry
gen same_exit_entry=1 if EventDate[_n-1]==EventDate ///
                        & EventCode[_n-1]==5 & EventCode==6
browse if same_exit_entry==1
* AJOUTE 6 heures à la date d'entrée
replace EventDate=EventDate + (6*60*60*1000) if same_exit_entry==1
drop same_exit_entry

*****
*CREATION VARIABLE datebeg(date de début)*
*****
cap drop datebeg
qui by IndividualId: gen double datebeg=cond(_n==1, DoB, EventDate[_n-1])
format datebeg %tc
lab var datebeg "Date of beginning"

order IndividualId LocationId DoB Sex datebeg EventDate EventCode
compress
save residency, replace

*****
* MATRICE POUR REPERER LES INCOHERENCES DE L'ORDRE DES EVENEMENTS :
*****
sort IndividualId EventDate EventCode
capture drop foll_EventCode
qui by IndividualId: gen foll_EventCode=EventCode[_n+1]
lab var foll_EventCode "Following event"
lab val foll_EventCode eventlab
tab EventCode foll_EventCode, missing

* Repérer les erreurs dans la matrice, calculer le pourcentage d'erreur
/* si > à 2 % => trop d'erreur pour une bonne analyse *cf. feuille Excel
"Incohérences" */

/*
**EXEMPLES POUR IDENTIFIER LES INDIVIDUS AVEC DES INCOHERENCES DANS LES
OBSERVATIONS
**et sauvegarde de la liste dans un fichier log
log using inconsistencies, replace
list IndividualId DoB EventDate EventCode foll_EventCode ///
        if EventCode==7 & foll_EventCode==1, sepyby(IndividualId)
br IndividualId DoB EventDate EventCode foll_EventCode if EventCode==7 &
foll_EventCode==1
*faire de même pour les autres incohérences,
/*si décès suivi d'un autre évènement,
    si l'évènement naissance est précédé par un autre évènement, etc.*
*/
*AFFICHE SI DECES SUIVI D'UN AUTRE EVENEMENT*
br IndividualId DoB EventDate EventCode foll_EventCode if EventCode==7 &
foll_EventCode==1

```

```

br IndividualId DoB EventDate EventCode foll_EventCode if EventCode==7 &
foll_EventCode==2
br IndividualId DoB EventDate EventCode foll_EventCode if EventCode==7 &
foll_EventCode==3
br IndividualId DoB EventDate EventCode foll_EventCode if EventCode==7 &
foll_EventCode==4
br IndividualId DoB EventDate EventCode foll_EventCode if EventCode==7 &
foll_EventCode==5
br IndividualId DoB EventDate EventCode foll_EventCode if EventCode==7 &
foll_EventCode==6
br IndividualId DoB EventDate EventCode foll_EventCode if EventCode==7 &
foll_EventCode==7
br IndividualId DoB EventDate EventCode foll_EventCode if EventCode==7 &
foll_EventCode==9
*AFFICHE SI EVENEMENT SUIVI D'UNE NAISSANCE*
br IndividualId DoB EventDate EventCode foll_EventCode if EventCode==1 &
foll_EventCode==2
br IndividualId DoB EventDate EventCode foll_EventCode if EventCode==2 &
foll_EventCode==2
br IndividualId DoB EventDate EventCode foll_EventCode if EventCode==3 &
foll_EventCode==2
br IndividualId DoB EventDate EventCode foll_EventCode if EventCode==4 &
foll_EventCode==2
br IndividualId DoB EventDate EventCode foll_EventCode if EventCode==5 &
foll_EventCode==2
br IndividualId DoB EventDate EventCode foll_EventCode if EventCode==6 &
foll_EventCode==2
br IndividualId DoB EventDate EventCode foll_EventCode if EventCode==9 &
foll_EventCode==2
*etc...*

***IDENTIFIE LES INDIVIDUS QUI ONT LA MEME DATE D'EVENEMENT COURANT ET PRECEDANT
capture drop same_datebeg
egen same_datebeg=max(datebeg==datebeg[_n-1] & datebeg!=DoB), by(IndividualId)
browse IndividualId DoB EventDate EventCode foll_EventCode if same_datebeg==1

log close

*/

```

ANNEXE A.2 : Programme « censure_OBE.do »

```

*****
* CENSURE A DROITE = DATE DE FIN D'OBSERVATION *
*****

* Coupe les données au 1er Janv 2012 00:00:00
***choix de la date de censure à droite = date de fin d'observation***
*ouverture fichier residency.dta
use "D:\user\Desktop\stage IRD\TABLES\residency.dta", clear
display %20.0f clock("1Jan2012", "DMY")
*1640995200000 (1erjan2012)

```

```

*replace évènement=OBL si évènement=OBE et date<1jan2012
      /*normalement aucun OBL car attribution du code 9 seulement si l'individu est
      considéré dans la zone en attendant le prochain passage (ici date du 1jan2015)*/
replace EventCode=19 if (EventCode==9 & EventDate<1640995200000)

*crée la variable lastrecord pour la dernière observation de chaque individu
sort IndividualId EventDate EventCode
cap drop lastrecord
qui by IndividualId: gen lastrecord=_n==_N
*analyse de la durée du dernier séjours
/* besoin de la durée de ce séjour pour couper celui-ci à la date choisie
   ici au 1jan2012 */
*** utiliser time0(datebeg) systématiquement ***
stset EventDate, id(IndividualId) failure(lastrecord==1) time0(datebeg)

* Crée une nouvelle observation au 1 jan 2012 00:00:00 (pour les obs supérieures à
cette date)
* svt pour les individus présents ds la zone = code 9
capture drop lastobs
stsplint lastobs, at(1640995200000)
sort IndividualId EventDate EventCode
*supprime les variables de durées
drop lastrecord
drop _*
*****
* Supprime les observations supérieures au 1er Jan 2012 00:00:00 (censure à droite)
* cad si les observations sont supérieures à la date du 01 Janv 2012
drop if EventDate>1640995200000

* Crée une ligne d'observation supplémentaire
* pour ceux qui sont morts ou ont émigrés avant le 01Jan2012 00:00:00
sort IndividualId EventDate EventCode
expand=2 if IndividualId!=IndividualId[_n+1] & EventDate<1640995200000 , ///
      gen(duplicate)
*recodage pour ces individus
sort IndividualId EventDate EventCode duplicate
qui by IndividualId: replace EventDate=1640995200000 if duplicate==1
* Pour tout les individus le dernier évènement est OBE
qui by IndividualId: replace EventCode=9 if _n==_N
qui by IndividualId: replace datebeg=cond(_n==1, DoB, EventDate[_n-1])
drop duplicate
drop lastobs
* => Toutes les dernières observations sont au 1 Jan 2012 00:00:00

*refaire tourner les labels et la MATRICE DES INCOHERENCES*
sort IndividualId EventDate EventCode
capture drop foll_EventCode
qui by IndividualId: gen foll_EventCode=EventCode[_n+1]
lab var foll_EventCode "Following event"
label define eventlab 1 "ENU" 2 "BTH" 3 "IMG" 4 "OMG" 5 "EXT" 6 "ENT" ///

```

```

7 "DTH" 8 "CHD" 9 "OBE" 10 "DLV" 18 "OBS" 19 "OBL" 20 "1Jan" 21 "NewAgeGroup",
modify
lab val foll_EventCode eventlab
tab EventCode foll_EventCode, missing
compress

save residency.dta, replace

```

ANNEXE A.3 : Programme « création_résidence.do »

```

*****
*   CREATION DE LA VARIABLE RESIDENCE                               *
*****
use "D:\user\Desktop\stage IRD\TABLES\residency.dta", clear
***Crée la variable residence (permet de repérer quand l'individu est résident de
la zone)***
sort IndividualId EventDate EventCode
br IndividualId DoB datebeg EventDate EventCode foll_EventCode

* 6 mois c'est :
display %20.0f (365.25 * 24 * 60 * 60 * 1000)/2
*15778800000*
capture drop residence
gen residence=.

***Recensement Initial
qui by IndividualId: replace residence=0 if EventCode==1 & (EventCode[_n+1]!=1)
***NAISSANCE
replace residence=0 if residence==. & EventCode==2
***SORTIE
replace residence=1 if residence==. & EventCode==5
***DECES
replace residence=1 if residence==. & EventCode==7
***ENTREE
**Cases avec ENTREE précédées par SORTIE d'une durée de moins de 180-jours
qui by IndividualId: replace residence=1 if residence==. & ///
(EventCode==6 & EventCode[_n-1]==5 & EventDate-EventDate[_n-1]<=15778800000)
**Cases avec ENTREE précédées par SORTIE d'une durée de plus de 180-jours
qui by IndividualId: replace residence=0 if residence==. & ///
(EventCode==6 & EventCode[_n-1]==5 & EventDate-EventDate[_n-1]>15778800000)
**Cases avec ENTREE précédées par EMIGRATION d'une durée de moins de 180-jours
qui by IndividualId: replace residence=1 if residence==. & ///
(EventCode==6 & EventCode[_n-1]==7 & EventDate-EventDate[_n-1]<=15778800000)
**Cases avec ENTREE précédées par EMIGRATION d'une durée de plus de 180-jours
qui by IndividualId: replace residence=0 if residence==. & ///
(EventCode==6 & EventCode[_n-1]==7 & EventDate-EventDate[_n-1]>15778800000)
**Cases ENTREE comme 1er évènement (issue normalement impossible)
qui by IndividualId: replace residence=0 if residence==. & ///
EventCode==6 & _n==1

```

```

***IMMIGRATION
**Cases avec IMG precedées de OMG d'une durée de moins de 180-jours
qui by IndividualId: replace residence=1 if residence==. & ///
  (EventCode==3 & EventCode[_n-1]==4 & EventDate-EventDate[_n-1]<=15778800000)
**Cases avec IMG precedées de OMG d'une durée de plus de 180-jours
qui by IndividualId: replace residence=0 if residence==. & ///
  (EventCode==3 & EventCode[_n-1]==4 & EventDate-EventDate[_n-1]>15778800000)
qui by IndividualId: replace residence=0 if residence==. & EventCode==3 &
EventCode[_n-1]==.

***EMIGRATION
qui by IndividualId: replace residence=1 if residence==. & EventCode==4
qui by IndividualId: replace residence=0 if residence==1 & ///
  (EventCode==4 & (EventCode[_n-1]==4 | EventCode[_n-1]==3) & ///
  EventDate-EventDate[_n-1]<=15778800000)

***OBL (dernière observation)
qui by IndividualId: replace residence=1 if residence==. & EventCode==19
* Remplace par non-residence si OBL est précédé de OMG or EXT or DTH
qui by IndividualId: replace residence=0 if residence==1 & EventCode==19 ///
  & (EventCode[_n-1]==4 | EventCode[_n-1]==5 |
EventCode[_n-1]==7)

***OBE
qui by IndividualId: replace residence=1 if EventCode==9 ///
  & EventCode[_n-1]!=4 & EventCode[_n-1]!=5 & EventCode[_n-1]!=7 &
EventCode[_n-1]!=19
qui by IndividualId: replace residence=0 if EventCode==9 ///
  & (EventCode[_n-1]==4 | EventCode[_n-1]==5 | EventCode[_n-1]==7 |
EventCode[_n-1]==19)

* Sélectionne les cases avec valeurs manquantes pour la résidence
capture drop misresidence
egen misresidence=max(residence==.), by(IndividualId)
sort IndividualId EventDate EventCode
br IndividualId datebeg EventDate EventCode residence if misres==1

tabulate residence EventCode, missing

*** Corriger les incohérences si possible
/* !!! Refaire tourner les indications de la variable résidence
  et des valeurs manquantes après chaque correction !!! */

** Exemples de corrections :
replace residence=0 if residence==. & misres==1 & EventCode==6
replace residence=0 if residence==. & misres==1 & EventCode==3

* recoder les valeurs manquantes "residence" en 0 (=non resident)
recode residence .=0

```

```

tabulate residence EventCode, missing

compress

save residency, replace

```

ANNEXE A.4 : Programme pour créer la table child.dta et pour merger la table residency et child « merge_fécondite.do »

```

*****
* TABLE FECONDITE : AJOUT DE L'EVENEMENT NAISSANCE *
*
*           DANS LE PARCOURS DES FEMMES *
*****
* PROGRAMME POUR MERGER LES EVENEMENTS NAISSANCES DANS LA TABLE STANDARD *
cd "D:\user\Desktop\stage IRD\TABLES"
use "D:\user\Desktop\stage IRD\TABLES\residency.dta", clear

use residency.dta, clear
keep IndividualId EventDate
keep if EventDate==1640995200000
duplicates drop IndividualId, force
sort IndividualId
save ind_residency.dta, replace

* CREATION DE LA TABLE child.dta *
/* Table des naissances enregistrées dans la zone
   contenant les variables suivantes :
   IndividualId ChildId ChildDoB ChildSex ...
   * NB: IndividualId c'est l'ID de la mère
   * NB: ChildId c'est l'Id de l'enfant
   */

use residency.dta, clear
save child.dta, replace
keep IndividualId IdMere DoB EventCode EventDate Sex
keep if EventCode==2
keep if IdMere>0
sort IdMere DoB
drop EventDate
rename DoB ChildDoB
rename IndividualId ChildId
rename IdMere IndividualId
rename Sex ChildSex
sort IndividualId ChildDoB
/*
* si la variable ChildDoB est sous le format "int": %td or %d
* transformer le format de la variable en "double": %tc
*(ici pas besoin)
format ChildDoB %td
local formatdate: format ChildDoB

```

```

if ("`formatdate'"=="%d" | "`formatdate'"=="%td") {
    replace ChildDoB=cofd(ChildDoB)
}
format ChildDoB %tc

*AJOUTE 12h00 à la date de naissance !! seulement si c'est pas déjà fait*
replace ChildDoB=ChildDoB + (12*60*60*1000)
*/
rename ChildDoB EventDate
*code naissance de l'enfant = 8*
replace EventCode=8
*supprime si la date de naissance est supérieure ou égale au 1er jan 2012
drop if EventDate>=1640995200000
sort IndividualId EventDate
order IndividualId EventDate EventCode
save child.dta, replace

use ind_residency.dta, clear
keep IndividualId EventDate
append using child
sort IndividualId EventDate
drop if EventDate==.
egen temp=max(EventDate==1640995200000), by(IndividualId)
br if temp==0
/*on supprime les naissances dont on n'a pas d'information
sur le parcours de la mere (13 individus qui sont entrés par naissance mais dont on
ne connaît pas l'Id de la mère*/
drop if temp==0
drop temp
save child.dta, replace
erase ind_residency.dta

**** MERGE TABLE residency AVEC LA TABLE des enfants ****
use residency, clear
cap drop _merge
merge m:m IndividualId EventDate using child.dta
sort IndividualId EventDate
format EventDate %tc
sort IndividualId EventDate EventCode

/* Enleve 6 heures à la date de naissance de l'enfant
si simultanéité avec d'autres évènements IMG, OMG, EXT, DTH, OBS et OBL */
capture drop duplicate
expand=2 if _merge==3 & EventDate<1640995200000, gen(duplicate)
sort IndividualId EventDate EventCode duplicate
replace EventDate=EventDate - (6*60*60*1000) if duplicate==1 & ///
(EventCode==3 | EventCode==4 | EventCode==5 | EventCode==7 |
EventCode==18 | EventCode==19)
replace EventCode=8 if duplicate==1
drop duplicate

```



```

sort IndividualId EventDate EventCode

label define eventlab 1 "ENU" 2 "BTH" 3 "IMG" 4 "OMG" 5 "EXT" 6 "ENT" ///
7 "DTH" 8 "CHD" 9 "OBE" 10 "DLV" 18 "OBS" 19 "OBL" 20 "1Jan" 21 "NewAgeGroup",
modify
lab val EventCode eventlab

* Réajuste le temps de début = variable datebeg
by IndividualId: replace datebeg=cond(_n==1, DoB, EventDate[_n-1])
* réajuste la variable résidence
replace residence=1 if EventCode==8

*AJOUTE LES INFOS SUR LES AUTRES VARIABLES *
by IndividualId : replace LocationId=LocationId[_N] if EventCode==8
by IndividualId : replace DoB=DoB[_N] if EventCode==8
by IndividualId : replace Sex=Sex[_N] if EventCode==8
by IndividualId : replace relg=relg[_N] if EventCode==8
by IndividualId : replace ethg=ethg[_N] if EventCode==8
by IndividualId : replace casteg=casteg[_N] if EventCode==8
by IndividualId : replace tim=tim[_N] if EventCode==8
by IndividualId : replace rel=rel[_N] if EventCode==8
by IndividualId : replace instruct=instruct[_N] if EventCode==8
by IndividualId : replace instg1=instg1[_N] if EventCode==8
by IndividualId : replace instg2=instg2[_N] if EventCode==8

***MATRICE DES INCOHERENCES ***
sort IndividualId EventDate EventCode
capture drop foll_EventCode
qui by IndividualId: gen foll_EventCode=EventCode[_n+1]
lab var foll_EventCode "Following event"
lab val foll_EventCode eventlab
tab EventCode foll_EventCode, missing

/*on remarque des incohérences*/

*SUPPRESSION des NAISSANCES considérées comme HORS ZONE*
*****
*pbm incohérences ENFANT puis RI
br if EventCode==8 & EventCode[_n+1]==1 & EventDate!=EventDate[_n+1]
drop if EventCode==8 & EventCode[_n+1]==1 & EventDate!=EventDate[_n+1]
sort IndividualId EventDate
*8 obs*
*pbm incohérences ENFANT puis IMMIGRATION
br if EventCode==8 & EventCode[_n+1]==3 & EventDate!=EventDate[_n+1]
drop if EventCode==8 & EventCode[_n+1]==3 & EventDate!=EventDate[_n+1]
sort IndividualId EventDate
drop if EventCode==8 & EventCode[_n+1]==3 & EventDate!=EventDate[_n+1]
sort IndividualId EventDate
drop if EventCode==8 & EventCode[_n+1]==3 & EventDate!=EventDate[_n+1]

```

```

sort IndividualId EventDate
drop if EventCode==8 & EventCode[_n+1]==3 & EventDate!=EventDate[_n+1]
sort IndividualId EventDate
drop if EventCode==8 & EventCode[_n+1]==3 & EventDate!=EventDate[_n+1]
sort IndividualId EventDate
drop if EventCode==8 & EventCode[_n+1]==3 & EventDate!=EventDate[_n+1]
sort IndividualId EventDate
drop if EventCode==8 & EventCode[_n+1]==3 & EventDate!=EventDate[_n+1]
sort IndividualId EventDate
drop if EventCode==8 & EventCode[_n+1]==3 & EventDate!=EventDate[_n+1]
sort IndividualId EventDate
drop if EventCode==8 & EventCode[_n+1]==3 & EventDate!=EventDate[_n+1]
sort IndividualId EventDate
drop if EventCode==8 & EventCode[_n+1]==3 & EventDate!=EventDate[_n+1]
sort IndividualId EventDate
drop if EventCode==8 & EventCode[_n+1]==3 & EventDate!=EventDate[_n+1]
sort IndividualId EventDate

*pbm incohérences DECES PUIS ENFANT (11 obs sup)
*!!répéter la commande jusqu'à effacement de toutes les obs = 0 obs deleted!!*
br if EventCode==8 & EventCode[_n-1]==7 & EventDate!=EventDate[_n-1]
drop if EventCode==8 & EventCode[_n-1]==7 & EventDate!=EventDate[_n-1]
sort IndividualId EventDate
drop if EventCode==8 & EventCode[_n-1]==7 & EventDate!=EventDate[_n-1]
sort IndividualId EventDate
drop if EventCode==8 & EventCode[_n-1]==7 & EventDate!=EventDate[_n-1]
sort IndividualId EventDate
drop if EventCode==8 & EventCode[_n-1]==7 & EventDate!=EventDate[_n-1]
sort IndividualId EventDate

*pbm incohérences EMIGRATION PUIS ENFANT (77sup)
br if EventCode==8 & EventCode[_n-1]==4 & EventDate!=EventDate[_n-1]
drop if EventCode==8 & EventCode[_n-1]==4 & EventDate!=EventDate[_n-1]
drop if EventCode==8 & EventCode[_n-1]==4 & EventDate!=EventDate[_n-1]
drop if EventCode==8 & EventCode[_n-1]==4 & EventDate!=EventDate[_n-1]
drop if EventCode==8 & EventCode[_n-1]==4 & EventDate!=EventDate[_n-1]
drop if EventCode==8 & EventCode[_n-1]==4 & EventDate!=EventDate[_n-1]
drop if EventCode==8 & EventCode[_n-1]==4 & EventDate!=EventDate[_n-1]
sort IndividualId EventDate

***MATRICE DES INCOHERENCES ***
sort IndividualId EventDate EventCode
capture drop foll_EventCode
qui by IndividualId: gen foll_EventCode=EventCode[_n+1]
lab var foll_EventCode "Following event"
lab val foll_EventCode eventlab
tab EventCode foll_EventCode, missing

* Réajuste le temps de début = variable datebeg
by IndividualId: replace datebeg=cond(_n==1, DoB, EventDate[_n-1])
* réajuste la variable résidence
replace residence=1 if EventCode==8

```

```

order IndividualId LocationId DoB Sex datebeg EventDate EventCode
erase child.dta
compress
save resid_fec, replace

```

ANNEXE A.5 : Programme pour la création du calendrier et des groupes d'âges "stspllit_calendrier_année.do"

```

*****
*   CREATION DU CALENDRIER PAR ANNEES   *
*****

*use "D:\user\Desktop\stage IRD\TABLES\residency.dta", clear

*table avec évènement naissance*
use "D:\user\Desktop\stage IRD\TABLES\resid_fec.dta", clear
* coupe les données à chaque 1 Jan 00:00:00
* de 1983 à 2011 (choix des dates ici)

***Sorties des valeurs numériques correspondant au dates
forval num=1983/2012 {
    display %20.0f clock("1Jan`num'", "DMY")
}

sort IndividualId EventDate EventCode
cap drop lastrecord
qui by IndividualId: gen lastrecord=_n==_N
*** ATTENTION: mettre time0(datebeg) systématiquement
*et faire l'analyse seulement si résidence==1, cad sur les période de résidence de
l'individu.
stset EventDate if residence==1, id(IndividualId) failure(lastrecord==1)
time0(datebeg)

* Crée les nouvelles observations pour chaque 1 Jan 00:00:00, pour chaque année
capture drop lastobs
capture drop calendar_year
stspllit calendar_year, at(725846400000, ///
    757382400000, ///
    789004800000, ///
    820540800000, ///
    852076800000, ///
    883612800000, ///
    915235200000, ///
    946771200000, ///
    978307200000, ///
    1009843200000, ///
    1041465600000, ///

```

```
1073001600000, ///  
1104537600000, ///  
1136073600000, ///  
1167696000000, ///  
1199232000000, ///  
1230768000000, ///  
1262304000000, ///  
1293926400000, ///  
1325462400000, ///  
1356998400000, ///  
1388534400000, ///  
1420156800000, ///  
1451692800000, ///  
1483228800000, ///  
1514764800000, ///  
1546387200000, ///  
1577923200000, ///  
1609459200000, ///  
    1640995200000)  
display %20.0f 365.25 * 24 * 60 * 60 * 1000  
* 1 an = 31557600000  
recode calendar_year 725846400000=1983 ///  
    757382400000=1984 ///  
    789004800000=1985 ///  
    820540800000=1986 ///  
    852076800000=1987 ///  
    883612800000=1988 ///  
    915235200000=1989 ///  
    946771200000=1990 ///  
    978307200000=1991 ///  
    1009843200000=1992 ///  
    1041465600000=1993 ///  
    1073001600000=1994 ///  
    1104537600000=1995 ///  
    1136073600000=1996 ///  
    1167696000000=1997 ///  
    1199232000000=1998 ///  
    1230768000000=1999 ///  
    1262304000000=2000 ///  
    1293926400000=2001 ///  
    1325462400000=2002 ///  
    1356998400000=2003 ///  
    1388534400000=2004 ///  
    1420156800000=2005 ///  
    1451692800000=2006 ///  
    1483228800000=2007 ///  
    1514764800000=2008 ///  
    1546387200000=2009 ///  
    1577923200000=2010 ///  

```

```

1609459200000=2011 ///
1640995200000=2012
sort IndividualId EventDate EventCode
drop lastrecord
drop _*

* Recode les EventCode
qui by IndividualId: replace EventCode=20 if EventCode==EventCode[_n+1] ///
    & calendar_year!=. & calendar_year!=calendar_year[_n+1] &
IndividualId==IndividualId[_n+1]

sort IndividualId EventDate EventCode
capture drop foll_EventCode
qui by IndividualId: gen foll_EventCode=EventCode[_n+1]
lab var foll_EventCode "Following event"
label define eventlab 1 "ENU" 2 "BTH" 3 "IMG" 4 "OMG" 5 "EXT" 6 "ENT" ///
7 "DTH" 8 "CHD" 9 "OBE" 10 "DLV" 18 "OBS" 19 "OBL" 20 "1Jan" 21 "NewAgeGroup",
modify
lab val foll_EventCode eventlab
lab val EventCode eventlab
tab EventCode foll_EventCode, missing

*** CREE LA VARIABLE AGE GROUP : 0 1 5 10 15 ... 85
display %20.0f 365.25 * 24 * 60 * 60 * 1000
display %20.0f 5 * 365.25 * 24 * 60 * 60 * 1000
display %20.0f 85 * 365.25 * 24 * 60 * 60 * 1000
sort IndividualId EventDate EventCode
cap drop lastrecord
qui by IndividualId: gen lastrecord=_n==_N
***
stset EventDate if residence==1, id(IndividualId) failure(lastrecord==1) ///
    time0(datebeg) origin(time DoB)

capture drop agegroup
stsplint agegroup if residence==1, at(0 31557600000 ///
15778800000 (15778800000)2682396000000)
replace agegroup=agegroup/31557600000
sort IndividualId EventDate EventCode
drop lastrecord
drop _*

* Recode les EventCode
label define eventlab 1 "ENU" 2 "BTH" 3 "IMG" 4 "OMG" 5 "EXT" 6 "ENT" ///
7 "DTH" 8 "CHD" 9 "OBE" 10 "DLV" 18 "OBS" 19 "OBL" 20 "1Jan" 21 "NewAgeGroup",
modify
qui by IndividualId: replace EventCode=21 if EventCode==EventCode[_n+1] ///
    & agegroup!=. & agegroup!=agegroup[_n+1] & IndividualId==IndividualId[_n+1]
capture drop foll_EventCode
qui by IndividualId: gen foll_EventCode=EventCode[_n+1]
lab var foll_EventCode "Following event"

```

```

lab val foll_EventCode eventlab
lab val EventCode eventlab
tab EventCode foll_EventCode, missing

compress
save resid_fec.dta, replace

```

ANNEXE A.6 : Programme pour l'analyse de la fécondité "analyse_fécondité.do"

```

*****
* ANALYSE DE LA FECONDITE_évènement renouvelable_ *
*****

clear
set mem 1000000
use "D:\user\Desktop\stage IRD\TABLES\resid_fec.dta", clear

*création variable de censure FECONDITE
gen byte delivery=EventCode==8
save resid_fec.dta, replace

*****
*TAUX BRUT DE NATALITE *
*****

use "D:\user\Desktop\stage IRD\TABLES\resid_fec.dta", clear
*calcul de la durée des séjours avant l'évènement naissance(fécondité) pour
l'ensemble de la population résidente
stset EventDate if residence==1, id(IndividualId) failure(delivery==1) ///
      origin(time DoB) time0(datebeg) exit(time .) scale(31557600000)
*Taux brut de natalité par année*
stptime, by(calendar) per(1000)
*création de la variable perio (période)*
cap drop perio
gen perio=0
replace perio=1 if (calendar>=1984 & calendar<=1988)
replace perio=2 if (calendar>=1989 & calendar<=1993)
replace perio=3 if (calendar>=1994 & calendar<=1998)
replace perio=4 if (calendar>=1999 & calendar<=2003)
replace perio=5 if (calendar>=2004 & calendar<=2008)
replace perio=6 if (calendar>=2009 & calendar<=2011)
*Taux brut de natalité par période*
stptime, by (perio) per(1000)
drop _*

*****
* TAUX GLOBAL DE FECONDITE *
*****

use "D:\user\Desktop\stage IRD\TABLES\resid_fec.dta", clear

```

```

stset EventDate if residence==1 & Sex==2 & agegroup>10 & agegroup<50,
id(IndividualId) failure(delivery==1) ///
    origin(time DoB) time0(datebeg) exit(time .) scale(31557600000)
stptime, by(calendar) per(1000)
cap drop perio
gen perio=0
replace perio=1 if (calendar>=1984 & calendar<=1988)
replace perio=2 if (calendar>=1989 & calendar<=1993)
replace perio=3 if (calendar>=1994 & calendar<=1998)
replace perio=4 if (calendar>=1999 & calendar<=2003)
replace perio=5 if (calendar>=2004 & calendar<=2008)
replace perio=6 if (calendar>=2009 & calendar<=2011)
stptime, by (perio) per(1000)
drop _*

*****
* TAUX DE FECONDITE _FEMMES_ *
*****

use "D:\user\Desktop\stage IRD\TABLES\resid_fec.dta", clear
stset EventDate if residence==1 & Sex==2, id(IndividualId) failure(delivery==1) ///
    origin(time DoB) time0(datebeg) exit(time .) scale(31557600000)
*courbe de survie*
sts graph, gwood tmax(100)
sts graph, hazard ci kernel(rectangle) width(1) ///
    xlab(15(5)50) ylab(,angle(horizontal))
sts graph, hazard ci kernel(rectangle) width(1) ///
    tmin(15) tmax(50) xlab(15(5)50) ylab(,angle(horizontal))
* moyenne d'âge à la maternité, (sur les femmes qui ont accouchées)
mean _t if _d==1
*28,4 ans*
* durée moyenne pour espérances de vie sans enfant des femmes qui ont des enfants
stci, rmean ccorr noshow

*****
*Taux de fécondité par année et groupe d'âge *
*****
stptime, by(calendar)at(15(5)50) per(1000) trim
*table de survie*
sts list, at(10(5)70) failure
* durée moyenne pour espérances de vie sans enfant des femmes qui ont des enfants
stci, rmean by(calendar_year)

*****
*ANALYSE FECONDITE PAR PERIODE*
*****
cap drop perio
gen perio=0
replace perio=1 if (calendar>=1984 & calendar<=1988)

```

```

replace perio=2 if (calendar>=1989 & calendar<=1993)
replace perio=3 if (calendar>=1994 & calendar<=1998)
replace perio=4 if (calendar>=1999 & calendar<=2003)
replace perio=5 if (calendar>=2004 & calendar<=2008)
replace perio=6 if (calendar>=2009 & calendar<=2011)
*****

sts graph, by(perio) tmax(80) hazard kernel(rectangle) ///
           ylog ylab(,angle(horizontal))
sts graph, by(perio) tmax(80)title("Evènement Naissance - Femmes") ///
           yline(.5) xline(15) ylog ylab(,angle(horizontal))
sts list, by (perio)at(10(5)60)
stptime, by (perio) at(15(5)50) per(1000) trim if perio!=0
stptime, by(perio) per(1000)

* durée moyenne espérances de vie sans enfant des femmes qui ont des enfants
stci, by(perio) rmean ccorr noshow

*****

```

ANNEXE A.7 : Programme pour la création de la variable vi_co_cu « création vi_vo_cu.do »

```

*****
*   CREATION DE LA VARIABLE VI-CO-CU   *
*   (identifiant des ménages)         *
*****

cd "D:\user\Desktop\stage IRD\TABLES"

use "D:\user\Desktop\stage IRD\ANALYSE_NIAKHAR\TABLES\resid_niakhar_11", clear
keep conc cuis vil
/*1 à 31 villages; */
drop if conc==.
drop if cuis==.
drop if vil==.
save info_menage.dta, replace

use "D:\user\Desktop\stage IRD\Tables_Bases_Niakhar\correspondance cuisis$.dta",
clear
rename vill_cp_village vil
rename cuis_cp_cuisine cuis
rename conc_cp_concession conc
rename cuis_id_cuis_enquete Ecuis
rename conc_id_conc_enquete Econc
drop if conc==.
drop if cuis==.
drop if vil==.
drop if Ecuis==.
drop if Econc==.

```



```
save vid_menage.dta, replace

gen vill=vil
tostring vill, replace
replace vill="0"+vill if vil<10
sort vil
gen concl=Econc
gen cuis1=Ecuis
tostring concl, replace
tostring cuis1, replace
replace cuis1="0"+cuis1 if Ecuis<10
gen vi_co_cu=vill+concl+cuis1
sort vi_co_cu

save vid_menage, replace

use "D:\user\Desktop\stage IRD\Tables_Bases_Niakhar\INDICATEURS.dta", clear
/*3133 obs*/
save indicateurs, replace

use indicateurs, clear
cap drop _merge
merge 1:m vi_co_cu using vid_menage
sort vi_co_cu

save Idmenage.dta, replace

br if _merge==3

use "D:\user\Desktop\stage IRD\TABLES\Idmenage.dta", clear
keep if _merge==3
cap drop _merge
rename vil LocationId
merge 1:m cuis conc LocationId using resid_fec.dta
sort vi_co_cu
order IndividualId LocationId DoB Sex datebeg EventDate EventCode vi_co_cu nomvil
conc cuis Econc Ecuis vill concl cuis1 isnv isnv2 israp israp2
sort IndividualId EventDate
save menage_resid_fec.dta, replace

tab _merge,m

keep if calendar==2003
keep if _merge==3
```

**ANNEXE A.8 : Programme pour l'analyse des déterminants de la fécondité
« analyse_fécondité.do »**

```

clear
set mem 1000000
cd "D:\user\Desktop\stage IRD\TABLES"

*****
* ANALYSE DES DETERMINANTS DE LA FECONDITE *
*****

/* analyse par religion, niveau instruction, niveau de vie et niveau agro-
pastorale */

use "D:\user\Desktop\stage IRD\TABLES\resid_fec.dta", clear

* FECONDITE ET RELIGION      *
*****

/*
relg (religion) (relgm=religion de la mère)
1 Musulmane
2 Chrétienne
3 Animiste
9 NSP
0 Données manquantes
*/

**Taux de fécondité par religion et groupe d'âge**
stset EventDate if residence==1 & Sex==2, id(IndividualId) failure(delivery==1) ///
      origin(time DoB) time0(datebeg) exit(time .) scale(31557600000)
stptime, by(relg)at(15(5)50) per(1000)

* FECONDITE DES FEMMES MUSULMANES *
stset EventDate if residence==1 & Sex==2 & relg==1, id(IndividualId)
failure(delivery==1) ///
      origin(time DoB) time0(datebeg) exit(time .) scale(31557600000)
*courbe de survie*
sts graph, hazard ci kernel(rectangle) width(1) ///
      tmin(15) tmax(50) xlab(15(5)50) ylab(,angle(horizontal))
* moyenne d'âge à la maternité, (sur les femmes qui ont accouchées)
mean _t if _d==1
*28,6 ans*
* durée moyenne pour espérances de vie sans enfant des femmes qui ont des enfants
stci, rmean ccorr noshow
*21,5 ans*
stptime, by(calendar)at(15(5)50) per(1000) trim
stptime, by(perio)at(15(5)50) per(1000) trim

* FECONDITE DES FEMMES CHRETIENNES *
stset EventDate if residence==1 & Sex==2 & relg==2, id(IndividualId)
failure(delivery==1) ///
      origin(time DoB) time0(datebeg) exit(time .) scale(31557600000)
*courbe de survie*

```

```

sts graph, gwood tmax(100)
sts graph, hazard ci kernel(rectangle) width(1) ///
            tmin(15) tmax(50) xlabel(15(5)50) ylabel(,angle(horizontal))
* moyenne d'âge à la maternité, (sur les femmes qui ont accouchées)
mean _t if _d==1
*28,71 ans*
* durée moyenne pour espérances de vie sans enfant des femmes qui ont des enfants
stci, rmean ccorr noshow
*21,6 ans*
stptime, by(calendar)at(15(5)50) per(1000) trim
stptime, by(perio)at(15(5)50) per(1000) trim

* FECONDITE DES FEMMES ANIMISTES *
stset EventDate if residence==1 & Sex==2 & relg==3, id(IndividualId)
failure(delivery==1) ///
            origin(time DoB) time0(datebeg) exit(time .) scale(31557600000)
*courbe de survie*
sts graph, gwood tmax(100)
sts graph, hazard ci kernel(rectangle) width(1) ///
            tmin(15) tmax(50) xlabel(15(5)50) ylabel(,angle(horizontal))
* moyenne d'âge à la maternité, (sur les femmes qui ont accouchées)
mean _t if _d==1
*29 ans*
* durée moyenne pour espérances de vie sans enfant des femmes qui ont des enfants
stci, rmean ccorr noshow
*21ans*
stptime, by(calendar)at(15(5)50) per(1000) trim
stptime, by(perio)at(15(5)50) per(1000) trim

*****
*TEST de significativité : le test du log-rank *
*****
/* pour tester les différences de séjour de plusieurs sous-populations)
   et ainsi voir l'allure des courbes
   avant => vérification de l'hypothèse de proportionnalité */
**TEST significativité va religion (M et C) sur fécondité**
stset EventDate if residence==1 & Sex==2, id(IndividualId) failure(delivery==1) ///
            origin(time DoB) time0(datebeg) exit(time .) scale(31557600000)
keep if relg==1 | relg==2
sts graph, by(relg) tmax(80)title("Evènement Naissance - Femmes")
sts graph, by (relg) hazard kernel(rectangle) ///
            tmin(15) tmax(50) xlabel(15(5)50) ylabel(,angle(horizontal))
sts test relg,logrank
/*la variable religion est non significative (Pr>Chi2 = 0.3805)*/

*****

```

```

* FECONDITE ET INSTRUCTION      *
*****
/*
instg2 (niveau d'instruction)
1 jms sco
2 Primaire Incomplet
3 Primaire Complet
4 Secondaire et +
5 Autre (coranique, arable, alphabétisation
9 NSP
0 Données manquantes
*/

/*
*STAT*
keep if Sex==2
keep if lastrecord==1
tab instg2 lastrecord,m
keep if perio==4
tab instg2 lastrecord,m
*/

use "D:\user\Desktop\stage IRD\TABLES\resid_fec.dta", clear
* FECONDITE DES FEMMES avec NIVEAU SUP ETUDE      *
*calcul de la durée des séjours avant l'évènement naissance (fécondité) pour les
femmes résidentes
stset EventDate if residence==1 & Sex==2 & instg2==4, id(IndividualId)
failure(delivery==1) ///
        origin(time DoB) time0(datebeg) exit(time .) scale(31557600000)
*courbe de survie*
sts graph, gwood tmax(100)
sts graph, hazard ci kernel(rectangle) width(1) ///
        tmin(15) tmax(50) xlabel(15(5)50) ylabel(,angle(horizontal))
* moyenne d'âge à la maternité, (sur les femmes qui ont accouchées)
mean _t if _d==1
*27,53 ans*
* durée moyenne pour espérances de vie sans enfant des femmes qui ont des enfants
stci, rmean ccorr noshow
*25,69 ans*
stptime, by(calendar)at(15(5)50) per(1000) trim

* FECONDITE DES FEMMES avec niveau PRIMAIRE COMPLET      *
*calcul de la durée des séjours avant l'évènement naissance (fécondité) pour les
femmes résidentes
stset EventDate if residence==1 & Sex==2 & instg2==3, id(IndividualId)
failure(delivery==1) ///
        origin(time DoB) time0(datebeg) exit(time .) scale(31557600000)
*courbe de survie*
sts graph, gwood tmax(100)
sts graph, hazard ci kernel(rectangle) width(1) ///
        tmin(15) tmax(50) xlabel(15(5)50) ylabel(,angle(horizontal))

```

```

* moyenne d'âge à la maternité, (sur les femmes qui ont accouchées)
mean _t if _d==1
*27,81 ans*
* durée moyenne pour espérances de vie sans enfant des femmes qui ont des enfants
stci, rmean ccorr noshw
*22,32 ans*
stptime, by(calendar)at(15(5)50) per(1000) trim

* FECONDITE DES FEMMES avec Niveau Primaire Incomplet *
*calcul de la durée des séjours avant l'évènement naissance (fécondité) pour les
femmes résidentes
stset EventDate if residence==1 & Sex==2 & instg2==2, id(IndividualId)
failure(delivery==1) ///
        origin(time DoB) time0(datebeg) exit(time .) scale(31557600000)
*courbe de survie*
sts graph, gwood tmax(100)
sts graph, hazard ci kernel(rectangle) width(1) ///
        tmin(15) tmax(50) xlab(15(5)50) ylab(,angle(horizontal))
* moyenne d'âge à la maternité, (sur les femmes qui ont accouchées)
mean _t if _d==1
*27,38ans*
* durée moyenne pour espérances de vie sans enfant des femmes qui ont des enfants
stci, rmean ccorr noshw
*21,55 ans*
stptime, by(calendar)at(15(5)50) per(1000) trim

* FECONDITE DES FEMMES JMS SCOLARISEES *
*calcul de la durée des séjours avant l'évènement naissance (fécondité) pour les
femmes résidentes
stset EventDate if residence==1 & Sex==2 & instg2==1, id(IndividualId)
failure(delivery==1) ///
        origin(time DoB) time0(datebeg) exit(time .) scale(31557600000)
*courbe de survie*
sts graph, gwood tmax(100)
sts graph, hazard ci kernel(rectangle) width(1) ///
        tmin(15) tmax(50) xlab(15(5)50) ylab(,angle(horizontal))
* moyenne d'âge à la maternité, (sur les femmes qui ont accouchées)
mean _t if _d==1
*28,89 ans*
* durée moyenne pour espérances de vie sans enfant des femmes qui ont des enfants
stci, rmean ccorr noshw
*21,17 ans*
stptime, by(calendar)at(15(5)50) per(1000) trim

* FECONDITE DES FEMMES avec niveau AUTRE (coranique, ...) *
*calcul de la durée des séjours avant l'évènement naissance (fécondité) pour les
femmes résidentes
stset EventDate if residence==1 & Sex==2 & instg2==5, id(IndividualId)
failure(delivery==1) ///
        origin(time DoB) time0(datebeg) exit(time .) scale(31557600000)
*courbe de survie*

```

```

sts graph, gwood tmax(100)
sts graph, hazard ci kernel(rectangle) width(1) ///
                tmin(15) tmax(50) xlab(15(5)50) ylab(,angle(horizontal))
* moyenne d'âge à la maternité, (sur les femmes qui ont accouchées)
mean _t if _d==1
*27,38ans*
* durée moyenne pour espérances de vie sans enfant des femmes qui ont des enfants
stci, rmean ccorr noshow
*21,55 ans*
stptime, by(calendar)at(15(5)50) per(1000) trim

*****
* TEST DE SIGNIFICATIVITE : log-rank *
*****
drop if instg2==0
drop if instg2==9
stset EventDate if residence==1 & Sex==2, id(IndividualId) failure(delivery==1) ///
                origin(time DoB) time0(datebeg) exit(time .) scale(31557600000)
**Taux de fécondité par niveau d'instruction et groupe d'âge**
stptime, by(instg2)at(15(5)50) per(1000)
sts graph, by(instg2) tmax(80)title("Evènement Naissance - Femmes")
sts graph, by (instg2) hazard kernel(rectangle) ///
                tmin(15) tmax(50) xlab(15(5)50) ylab(,angle(horizontal))
sts test instg2,logrank
*=> effet significatif du niveau d'instruction sur le risque de connaître des
naissances et sa durée

keep if instg2==1 | instg2==4
sts graph, by(instg2) tmax(80)title("Evènement Naissance - Femmes")
sts test instg2,logrank

*****
* ANALYSE FECONDITE ET NIVEAU DE VIE *
*****
use "D:\user\Desktop\stage IRD\TABLES\menage_resid_fec.dta", clear
tab _merge,m
* sur la période 1999-2003 *
save table_analyse-perio_fec.dta, replace
keep if _merge==3
cap drop perio
gen perio=0
replace perio=1 if (calendar>=1984 & calendar<=1988)
replace perio=2 if (calendar>=1989 & calendar<=1993)
replace perio=3 if (calendar>=1994 & calendar<=1998)
replace perio=4 if (calendar>=1999 & calendar<=2003)
replace perio=5 if (calendar>=2004 & calendar<=2008)
replace perio=6 if (calendar>=2009 & calendar<=2011)

```

```

keep if perio==4

gen byte delivery=EventCode==8
stset EventDate if residence==1 & Sex==2, id(IndividualId) failure(delivery==1) ///
      origin(time DoB) time0(datebeg) exit(time .) scale(31557600000)

*Regroupement en 2 classes selon le niveau de vie "pauvres" et "non pauvres" *
cap drop G_isnv2
gen G_isnv2=.
replace G_isnv2=0 if isnv2>=0 & isnv2<0.2809
replace G_isnv2=1 if isnv2>=0.2809 & isnv2<=1
tab G_isnv2,m
sort IndividualId EventDate EventCode
cap drop lastrecord
qui by IndividualId: gen lastrecord=_n==_N
tab G_isnv2 lastrecord
*COURBES*
*courbe de survie*
sts graph, by(G_isnv2) tmax(80)title("Evènement Naissance - Femmes")
sts graph, by (G_isnv2) hazard kernel(rectangle) ///
      tmin(15) tmax(50) xlab(15(5)50) ylab(,angle(horizontal))
*TEST LOGRANK effet du niveau de vie sur les naissances*
sts test G_isnv2, logrank
*effet significatif au seuil de 1% (0,002)*
keep if agegroup>=15 & agegroup<=45
stptime, by (perio) per(1000)
*taux global de fécondité pour la période 1999-2003 en fonction du niveau de vie*
stptime, by (G_isnv2) per(1000)

*Regroupement en 3 classes selon le niveau de vie "très pauvre" "pauvres" et "non
pauvres" *
cap drop G3_isnv2
gen G3_isnv2=.
replace G3_isnv2=1 if isnv2>=0 & isnv2<=0.148
replace G3_isnv2=2 if isnv2>0.148 & isnv2<=0.424
replace G3_isnv2=3 if isnv2>0.424 & isnv2<=1
tab G3_isnv2,m
tab G3_isnv2
sort IndividualId EventDate EventCode
/*stat
cap drop lastrecord
qui by IndividualId: gen lastrecord=_n==_N
tab G3_isnv2 lastrecord
keep if Sex==2 & lastrecord==1
tab G3_isnv2*/
*COURBES*
*courbe de survie*
sts graph, by(G3_isnv2) tmax(80)title("Evènement Naissance - Femmes")

```

```

sts graph, by (G3_isnv2) hazard kernel(rectangle) ///
                tmin(15) tmax(50) xlab(15(5)50) ylab(,angle(horizontal))
*TEST LOGRANK effet du niveau de vie sur les naissances*
sts test G3_isnv2, logrank
*effet significatif au seuil de 1% (0.0044)*
keep if agegroup>=15 & agegroup<=45
stptime, by (perio) per(1000)
*taux globaux de fécondité pour la période 1999-2003 en fonction du niveau de vie*
stptime, by (G3_isnv2) per(1000)

*Regroupement en 5 classes selon le niveau de vie *
cap drop G5_isnv2
gen G5_isnv2=.
replace G5_isnv2=1 if isnv2>=0 & isnv2<=0.058
replace G5_isnv2=2 if isnv2>0.058 & isnv2<=0.208
replace G5_isnv2=3 if isnv2>0.208 & isnv2<=0.424
replace G5_isnv2=4 if isnv2>0.424 & isnv2<=0.668
replace G5_isnv2=5 if isnv2>0.668 & isnv2<=1
tab G5_isnv2,m
tab G5_isnv2
sort IndividualId EventDate EventCode
/*stat
cap drop lastrecord
qui by IndividualId: gen lastrecord=_n==_N
tab G3_isnv2 lastrecord
keep if Sex==2 & lastrecord==1
tab G3_isnv2*/

*COURBES*
*courbe de survie*
sts graph, by(G5_isnv2) tmax(80)title("Evènement Naissance - Femmes")
sts graph, by (G5_isnv2) hazard kernel(rectangle) ///
                tmin(15) tmax(50) xlab(15(5)50) ylab(,angle(horizontal))
*TEST LOGRANK effet du niveau de vie sur les naissances*
sts test G5_isnv2, logrank
*effet significatif au seuil de 1% (0.0004)*
keep if agegroup>=15 & agegroup<=45
stptime, by (perio) per(1000)
*taux globaux de fécondité pour la période 1999-2003 en fonction du niveau de vie*
stptime, by (G5_isnv2) per(1000)

*****
* ANALYSE ANNEE 2003 *
use "D:\user\Desktop\stage IRD\TABLES\menage_resid_fec.dta", clear
tab _merge,m
keep if calendar==2003
keep if _merge==3
graph bar isnv2, by(Sex)
graph bar isnv2 if Sex==2, by(agegroup)

```



```

save table_analyse_fec.dta, replace
gen byte delivery=EventCode==8
stset EventDate if residence==1 & Sex==2, id(IndividualId) failure(delivery==1) ///
      origin(time DoB) time0(datebeg) exit(time .) scale(31557600000)

*Regroupement en 2 classes selon le niveau de vie "pauvres" et "non pauvres" *
cap drop G_isnv2
gen G_isnv2=.
replace G_isnv2=0 if isnv2>=0 & isnv2<0.2809
replace G_isnv2=1 if isnv2>=0.2809 & isnv2<=1
tab G_isnv2,m
sort IndividualId EventDate EventCode
cap drop lastrecord
qui by IndividualId: gen lastrecord=_n==_N
tab G_isnv2 lastrecord
*COURBES*
sts graph, by (G_isnv2) hazard kernel(rectangle) ///
      tmin(15) tmax(50) xlab(15(5)50) ylab(,angle(horizontal))
*TEST LOGRANK effet du niveau de vie sur les naissances*
sts test G_isnv2, logrank
*effet significatif au seuil de 10%*

*****
* ANALYSE FECONDITE ET NIVEAU DE RICHESSE AGRO-PASTORALE *
*****

*Regroupement en 3 classes selon le niveau de richesse agro-pastorale "très pauvre"
"pauvres" et "non pauvres" *
cap drop G3_israp2
gen G3_israp2=.
replace G3_israp2=1 if israp2>=0 & israp2<=0.352
replace G3_israp2=2 if israp2>0.352 & israp2<=0.658
replace G3_israp2=3 if israp2>0.658 & israp2<=1
tab G3_israp2,m
tab G3_israp2
sort IndividualId EventDate EventCode
/*stat
cap drop lastrecord
qui by IndividualId: gen lastrecord=_n==_N
tab G3_israp2 lastrecord
keep if Sex==2 & lastrecord==1
tab G3_israp2*/
*COURBES*
*courbe de survie*
sts graph, by(G3_israp2) tmax(80)title("Evènement Naissance - Femmes")
sts graph, by (G3_israp2) hazard kernel(rectangle) ///
      tmin(15) tmax(50) xlab(15(5)50) ylab(,angle(horizontal))
*TEST LOGRANK effet du niveau de richesse agro-pastorale sur les naissances*
sts test G3_israp2, logrank
*effet significatif au seuil de 1% (0.0052)*

```

```

keep if agegroup>=15 & agegroup<=45
stptime, by (perio) per(1000)
*taux globaux de fécondité pour la période 1999-2003 en fonction du niveau de
richesse agro-pastorale*
stptime, by (G3_israp2) per(1000)

*****
*test selon le sexe de l'enfant*
*****
*use "D:\user\Desktop\stage IRD\TABLES\resid_fec.dta", clear
use "D:\user\Desktop\stage IRD\TABLES\menage_resid_fec.dta", clear
gen byte delivery=EventCode==8
tab ChildSex lastrecord,m
*50,67% (81628) de naissance garçon 49,33% (79464) de naissance de fille*
stset EventDate if residence==1 & Sex==2, id(IndividualId) failure(delivery==1) ///
        origin(time DoB) time0(datebeg) exit(time .) scale(31557600000)
*courbe de survie*
sts graph, by (ChildSex) hazard kernel(rectangle) width(1) ///
        tmin(15) tmax(50) xlabel(15(5)50) ylabel(,angle(horizontal))
sts test ChildSex
*=> pas de différence significative (0.96) entre la durée pour avoir un garçon ou
une fille
*****

*****
**      MODELE DE COX          **
*****
use "D:\user\Desktop\stage IRD\TABLES\menage_resid_fec.dta", clear
tab _merge,m
* sur la période 1999-2003 *
keep if _merge==3
cap drop perio
gen perio=0
replace perio=1 if (calendar>=1984 & calendar<=1988)
replace perio=2 if (calendar>=1989 & calendar<=1993)
replace perio=3 if (calendar>=1994 & calendar<=1998)
replace perio=4 if (calendar>=1999 & calendar<=2003)
replace perio=5 if (calendar>=2004 & calendar<=2008)
replace perio=6 if (calendar>=2009 & calendar<=2011)
keep if perio==4
gen byte delivery=EventCode==8
stset EventDate if residence==1 & Sex==2, id(IndividualId) failure(delivery==1) ///
        origin(time DoB) time0(datebeg) exit(time .) scale(31557600000)

**variables niveau instruction, niveau de vie et niveau agro-pastorale**
sort IndividualId EventDate EventCode
cap drop lastrecord
qui by IndividualId: gen lastrecord=_n==_N
tab instg2,m
drop if instg2==9 | instg2==0

```

```
*creation des va indicatrices*
*instruction*
/*
instg2 (niveau d'instruction)
1 jms sco
2 Primaire Incomplet
3 Primaire Complet
4 Secondaire et +
5 Autre (coranique, arable, alphabétisation
9 NSP
0 Données manquantes
*/
*cap drop edu0
*gen edu0=1 if instg2==1
*replace edu0=0 if instg2!=1
cap drop edu1
gen edu1=1 if instg2==2
replace edu1=0 if instg2!=2
cap drop edu2
gen edu2=1 if instg2==3
replace edu2=0 if instg2!=3
cap drop edu3
gen edu3=1 if instg2==4
replace edu3=0 if instg2!=4
cap drop eduA
gen eduA=1 if instg2==5
replace eduA=0 if instg2!=5

*va isnv2 *
*Regroupement en 3 classes selon le niveau de vie "très pauvre" "pauvres" et "non
pauvres" *
cap drop G3_isnv2
gen G3_isnv2=.
replace G3_isnv2=1 if isnv2>=0 & isnv2<=0.148
replace G3_isnv2=2 if isnv2>0.148 & isnv2<=0.424
replace G3_isnv2=3 if isnv2>0.424 & isnv2<=1
tab G3_isnv2,m

sort IndividualId EventDate EventCode
cap drop nv1
gen nv1=1 if G3_isnv2==2
replace nv1=0 if G3_isnv2!=2
cap drop nv2
gen nv2=1 if G3_isnv2==3
replace nv2=0 if G3_isnv2!=3

*va israp2 *
*Regroupement en 3 classes selon le niveau de vie "très pauvre" "pauvres" et "non
pauvres" *
cap drop G3_israp2
```

```
gen G3_israp2=.
replace G3_israp2=1 if israp2>=0 & israp2<=0.352
replace G3_israp2=2 if israp2>0.352 & israp2<=0.658
replace G3_israp2=3 if israp2>0.658 & israp2<=1
tab G3_israp2,m
sort IndividualId EventDate EventCode

cap drop rap1
gen rap1=1 if G3_israp2==2
replace rap1=0 if G3_israp2!=2
cap drop rap2
gen rap2=1 if G3_israp2==3
replace rap2=0 if G3_israp2!=3

*modele de cox*
*modele de ref : edu0 nv0 rap0
stcox edu1 edu2 edu3 eduA nv1 nv2 rap1 rap2
*=> influence significative (positive) des variables instruction, niveau de vie et
richesse agro-pastorale
* sur le risque d'avoir des enfants.

*modele de cox avec interactions de variables (edu et nv)*
*variables d'interactions*
gen edu1nv1=edu1*nv1
gen edu1nv2=edu1*nv2
gen edu2nv1=edu2*nv1
gen edu2nv2=edu2*nv2
gen edu3nv2=edu3*nv2
gen edu3nv1=edu3*nv1
*modele*
stcox edu1 edu2 edu3 eduA nv1 nv2 edu1nv1 edu1nv2 edu2nv1 edu2nv2 edu3nv1 edu3nv2
```

ANNEXES B : LES FEUILLES DE CALCUL SOUS EXCEL (Feuilles-outilles)

ANNEXE B.1 : Matrice des incohérences

Aperçu de la feuille Excel : Matrice_incohérences du fichier Matrice_Incohérences.xlsx

Matrice des incohérences									
Event occurred	Following event							Total	
	BTH	IMG	OMG	EXT	ENT	DTH	OBE		
ENU	0	1	8955	4338	0	4345	5133	0	22772
BTH	0	0	6824	5416	4	5616	20241	2	38103
IMG	0	6	12449	3985	56	1235	9751	5	27487
OMG	1	10368	4	1	51	1	9	2171	32145
EXT	1	7	7	14	17390	1	150	139	17709
ENT	1	4	3904	3953	116	1063	8601	0	17642
DTH	0	8	0	0	3	0	2	12248	12261
OBE	0	2	0	0	0	0	0	43885	43887
Total	3	10396	32143	17707	17620	12261	43887	77989	212006
données avec incohérences									
INCOHERENCES									
SOMME	457								
%	0,21555994								
0,22 %									
si le % d'erreur est < 2 % ok pour analyse sinon problème dans la qualité des données									

ANNEXE B.2 : Feuilles-outils pour le calcul des indicateurs de la fécondité

Aperçu de la feuille Excel : sorties_tx_fécondité du fichier outils_analyse_FEC.xlsx

Sorties STATA (stptime)							!!! Remplacer les points par des virgules !!!		
Analysée la fécondité de 1983 à 2011									
taux de fécondité par année et groupe d'âge 15-49 ans									
Année	Groupe d'âge	Personne _année	Naissance	Taux de fécondité	Intervalle de confiance à 95%		TAUX de fécondité en ‰		
Sorties Stata à coller ci-dessous							(copier-coller de la colonne du taux de fécondité dans les cellules ci-dessous)		
1983							1983		
(15	- 20]	660,83054	114	172,51019	143,5795	207,2703	15-19	172,51	
(20	- 25]	621,47605	198	318,59635	277,1715	366,2124	20-24	318,60	
(25	- 30]	549,09224	187	340,56209	295,0871	393,045	25-29	340,56	
(30	- 35]	373,33181	117	313,39413	261,4552	375,6509	30-34	313,39	
(35	- 40]	314,72234	69	219,24087	173,1605	277,5838	35-39	219,24	
(40	- 45]	348,12319	43	123,51949	91,60696	166,5492	40-44	123,52	
(45	- 50]	325,10913	8	24,607122	12,30596	49,20463	45-49	24,61	
1984							1984		
(15	- 20]	1001,0473	194	193,79704	168,3583	223,0796	15-19	193,80	
(20	- 25]	955,82736	293	306,54071	273,376	343,7288	20-24	306,54	
(25	- 30]	894,63008	289	323,03855	287,8616	362,5142	25-29	323,04	
(30	- 35]	606,54291	183	301,7099	261,0159	348,7484	30-34	301,71	
(35	- 40]	510,88763	144	281,86237	239,3887	331,872	35-39	281,86	
(40	- 45]	512,90471	67	130,62855	102,8128	165,9697	40-44	130,63	
(45	- 50]	498,33815	11	22,073365	12,22423	39,858	45-49	22,07	
1985							1985		
(15	- 20]	916,70846	142	154,90203	131,4092	182,5948	15-19	154,90	
(20	- 25]	939,667	263	279,88639	248,0244	315,8414	20-24	279,89	
(25	- 30]	927,09371	292	314,96277	280,8319	353,2417	25-29	314,96	
(30	- 35]	618,23573	170	274,97602	236,5978	319,5796	30-34	274,98	
(35	- 40]	510,04331	113	221,54981	184,2456	266,407	35-39	221,55	
(40	- 45]	501,60222	64	127,59114	99,86662	163,0124	40-44	127,59	
(45	- 50]	507,34762	13	25,623457	14,87842	44,12844	45-49	25,62	
1986							1986		
(15	- 20]	878,93756	147	167,24738	142,2831	196,5917	15-19	167,25	
(20	- 25]	929,65374	310	333,45749	298,329	372,7224	20-24	333,46	
(25	- 30]	977,07145	341	349,00211	313,8578	388,0817	25-29	349,00	
(30	- 35]	651,91682	220	337,46637	295,694	385,1399	30-34	337,47	
(35	- 40]	527,34685	130	246,51707	207,583	292,7536	35-39	246,52	
(40	- 45]	490,5424	54	110,08223	84,31083	143,7312	40-44	110,08	
(45	- 50]	509,2577	20	39,272848	25,33716	60,8733	45-49	39,27	
1987							1987		
(15	- 20]	909,96972	152	167,03853	142,4869	195,8205	15-19	167,04	
(20	- 25]	936,27842	311	332,16615	297,2269	371,2125	20-24	332,17	
(25	- 30]	938,88101	294	313,13872	279,3146	351,0588	25-29	313,14	
(30	- 35]	716,39794	211	294,52904	257,3529	337,0756	30-34	294,53	
(35	- 40]	527,31056	136	257,91251	218,013	305,1141	35-39	257,91	
(40	- 45]	497,66535	76	152,71306	121,9654	191,2122	40-44	152,71	
(45	- 50]	498,34878	16	32,106028	19,66918	52,40672	45-49	32,11	

Aperçu de la feuille Excel : indicateurs_FECONDITE du fichier outils_analyse_FEC.xlsx

Indicateurs Fécondité_Niakhar_											
Année	ISF	Taux de fécondité %							Taux global de fécondité	Taux brut de natalité	Age Moyen à la maternité
		15-19	20-24	25-29	30-34	35-39	40-44	45-49			
1984	7,80	193,80	306,54	323,04	301,71	281,86	130,63	22,07	237,50	50,77	29,59
1985	7,00	154,90	279,89	314,96	274,98	221,55	127,59	25,62	214,71	45,39	29,69
1986	7,92	167,25	333,46	349,00	337,47	246,52	110,08	39,27	246,68	51,63	29,55
1987	7,75	167,04	332,17	313,14	294,53	257,91	152,71	32,11	238,36	49,02	29,86
1988	7,54	128,78	331,23	332,98	316,92	235,25	131,59	31,09	231,70	47,40	29,88
1989	7,97	167,67	319,35	358,02	297,98	251,02	172,60	26,61	242,84	48,81	29,92
1990	7,15	134,20	313,05	285,85	277,47	255,96	136,17	27,33	217,19	43,08	30,04
1991	7,41	154,79	327,29	308,19	325,95	242,51	100,18	23,80	230,94	44,79	29,42
1992	7,30	134,96	340,73	315,39	291,24	222,72	127,29	27,55	223,51	43,17	29,62
1993	7,67	131,67	332,89	343,08	309,02	252,65	142,83	21,82	232,81	44,76	29,89
1994	7,01	136,35	297,39	292,72	326,20	225,81	88,00	36,52	213,36	41,72	29,70
1995	6,73	131,08	262,28	288,96	304,68	209,05	122,40	27,93	202,63	39,93	30,02
1996	7,06	113,55	289,39	335,81	290,14	245,85	114,92	22,94	211,06	41,97	29,98
1997	6,82	126,65	296,13	295,72	294,40	223,30	111,21	16,89	205,96	40,97	29,67
1998	6,48	105,84	273,09	288,77	271,61	222,52	122,90	10,45	192,96	38,92	29,98
1999	6,63	105,01	258,16	314,12	302,85	222,33	116,18	8,04	195,30	40,08	29,99
2000	6,96	84,24	287,43	332,52	302,83	255,36	111,82	17,33	203,49	41,85	30,24
2001	6,49	100,51	277,87	283,78	279,41	233,16	103,05	20,27	191,12	39,71	30,03
2002	6,90	95,83	280,62	308,25	324,23	213,71	130,36	26,62	198,98	41,85	30,32
2003	6,52	86,64	278,79	284,65	285,86	242,78	94,37	31,08	188,99	39,58	30,29
2004	6,51	72,10	262,48	313,70	290,35	227,77	117,21	17,79	186,34	39,64	30,43
2005	6,54	73,98	253,09	290,64	310,22	258,57	98,80	22,30	185,23	39,82	30,60
2006	6,42	65,09	252,26	297,86	297,87	222,65	121,79	25,52	180,62	39,06	30,73
2007	6,27	66,24	251,27	269,78	280,86	243,85	125,18	16,21	178,84	38,54	30,79
2008	6,43	59,26	225,81	316,35	288,91	246,83	124,06	25,38	182,98	39,54	31,04
2009	6,32	69,18	227,12	303,14	297,38	229,54	110,14	26,64	183,98	39,62	30,78
2010	6,43	63,33	232,46	282,74	287,17	266,91	129,17	24,67	185,48	39,77	31,18
2011	5,69	58,00	183,17	274,72	262,35	233,51	103,93	22,11	165,26	35,21	31,15

Les formules :

ISF = (somme des taux de fécondité *5)/1000

AMM = (17,5*taux(15-19)+22,5*taux(20-24)+27,5*taux(25-29)+32,5*taux(30-34)+37,5*taux(35-39)+42,5*taux(40-44)+47,5*taux(45-49))/sommedestaux

Dans ce fichier Excel d'autres feuilles sont créées afin de permettre la représentation graphique de ces données (ISF, AMM, Taux de fécondité, etc.). Il en est de même avec les résultats obtenus par période et non par année.

ANNEXES C : LES SORTIES STATA

ANNEXE C.1 : Table de survie par groupe d'âge des femmes de ne pas avoir d'enfant sur l'ensemble de la période

```
. sts list, at(10(5)60)
```

```
failure _d: delivery == 1
analysis time _t: (EventDate-origin)/3.16e+10
origin: time DoB
exit on or before: time .
id: IndividualId
```

Time	Beg. Total	Fail	Survivor Function	Std. Error	[95% Conf. Int.]	
10	11984	6	0.9996	0.0002	0.9991	0.9998
15	8737	59	0.9932	0.0008	0.9914	0.9947
20	7491	4075	0.5862	0.0049	0.5766	0.5957
25	6283	9375	0.1484	0.0024	0.1436	0.1532
30	5145	8716	0.0321	0.0007	0.0306	0.0335
35	4246	6936	0.0073	0.0002	0.0069	0.0077
40	3603	4660	0.0022	0.0001	0.0021	0.0024
45	3361	2079	0.0012	0.0000	0.0011	0.0013
50	3062	388	0.0011	0.0000	0.0010	0.0012
55	2586	30	0.0011	0.0000	0.0010	0.0012
60	2275	1	0.0011	0.0000	0.0010	0.0012

Note: survivor function is calculated over full data and evaluated at indicated times; it is not calculated from aggregates shown at left.

ANNEXE C.2 : Table de survie par groupe d'âge des femmes d'avoir un enfant sur l'ensemble de la période

```
. sts list, at(10(5)70) failure
```

```
failure _d: delivery == 1
analysis time _t: (EventDate-origin)/3.16e+10
origin: time DoB
exit on or before: time .
id: IndividualId
```

Time	Beg. Total	Fail	Failure Function	Std. Error	[95% Conf. Int.]	
10	11984	6	0.0004	0.0002	0.0002	0.0009
15	8737	59	0.0068	0.0008	0.0053	0.0086
20	7491	4075	0.4138	0.0049	0.4043	0.4234
25	6283	9375	0.8516	0.0024	0.8468	0.8564
30	5145	8716	0.9679	0.0007	0.9665	0.9694
35	4246	6936	0.9927	0.0002	0.9923	0.9931
40	3603	4660	0.9978	0.0001	0.9976	0.9979
45	3361	2079	0.9988	0.0000	0.9987	0.9989
50	3062	388	0.9989	0.0000	0.9988	0.9990
55	2586	30	0.9989	0.0000	0.9988	0.9990
60	2275	1	0.9989	0.0000	0.9988	0.9990
65	1932	0	0.9989	0.0000	0.9988	0.9990
70	1561	2	0.9989	0.0000	0.9988	0.9990

ANNEXE C.2 : Table de survie des femmes sans connaître l'évènement naissance, par période

```
. sts list, by (perio)at(10(5)60)
```

```
      failure _d: delivery == 1
analysis time _t: (EventDate-origin)/3.16e+10
      origin: time DoB
exit on or before: time .
      id: IndividualId
```

Time	Beg. Total	Fail	Survivor Function	Std. Error	[95% Conf. Int.]	
perio=1						
10	1657	0	1.0000	.	.	.
15	922	23	0.9770	0.0047	0.9656	0.9847
20	980	758	0.4354	0.0130	0.4098	0.4606
25	929	1477	0.0892	0.0045	0.0806	0.0984
30	850	1510	0.0174	0.0011	0.0152	0.0197
35	557	1031	0.0038	0.0003	0.0032	0.0044
40	494	650	0.0011	0.0001	0.0009	0.0013
45	512	324	0.0006	0.0001	0.0005	0.0007
50	486	76	0.0005	0.0001	0.0004	0.0006
55	414	3	0.0005	0.0001	0.0004	0.0006
60	364	0	0.0005	0.0001	0.0004	0.0006
perio=2						
10	1819	2	0.9991	0.0006	0.9966	0.9998
15	1192	11	0.9904	0.0027	0.9834	0.9945
20	826	705	0.4663	0.0133	0.4400	0.4922
25	849	1371	0.0910	0.0048	0.0819	0.1006
30	802	1314	0.0181	0.0012	0.0158	0.0207
35	770	1253	0.0040	0.0003	0.0034	0.0047
40	533	761	0.0012	0.0001	0.0010	0.0014
45	474	331	0.0006	0.0001	0.0005	0.0007
50	480	60	0.0005	0.0001	0.0004	0.0007
55	451	11	0.0005	0.0001	0.0004	0.0006
60	369	0	0.0005	0.0001	0.0004	0.0006
perio=3						
10	2169	2	0.9992	0.0006	0.9968	0.9998
15	1348	4	0.9965	0.0015	0.9921	0.9985
20	1216	760	0.5366	0.0121	0.5126	0.5599
25	861	1459	0.1288	0.0056	0.1180	0.1400
30	782	1203	0.0287	0.0018	0.0253	0.0323
35	758	1108	0.0064	0.0005	0.0055	0.0075
40	744	896	0.0021	0.0002	0.0018	0.0025
45	506	335	0.0012	0.0001	0.0010	0.0014
50	461	54	0.0011	0.0001	0.0009	0.0013
55	459	5	0.0011	0.0001	0.0009	0.0013
60	429	0	0.0011	0.0001	0.0009	0.0013
perio=4						
10	2249	0	1.0000	.	.	.
15	1715	8	0.9957	0.0015	0.9914	0.9979
20	1387	684	0.6056	0.0116	0.5825	0.6278
25	1161	1766	0.1504	0.0058	0.1393	0.1618
30	774	1459	0.0327	0.0018	0.0292	0.0363
35	722	1089	0.0073	0.0005	0.0063	0.0084
40	700	813	0.0023	0.0002	0.0019	0.0027
45	721	423	0.0013	0.0001	0.0011	0.0015
50	479	61	0.0012	0.0001	0.0010	0.0014
55	430	3	0.0012	0.0001	0.0010	0.0014
60	426	0	0.0012	0.0001	0.0010	0.0014
perio=5						
10	2256	0	1.0000	.	.	.
15	2126	3	0.9986	0.0008	0.9957	0.9995
20	1719	636	0.7004	0.0099	0.6805	0.7192
25	1400	1904	0.2000	0.0064	0.1876	0.2127
30	1113	1842	0.0451	0.0021	0.0411	0.0495
35	742	1361	0.0105	0.0006	0.0093	0.0118
40	701	853	0.0032	0.0002	0.0027	0.0037
45	686	396	0.0018	0.0001	0.0015	0.0021
50	696	80	0.0016	0.0001	0.0013	0.0018
55	454	3	0.0016	0.0001	0.0013	0.0018
60	399	1	0.0016	0.0001	0.0013	0.0018
perio=6						
10	1694	1	0.9995	0.0005	0.9964	0.9999
15	1310	9	0.9930	0.0022	0.9869	0.9962
20	1239	418	0.7203	0.0114	0.6972	0.7420
25	951	1201	0.2445	0.0086	0.2278	0.2615
30	747	1207	0.0581	0.0032	0.0521	0.0645
35	633	973	0.0142	0.0010	0.0124	0.0163
40	382	617	0.0043	0.0004	0.0036	0.0050
45	398	229	0.0024	0.0002	0.0020	0.0029
50	411	48	0.0021	0.0002	0.0018	0.0026
55	342	4	0.0021	0.0002	0.0017	0.0025
60	252	0	0.0021	0.0002	0.0017	0.0025

Note: survivor function is calculated over full data and evaluated at indicated times; it is not calculated from aggregates shown at left.

```
. end of do-file
```

```
.
```

ANNEXE C.3 : Personne-année, Nombre de naissance, Taux de fécondité (15-49 ans) par année seulement et par année et groupe d'âge quinquennal

➤ Par année

Calendar person-time	failures	rate	[95% Conf.Int]	
1983	3191,7669	735	230,27998	214,2196 247,5445
1984	4976,8068	1182	237,50169	224,3408 251,4347
1985	4918,2683	1056	214,70972	202,1426 228,0582
1986	4961,9857	1224	246,67544	233,2362 260,8891
1987	5021,8225	1197	238,35968	225,2319 252,2526
1988	5053,8628	1171	231,70395	218,8059 245,3624
1989	5015,7512	1218	242,83501	229,5733 256,8628
1990	5069,3445	1101	217,18784	204,7305 230,4032
1991	5001,3975	1155	230,93545	217,9939 244,6452
1992	5051,2742	1129	223,50796	210,8434 236,9332
1993	5094,2371	1186	232,81209	219,9322 246,4462
1994	5352,4718	1142	213,35937	201,3369 226,0997
1995	5517,5367	1118	202,62665	191,0906 214,8591
1996	5742,5646	1212	211,05553	199,5017 223,2785
1997	5816,6485	1198	205,96053	194,6218 217,9599
1998	5970,067	1152	192,96266	182,1355 204,4335
1999	6067,5633	1185	195,30081	184,4917 206,7432
2000	6216,531	1265	203,48969	192,5795 215,018
2001	6430,6593	1229	191,11571	180,724 202,1049
2002	6664,0015	1326	198,97955	188,5528 209,9829
2003	6830,9311	1291	188,99327	178,9601 199,589
2004	7169,7823	1336	186,3376	176,6089 196,6022
2005	7477,1781	1385	185,23031	175,7276 195,2469
2006	7706,7669	1392	180,62049	171,377 190,3626
2007	7990,5764	1429	178,83566	169,7997 188,3525
2008	8301,5898	1519	182,977	174,0029 192,4139
2009	8539,0362	1571	183,97861	175,1023 193,3049
2010	8793,3037	1631	185,48205	176,6953 194,7057
2011	9034,4404	1493	165,2565	157,083 173,8553
Total	178978,17	36228	202,41575	200,3421 204,5109

➤ Par année et groupe d'âge

Groupe d'âge	Personne	Naissance	Taux de	Intervalle de confiance
	_année		fécondité	à 95%
1983				
(15 - 20]	660,83054	114	172,51019	143,5795 207,2703
(20 - 25]	621,47605	198	318,59635	277,1715 366,2124
(25 - 30]	549,09224	187	340,56209	295,0871 393,045
(30 - 35]	373,33181	117	313,39413	261,4552 375,6509
(35 - 40]	314,72234	69	219,24087	173,1605 277,5838
(40 - 45]	348,12319	43	123,51949	91,60696 166,5492
(45 - 50]	325,10913	8	24,607122	12,30596 49,20463
1984				
(15 - 20]	1001,0473	194	193,79704	168,3583 223,0796

(20	-	25]	955,82736	293	306,54071	273,376	343,7288
(25	-	30]	894,63008	289	323,03855	287,8616	362,5142
(30	-	35]	606,54291	183	301,7099	261,0159	348,7484
(35	-	40]	510,88763	144	281,86237	239,3887	331,872
(40	-	45]	512,90471	67	130,62855	102,8128	165,9697
(45	-	50]	498,33815	11	22,073365	12,22423	39,858
1985							
(15	-	20]	916,70846	142	154,90203	131,4092	182,5948
(20	-	25]	939,667	263	279,88639	248,0244	315,8414
(25	-	30]	927,09371	292	314,96277	280,8319	353,2417
(30	-	35]	618,23573	170	274,97602	236,5978	319,5796
(35	-	40]	510,04331	113	221,54981	184,2456	266,407
(40	-	45]	501,60222	64	127,59114	99,86662	163,0124
(45	-	50]	507,34762	13	25,623457	14,87842	44,12844
1986							
(15	-	20]	878,93756	147	167,24738	142,2831	196,5917
(20	-	25]	929,65374	310	333,45749	298,329	372,7224
(25	-	30]	977,07145	341	349,00211	313,8578	388,0817
(30	-	35]	651,91682	220	337,46637	295,694	385,1399
(35	-	40]	527,34685	130	246,51707	207,583	292,7536
(40	-	45]	490,5424	54	110,08223	84,31083	143,7312
(45	-	50]	509,2577	20	39,272848	25,33716	60,8733
1987							
(15	-	20]	909,96972	152	167,03853	142,4869	195,8205
(20	-	25]	936,27842	311	332,16615	297,2269	371,2125
(25	-	30]	938,88101	294	313,13872	279,3146	351,0588
(30	-	35]	716,39794	211	294,52904	257,3529	337,0756
(35	-	40]	527,31059	136	257,91251	218,013	305,1141
(40	-	45]	497,66535	76	152,71306	121,9654	191,2122
(45	-	50]	498,34878	16	32,106028	19,66918	52,40672
1988							
(15	-	20]	955,0956	123	128,78292	107,9215	153,6769
(20	-	25]	905,71132	300	331,23137	295,7926	370,916
(25	-	30]	882,92391	294	332,98453	297,0168	373,3079
(30	-	35]	779,3769	247	316,91984	279,762	359,0129
(35	-	40]	539,85982	127	235,24625	197,6928	279,9333
(40	-	45]	478,77091	63	131,58694	102,7949	168,4434
(45	-	50]	514,57659	16	31,093525	19,04889	50,75401
1989							
(15	-	20]	948,30073	159	167,66833	143,5312	195,8645
(20	-	25]	879,92095	281	319,34687	284,1085	358,9559
(25	-	30]	849,10982	304	358,022	319,9558	400,6171
(30	-	35]	808,77379	241	297,98196	262,639	338,0809
(35	-	40]	561,71783	141	251,01571	212,8221	296,0636
(40	-	45]	480,86894	83	172,6042	139,1938	214,0341
(45	-	50]	488,51744	13	26,611128	15,45192	45,82939
1990							
(15	-	20]	983,5743	132	134,2044	113,1564	159,1675
(20	-	25]	856,09352	268	313,04991	277,727	352,8653
(25	-	30]	839,59329	240	285,85269	251,8822	324,4047
(30	-	35]	850,54535	236	277,46904	244,2339	315,2267
(35	-	40]	582,12637	149	255,95817	217,9896	300,5399
(40	-	45]	484,67694	66	136,17318	106,9833	173,3274
(45	-	50]	475,60389	13	27,333671	15,87147	47,07375
1991							
(15	-	20]	943,21111	146	154,79037	131,6128	182,0497

(20	-	25]	821,89593	269	327,29204	290,4269	368,8367
(25	-	30]	804,70107	248	308,18898	272,1233	349,0346
(30	-	35]	877,43331	286	325,9507	290,2815	366,0029
(35	-	40]	602,02493	146	242,51487	206,2018	285,2229
(40	-	45]	489,10795	49	100,18238	75,71661	132,5536
(45	-	50]	462,09314	11	23,804725	13,18306	42,98432
1992							
(15	-	20]	992,87776	134	134,96123	113,94	159,8607
(20	-	25]	815,89309	278	340,73092	302,9423	383,2333
(25	-	30]	805,36438	254	315,38519	278,8895	356,6567
(30	-	35]	830,92923	242	291,24021	256,7639	330,3457
(35	-	40]	651,0557	145	222,7152	189,2611	262,0827
(40	-	45]	487,06766	62	127,29238	99,24295	163,2695
(45	-	50]	471,83899	13	27,551771	15,99811	47,44936
1993							
(15	-	20]	1017,6847	134	131,67143	111,1627	155,9639
(20	-	25]	826,10482	275	332,88754	295,7796	374,651
(25	-	30]	781,16139	268	343,07891	304,3677	386,7136
(30	-	35]	802,54615	248	309,0165	272,854	349,9718
(35	-	40]	712,43433	180	252,65487	218,3146	292,3967
(40	-	45]	497,08995	71	142,83129	113,189	180,2364
(45	-	50]	458,24859	10	21,822217	11,74155	40,55762
1994							
(15	-	20]	1129,4234	154	136,35277	116,432	159,6819
(20	-	25]	884,37109	263	297,38647	263,5323	335,5896
(25	-	30]	806,21982	236	292,72414	257,6618	332,5577
(30	-	35]	778,66772	254	326,19819	288,4512	368,8847
(35	-	40]	766,13177	173	225,80972	194,5481	262,0947
(40	-	45]	522,73337	46	87,998973	65,91356	117,4845
(45	-	50]	465,52829	17	36,517652	22,70158	58,74211
1995							
(15	-	20]	1205,3686	158	131,08024	112,1552	153,1987
(20	-	25]	945,54158	248	262,28355	231,5899	297,0451
(25	-	30]	802,88691	232	288,95726	254,0677	328,638
(30	-	35]	748,31758	228	304,68347	267,5942	346,9134
(35	-	40]	803,63878	168	209,04915	179,7119	243,1756
(40	-	45]	547,3982	67	122,39719	96,33423	155,5114
(45	-	50]	465,52005	13	27,925758	16,21527	48,09343
1996							
(15	-	20]	1294,6271	147	113,54621	96,59768	133,4684
(20	-	25]	1029,7436	298	289,39243	258,3321	324,1873
(25	-	30]	792,10648	266	335,81344	297,7882	378,6941
(30	-	35]	730,6696	212	290,14482	253,6028	331,9522
(35	-	40]	841,99108	207	245,84583	214,536	281,725
(40	-	45]	574,3371	66	114,91509	90,28205	146,2691
(45	-	50]	479,47155	11	22,941924	12,70524	41,42636
1997							
(15	-	20]	1279,1245	162	126,64913	108,5739	147,7335
(20	-	25]	1114,3895	330	296,12625	265,8397	329,8633
(25	-	30]	787,90717	233	295,72012	260,0859	336,2366
(30	-	35]	737,10038	217	294,39681	257,721	336,2919
(35	-	40]	797,12568	178	223,3023	192,7937	258,6387
(40	-	45]	629,46713	70	111,20517	87,9806	140,5604
(45	-	50]	473,61156	8	16,89148	8,447389	33,77636
1998							
(15	-	20]	1313,337	139	105,83727	89,62737	124,9789

(20	-	25]	1171,7659	320	273,09209	244,7516	304,7142
(25	-	30]	817,26629	236	288,76757	254,1791	328,0627
(30	-	35]	725,30387	197	271,6103	236,2113	312,3143
(35	-	40]	763,98218	170	222,51828	191,4615	258,6127
(40	-	45]	699,73395	86	122,90385	99,48969	151,8284
(45	-	50]	478,31648	5	10,45333	4,350967	25,11444
1999							
(15	-	20]	1333,249	140	105,00664	88,97687	123,9243
(20	-	25]	1204,671	311	258,16178	231,0068	288,5089
(25	-	30]	840,43869	264	314,12166	278,4262	354,3935
(30	-	35]	726,42562	220	302,85276	265,3649	345,6364
(35	-	40]	724,15469	161	222,32819	190,5068	259,4649
(40	-	45]	740,20681	86	116,18375	94,04982	143,5267
(45	-	50]	497,52193	4	8,0398466	3,017498	21,42143
2000							
(15	-	20]	1353,2545	114	84,241363	70,11373	101,2157
(20	-	25]	1255,9375	361	287,43469	259,2621	318,6686
(25	-	30]	890,17812	296	332,51772	296,7151	372,6404
(30	-	35]	719,88384	218	302,82664	265,1816	345,8157
(35	-	40]	700,95995	179	255,36409	220,5657	295,6525
(40	-	45]	778,0584	87	111,8168	90,62514	137,9639
(45	-	50]	519,21417	9	17,333888	9,019078	33,31423
2001							
(15	-	20]	1432,7402	144	100,50671	85,36141	118,3392
(20	-	25]	1270,3806	353	277,86949	250,3433	308,4223
(25	-	30]	976,09642	277	283,78344	252,257	319,2499
(30	-	35]	715,78231	200	279,41456	243,254	320,9505
(35	-	40]	686,23837	160	233,15514	199,6878	272,2315
(40	-	45]	805,42545	83	103,05113	83,10387	127,7863
(45	-	50]	542,78046	11	20,266021	11,22333	36,59446
2002							
(15	-	20]	1554,8207	149	95,830984	81,61553	112,5224
(20	-	25]	1300,6899	365	280,6203	253,2592	310,9374
(25	-	30]	1034,8679	319	308,25191	276,2152	344,0043
(30	-	35]	724,80308	235	324,226	285,313	368,4462
(35	-	40]	687,84485	147	213,71099	181,8113	251,2076
(40	-	45]	759,42273	99	130,36218	107,0539	158,7452
(45	-	50]	601,0525	16	26,619971	16,30824	43,45182
2003							
(15	-	20]	1581,3206	137	86,63645	73,27867	102,4292
(20	-	25]	1348,6677	376	278,79365	251,9913	308,4468
(25	-	30]	1064,4571	303	284,65216	254,3397	318,5772
(30	-	35]	755,60702	216	285,86288	250,1732	326,644
(35	-	40]	683,74312	166	242,78124	208,5206	282,6709
(40	-	45]	720,58385	68	94,367921	74,40481	119,6872
(45	-	50]	675,66669	21	31,080413	20,26466	47,6688
2004							
(15	-	20]	1692,1171	122	72,09903	60,37613	86,0981
(20	-	25]	1409,6269	370	262,4808	237,0531	290,636
(25	-	30]	1134,8574	356	313,69579	282,7451	348,0345
(30	-	35]	799,03114	232	290,35164	255,2937	330,2238
(35	-	40]	702,4557	160	227,77237	195,0777	265,9466
(40	-	45]	699,59811	82	117,21015	94,39864	145,5341
(45	-	50]	730,85435	13	17,787402	10,32837	30,63327
2005							
(15	-	20]	1811,2443	134	73,982291	62,45902	87,63153

(20	-	25]	1446,1344	366	253,0885	228,4438 280,3919
(25	-	30]	1221,4286	355	290,64327	261,9288 322,5057
(30	-	35]	857,45622	266	310,21992	275,0928 349,8325
(35	-	40]	699,99501	181	258,57327	223,5189 299,1252
(40	-	45]	678,15685	67	98,797203	77,75957 125,5265
(45	-	50]	762,40672	17	22,29781	13,86167 35,86814
2006						
(15	-	20]	1904,9863	124	65,092331	54,58706 77,61935
(20	-	25]	1502,3892	379	252,26487	228,1043 278,9845
(25	-	30]	1225,4015	365	297,86156	268,8194 330,0414
(30	-	35]	933,29847	278	297,86827	264,8333 335,024
(35	-	40]	700,65775	156	222,64793	190,3128 260,477
(40	-	45]	656,87588	80	121,78861	97,82265 151,6261
(45	-	50]	783,81804	20	25,516126	16,46191 39,55025
2007						
(15	-	20]	1992,652	132	66,243378	55,85408 78,56517
(20	-	25]	1615,808	406	251,26748	227,9774 276,9368
(25	-	30]	1260,2675	340	269,784	242,5791 300,0399
(30	-	35]	1004,0627	282	280,85896	249,9194 315,6288
(35	-	40]	709,44203	173	243,85361	210,0939 283,0381
(40	-	45]	671,03968	84	125,17889	101,0782 155,0261
(45	-	50]	740,19499	12	16,211944	9,20692 28,54669
2008						
(15	-	20]	2092,5292	124	59,258433	49,69469 70,66271
(20	-	25]	1696,0837	383	225,81432	204,2947 249,6007
(25	-	30]	1346,6069	426	316,35067	287,6921 347,8641
(30	-	35]	1048,7534	303	288,91445	258,1481 323,3475
(35	-	40]	741,38967	183	246,83376	213,5413 285,3167
(40	-	45]	669,0043	83	124,06497	100,0501 153,844
(45	-	50]	709,34609	18	25,375484	15,98763 40,27583
2009						
(15	-	20]	2153,6802	149	69,183904	58,92125 81,23407
(20	-	25]	1783,2295	405	227,11602	206,0399 250,3481
(25	-	30]	1375,6216	417	303,13569	275,3934 333,6726
(30	-	35]	1102,9676	328	297,37955	266,8771 331,3682
(35	-	40]	775,46422	178	229,53993	198,1792 265,8634
(40	-	45]	671,88089	74	110,13857	87,69791 138,3215
(45	-	50]	675,59088	18	26,643344	16,78644 42,28817
2010						
(15	-	20]	2195,0187	139	63,325202	53,6264 74,77812
(20	-	25]	1884,1811	438	232,46174	211,6798 255,2839
(25	-	30]	1393,5196	394	282,73733	256,1535 312,0801
(30	-	35]	1170,0395	336	287,16979	258,0488 319,5771
(35	-	40]	839,22896	224	266,91167	234,1501 304,2572
(40	-	45]	665,78815	86	129,17022	104,5623 159,5695
(45	-	50]	648,62483	16	24,667573	15,11214 40,26492
2011						
(15	-	20]	2241,3842	130	57,999872	48,83956 68,87829
(20	-	25]	1954,4661	358	183,17023	165,1458 203,1619
(25	-	30]	1441,4502	396	274,72333	248,9551 303,1587
(30	-	35]	1177,8044	309	262,35257	234,6725 293,2976
(35	-	40]	920,71562	215	233,51401	204,2968 266,9097
(40	-	45]	663,93672	69	103,92557	82,08235 131,5816
(45	-	50]	633,17997	14	22,110617	13,09507 37,3331

ANNEXE C.4 : Personne-année, Nombre de naissance, Taux global de fécondité (15-49 ans) par période, période et groupe d'âge

➤ Par période

perio	person-time	failures	rate	[95% Conf.]	
1	24932,746	5830	233,82904	227,9032	239,909
2	25232,005	5789	229,43084	223,5962	235,4178
3	28399,289	5822	205,00513	199,8062	210,3393
4	32209,686	6296	195,46915	190,7	200,3576
5	38645,893	7061	182,71023	178,4979	187,022
6	26366,78	4695	178,06497	173,0437	183,2319

➤ Par période et groupe d'âge

Groupe d'âge		Personne	Naissance	Taux de	Intervalle de confiance	
		_année		fécondité	à 95%	
1 (1984-1988)						
(15	-	20]	4661.7586	758	162.59958	151.4267 174.5969
(20	-	25]	4667.1378	1477	316.46805	300.7333 333.0261
(25	-	30]	4620.6002	1510	326.79737	310.7231 343.7032
(30	-	35]	3372.4703	1031	305.71062	287.608 324.9527
(35	-	40]	2615.4482	650	248.52337	230.1338 268.3824
(40	-	45]	2481.4856	324	130.56695	117.0966 145.5869
(45	-	50]	2527.8688	76	30.064851	24.01151 37.64424
2 (1989-1993)						
(15	-	20]	4885.6486	705	144.30019	134.0321 155.3549
(20	-	25]	4199.9083	1371	326.4357	309.6057 344.1805
(25	-	30]	4079.9299	1314	322.06435	305.113 339.9575
(30	-	35]	4170.2278	1253	300.4632	284.2788 317.569
(35	-	40]	3109.3592	761	244.74497	227.9595 262.7664
(40	-	45]	2438.8114	331	135.72185	121.8607 151.1597
(45	-	50]	2356.3021	60	25.463628	19.77109 32.79517
3 (1994-1998)						
(15	-	20]	6221.8806	760	122.14956	113.7668 131.15
(20	-	25]	5145.8117	1459	283.53155	269.3499 298.4599
(25	-	30]	4006.3867	1203	300.27057	283.7732 317.727
(30	-	35]	3720.0592	1108	297.84473	280.8135 315.9088
(35	-	40]	3972.8695	896	225.52968	211.2356 240.7911
(40	-	45]	2973.6698	335	112.65541	101.2153 125.3886
(45	-	50]	2362.4479	54	22.857647	17.50643 29.84457
4 (1999-2003)						
(15	-	20]	7255.3849	684	94.274805	87.46799 101.6113
(20	-	25]	6380.3467	1766	276.78747	264.1747 290.0024
(25	-	30]	4806.0382	1459	303.57645	288.3922 319.5602
(30	-	35]	3642.5019	1089	298.97033	281.7307 317.2649
(35	-	40]	3482.941	813	233.42342	217.9172 250.033
(40	-	45]	3803.6972	423	111.20759	101.0992 122.3267
(45	-	50]	2836.2358	61	21.50738	16.7341 27.64221
5 (2004-2008)						
(15	-	20]	9493.5289	636	66.993002	61.98364 72.4072
(20	-	25]	7670.0422	1904	248.23853	237.335 259.643
(25	-	30]	6188.5619	1842	297.64589	284.359 311.5537
(30	-	35]	4642.6019	1361	293.15458	277.9865 309.1503

(35	-	40]	3553.9402	853	240.0153	224.4369	256.675
(40	-	45]	3374.6748	396	117.34464	106.3381	129.4905
(45	-	50]	3726.6202	80	21.467173	17.24279	26.7265
6 (2009-2011)							
(15	-	20]	6590.083	418	63.428639	57.63042	69.81022
(20	-	25]	5621.8766	1201	213.62973	201.883	226.0599
(25	-	30]	4210.5914	1207	286.65807	270.934	303.2947
(30	-	35]	3450.8115	973	281.96267	264.7911	300.2478
(35	-	40]	2535.4088	617	243.35326	224.8895	263.333
(40	-	45]	2001.6058	229	114.40814	100.5097	130.2285
(45	-	50]	1957.3957	48	24.52238	18.48001	32.5404

ANNEXE C.5 : Age moyen des femmes à la 1ère naissance

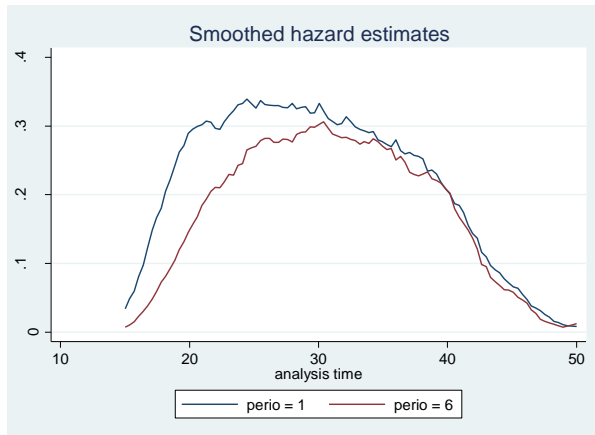
```
. stci, rmean by(calendar_year)
```

```
failure _d: delivery == 1
analysis time _t: (EventDate-origin)/3.16e+10
origin: time DoB
exit on or before: time .
id: IndividualId
```

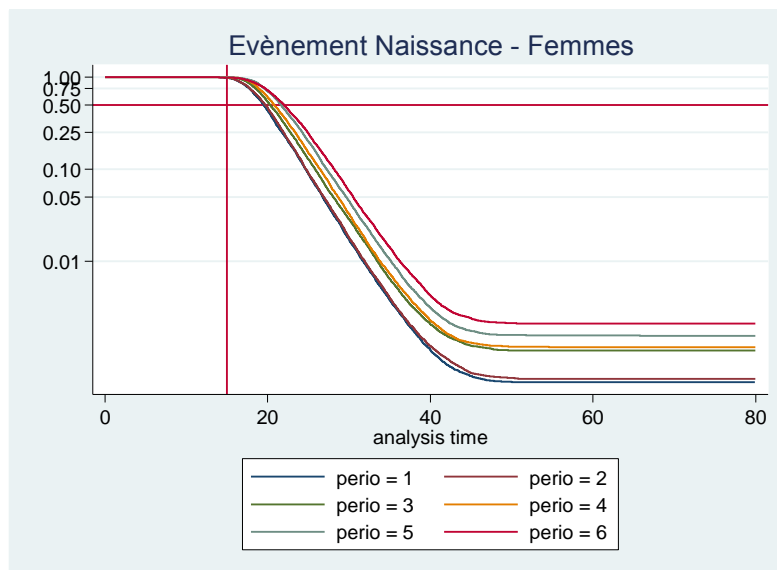
calendar_y~r	no. of subjects	restricted mean	Std. Err.	[95% Conf. Interval]	
0	1	11.73919(*)	0	11.7392	11.7392
1983	12387	20.0911(*)	.2320057	19.6364	20.5458
1984	12804	19.81609(*)	.1911535	19.4414	20.1907
1985	12873	20.52486(*)	.2197124	20.0942	20.9555
1986	13191	19.95696(*)	.2004411	19.5641	20.3498
1987	13593	20.07106(*)	.1873931	19.7038	20.4383
1988	13844	20.46814(*)	.1968638	20.0823	20.854
1989	13880	19.86691(*)	.199218	19.4764	20.2574
1990	14179	20.51459(*)	.2081918	20.1065	20.9226
1991	14101	20.12055(*)	.1966011	19.7352	20.5059
1992	14194	20.44788(*)	.189144	20.0772	20.8186
1993	14391	20.63596(*)	.1861473	20.2711	21.0008
1994	14929	20.74577(*)	.1897954	20.3738	21.1178
1995	15115	21.10971(*)	.1923273	20.7328	21.4867
1996	15578	21.18568(*)	.1812501	20.8304	21.5409
1997	15720	21.11269(*)	.1753026	20.7691	21.4563
1998	16130	21.50087(*)	.1907411	21.127	21.8747
1999	16076	21.56345(*)	.1874006	21.1961	21.9307
2000	16305	21.75761(*)	.177091	21.4105	22.1047
2001	16728	21.61527(*)	.1762563	21.2698	21.9607
2002	17099	21.56675(*)	.1698898	21.2338	21.8997
2003	17360	21.75223(*)	.1770136	21.4053	22.0992
2004	17987	22.17835(*)	.1682676	21.8486	22.5081
2005	18673	22.23053(*)	.1707902	21.8958	22.5653
2006	19141	22.35369(*)	.1738044	22.013	22.6943
2007	19729	22.60753(*)	.1659488	22.2823	22.9328
2008	20451	22.72876(*)	.1643016	22.4067	23.0508
2009	21249	22.47549(*)	.1681998	22.1458	22.8052
2010	21850	22.57995(*)	.1681927	22.2503	22.9096
2011	22418	23.38384(*)	.1937496	23.0041	23.7636
total	40181	21.46342(*)	.0347819	21.3952	21.5316

(*) largest observed analysis time is censored, mean is underestimated

ANNEXE C.6 : Evolution de la répartition des naissances par groupe d'âge des femmes selon la période de naissance, période 1 1984-1988 et période 6 2009-2011.



ANNEXE C.7 : Courbes de Kaplan-Meier, probabilité de survie des femmes n'ayant pas connu l'évènement naissance à chaque âge selon les six périodes



ANNEXE C.8 : Analyse fécondité et religion, test du log-rang et courbe de survie

- TEST LOG RANK

On teste ici l'effet de la variable religion sur le niveau de fécondité des femmes « musulmanes » « chrétiennes » et « animistes ».

```
. sts test relg,logrank
```

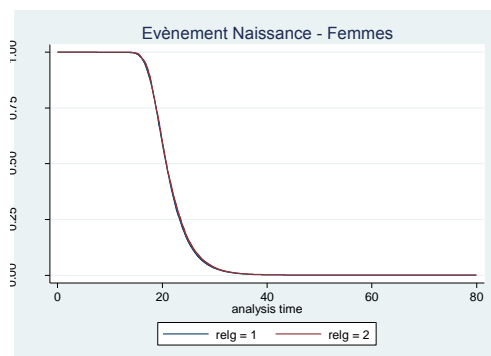
```
failure _d: delivery == 1
analysis time _t: (EventDate-origin)/3.16e+10
origin: time DoB
exit on or before: time .
id: IndividualId
```

Log-rank test for equality of survivor functions

relg	Events observed	Events expected
1	26273	26287.19
2	7195	7216.95
3	740	703.86
Total	34208	34208.00

chi2(2) = 1.93
Pr>chi2 = 0.3805

- Courbe de survie par religion « musulmane » et « chrétienne » (Kaplan-Meier)



- TEST LOG RANK pour tester l'effet de la variable religion sur le niveau de fécondité des femmes « musulmanes » et « chrétiennes ».

```
. sts test relg,logrank
```

```
failure _d: delivery == 1
analysis time _t: (EventDate-origin)/3.16e+10
origin: time DoB
exit on or before: time .
id: IndividualId
```

Log-rank test for equality of survivor functions

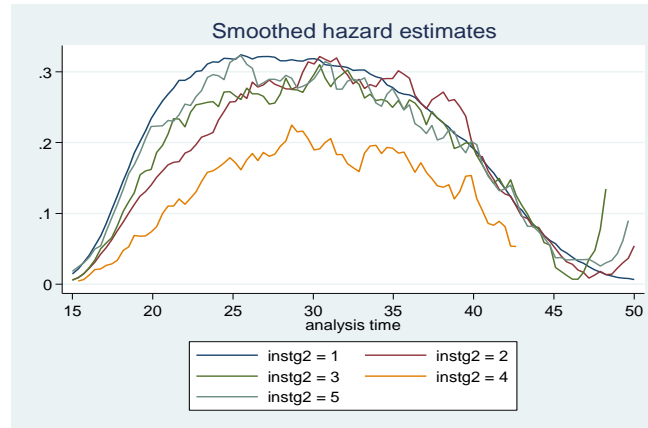
relg	Events observed	Events expected
1	26273	26258.90
2	7195	7209.10
Total	33468	33468.00

chi2(1) = 0.04
Pr>chi2 = 0.8513

ANNEXE C.9 : Analyse fécondité et instruction, test du log-rang

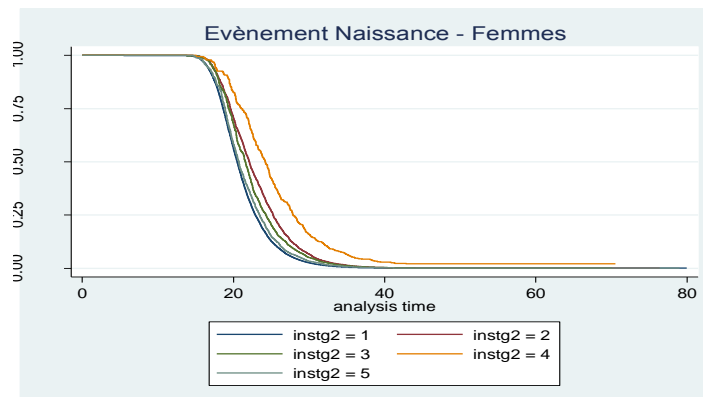
Niveaux : 1 = jamais scolarisé, 2 = primaire incomplet, 3 = primaire complet, 4 = secondaire et plus, 5 = autres (coraniques, arabes, analphabète,...)

Figure : Intensité de connaître l'évènement naissance pour une femme, selon son niveau d'instruction, zone de Niakhar, 1983-2011.



Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

Figure : Courbes de survie (KM) de ne pas connaître l'évènement naissance pour une femme, selon son niveau d'instruction, zone de Niakhar, 1983-2011.



Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

- TEST LOG RANK pour tester l'effet de la variable niveau d'instruction sur le niveau de fécondité des femmes.

```
. sts test instg2,logrank
      failure _d: delivery == 1
      analysis time _t: (EventDate-origin)/3.16e+10
      origin: time DoB
      exit on or before: time .
      id: IndividualId
```

Log-rank test for equality of survivor functions

instg2	Events observed	Events expected
1	28939	27863.72
2	2498	3070.64
3	1204	1359.16
4	357	663.60
5	1155	1195.88
Total	34153	34153.00
	chi2(4) =	315.35
	Pr>chi2 =	0.0000

ANNEXE C.10 : Analyse fécondité et niveau de vie, test du log-rang

- Niveau de vie en deux classes « pauvres » et « non pauvres »

La classe des ménages pauvre comprend 79 % des individus et la classe des ménages non pauvre se compose de 21 % des individus, en 2003. Ainsi plus des trois quart des individus sont considérés comme pauvre.

On observe un niveau de fécondité plus élevé pour les femmes « pauvres », avec un taux de fécondité de 201 enfants pour 1000 femmes, que pour les femmes « non pauvres » où le taux est de 181 enfants pour 1000 femmes (Tableau qui suit), soit un écart de plus de 20 points.

Tableau : Taux global de fécondité en fonction du niveau de vie des femmes, Niakhar, 1999-2003

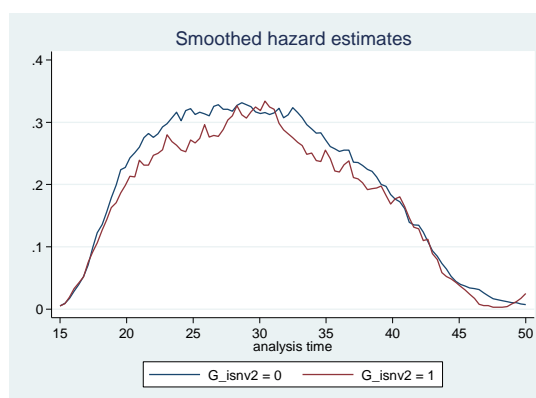
Niveau de vie des femmes	Personne-année (période)	Naissance	Taux global de fécondité ‰
« pauvre »	21511	4322	201
« non pauvre »	6075	1099	181
Total	27586	5421	197

Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

Pour comparer les niveaux de fécondité en tenant compte de l'effet du temps, c'est-à-dire de la durée de vie des femmes sans connaître l'évènement naissance, nous appliquons le test du log-rank afin de savoir s'il existe un effet significatif ou non de la variable ISNV sur la variable de censure delivery. Nous pouvons aussi avant cela observer l'allure des courbes représentant l'intensité des naissances des femmes entre 15 et 49 ans.

La courbe représentant les femmes ayant un niveau de vie plus élevé (en rouge) est souvent inférieure à la courbe représentant les femmes considérées comme « pauvres » (en bleue) (Figure qui suit). On remarque cette tendance surtout pour les femmes entre 18 et 28 ans, puis pour les femmes entre 32 et 40 ans.

Figure : Intensité de connaître l'évènement naissance pour une femme, selon son niveau de vie, zone de Niakhar, 1999-2003.



Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

D'après le test du log-rank, il y a un effet significatif au seuil de 1 % de la variable niveau de vie sur la variable fécondité (delivery) des femmes. Dans l'ensemble les femmes appartenant à la classe des

ménages non pauvres ont moins de « risque » d'avoir des enfants que les femmes appartenant à la classe des ménages pauvres.

- TESTS LOG RANK

1998-2003

- Niveau de vie répartie 2 classes :

```
. sts test G_isnv2, logrank
```

```
      failure _d:  delivery == 1
analysis time _t:  (EventDate-origin)/3.16e+10
      origin:    time DoB
exit on or before: time
      id:        IndividualId
```

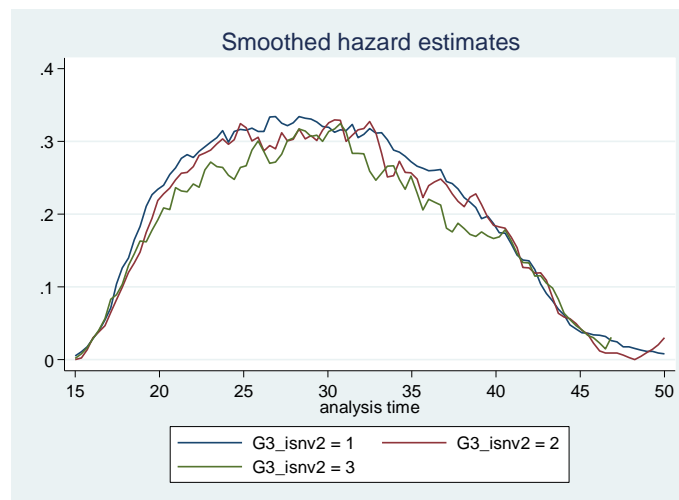
Log-rank test for equality of survivor functions

G_isnv2	Events observed	Events expected
0	4330	4235.79
1	1102	1196.21
Total	5432	5432.00

chi2(1) = 9.53
Pr>chi2 = 0.0020

- Niveau de vie répartie 3 classes :

Graphique : Intensité de connaître l'évènement naissance pour une femme, selon son niveau de vie, zone de Niakhar, 1999-2003.



Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

```

- Test du log-rank :
. sts test G3_isnv2, logrank

      failure _d:  delivery == 1
      analysis time _t:  (EventDate-origin)/3.16e+10
      origin:  time DoB
      exit on or before:  time
      id:  IndividualId

```

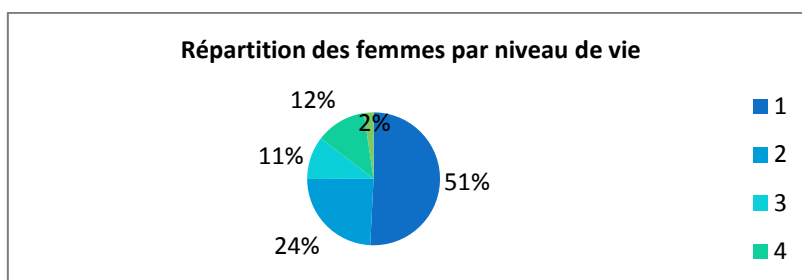
Log-rank test for equality of survivor functions

G3_isnv2	Events observed	Events expected
1	3692	3587.35
2	989	1017.69
3	751	826.97
Total	5432	5432.00
	chi2(2) =	10.86
	Pr>chi2 =	0.0044

- Niveau de vie suivant la classification en 5 classes

La part des femmes les plus pauvres (Groupe 1) reste la plus importante avec 51% des femmes. La classe composée des femmes les plus riches se compose seulement de 2% des femmes. Plus le niveau de vie augmente, moins il y a de femmes (Figure qui suit).

Figure : Répartition des femmes selon le niveau de vie, zone de Niakhar, 1999-2003



Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

Le taux de fécondité le plus bas correspond au taux des femmes les plus riches (148‰) (Tableau qui suit). Le plus élevé est celui des femmes considérées comme peu pauvre (204‰) alors que celui des femmes les plus pauvres est de 202 enfants pour 1000 femmes. Les taux ne se répartissent donc pas en ordre croissant selon l'ordre croissant du niveau de vie des femmes. Cependant la tendance générale reste la même que précédemment, les femmes des classes les plus pauvres ont plus d'enfants que les femmes des classes les plus riches.

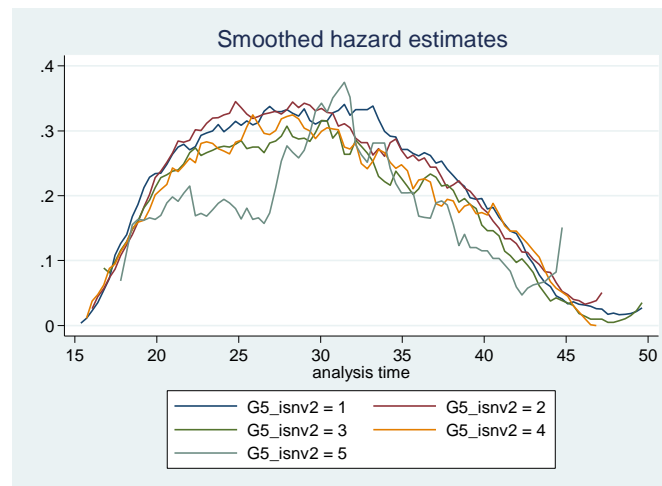
Tableau : Taux global de fécondité en fonction du niveau de vie des femmes, Niakhar, 1999-2003

Niveau de vie des femmes	Personne-année (période)	Naissance	Taux global de fécondité ‰
1 « très pauvre »	13686	2760	202
2 « peu pauvre »	6728	1375	204
3 « pauvre »	2984	536	180
4 « moyen pauvre »	3499	648	185
5 « riche »	689	102	148
Total	27586	5421	197

Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

Avec cette répartition des femmes en cinq classes de niveau de vie on observe quelque chose de nouveau que nous n'avions pas pu voir avec les autres répartitions, les femmes les plus riches (G5_isnv2=5) ont dans l'ensemble leurs enfants à un âge plus avancé, après 30 ans (Figure qui suit).

Figure : Intensité de connaître l'évènement naissance pour une femme, selon son niveau de vie, zone de Niakhar, 1999-2003.



Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

- Test du log-rank : niveau de vie répartie 5 classes :

```
. sts test G5_isnv2, logrank
```

```
failure _d: delivery == 1
analysis time _t: (EventDate-origin)/3.16e+10
origin: time DoB
exit on or before: time .
id: IndividualId
```

Log-rank test for equality of survivor functions

G5_isnv2	Events observed	Events expected
1	2760	2665.07
2	1375	1340.16
3	536	590.47
4	648	688.79
5	102	136.51
Total	5421	5421.00
	chi2(4) =	20.48
	Pr>chi2 =	0.0004

➤ Niveau de fécondité et niveau de vie des femmes en 2003

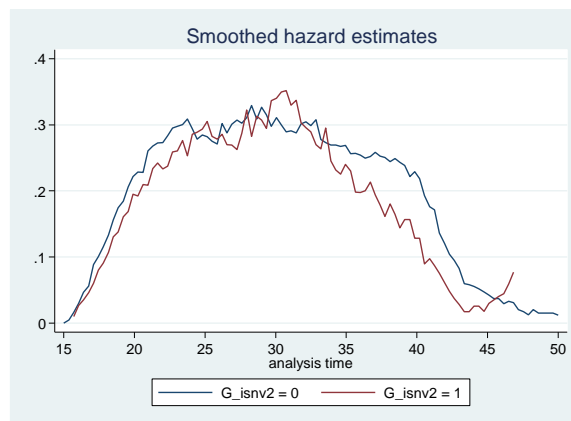
FECONDITE ET NIVEAU DE VIE EN 2003 DANS LA ZONE DE NIAKHAR

L'étude de l'effet du niveau de vie des ménages sur la fécondité en 2003 se fait sur un échantillon de 15606 femmes de tout âge. Cette année-là 1130 enfants sont nés.

La courbe représentant les femmes ayant un niveau de vie plus élevé (en rouge) est souvent inférieure à la courbe représentant les femmes considérées comme « pauvres » (en bleue), ce qui laisse penser que les femmes les plus riches ont moins d'enfants que les femmes les plus pauvres (Figure 20). On remarque cette tendance surtout pour les femmes entre 20 et 25 ans, puis pour les femmes aux âges plus élevés, après 35 ans.

D'après le test du log-rank l'effet de la variable niveau de vie sur la fécondité des femmes est significatif au seuil de 10%. Les femmes appartenant à la classe des ménages non pauvres ont moins de « risque » d'avoir des enfants que les femmes appartenant à la classe des ménages pauvres (Annexe C.10).

Figure : Intensité des naissances par âge de la femme, en fonction du niveau de vie du ménage, 2003, zone de Niakhar.



Exploitation des données du suivi démographique de la zone de Niakhar, 1983 à 2011

- Test du log-rank

```
. sts test G_isnv2, logrank
      failure _d:  delivery == 1
      analysis time _t:  (EventDate-origin)/3.16e+10
      origin:  time DoB
      exit on or before:  time .
      id:  IndividualId
```

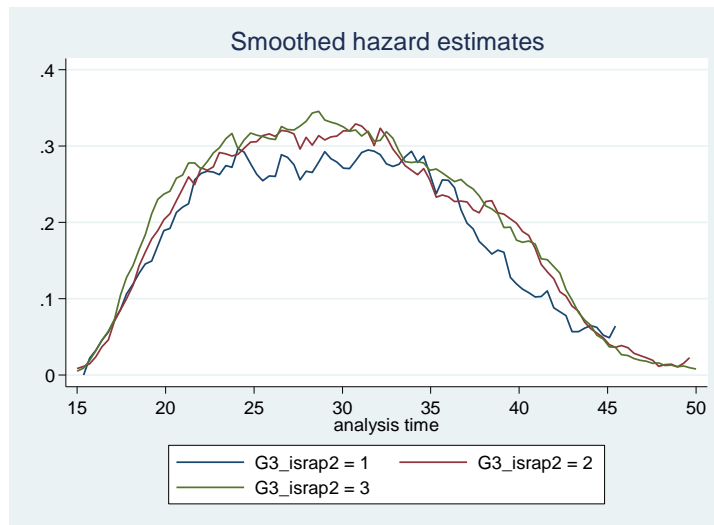
Log-rank test for equality of survivor functions

G_isnv2	Events observed	Events expected
0	900	876.18
1	230	253.82
Total	1130	1130.00

chi2(1) = 2.91
Pr>chi2 = 0.0883

ANNEXE C.11 : Analyse fécondité et niveau de richesse agro-pastorale, test du log-rang

Figure : Intensité des naissances selon le niveau de richesse agro-pastorale (1= « très pauvre », 2= « pauvre » et 3= « riche »)



```

- Test du log-rank
. sts test G3_israp2, logrank
      failure _d:  delivery == 1
      analysis time _t: (EventDate-origin)/3.16e+10
      origin: time DoB
      exit on or before: time
      id: IndividualId
    
```

Log-rank test for equality of survivor functions

G3_israp2	Events observed	Events expected
1	390	444.56
2	1545	1586.47
3	3497	3400.97
Total	5432	5432.00
	chi2(2) =	10.52
	Pr>chi2 =	0.0052

ANNEXE C.12 : Modèle de Cox

➤ SORTIE MODELE DE COX sans interaction

```
. stset EventDate if residence==1 & Sex==2, id(IndividualId) failure(delivery==1) ///
>      origin(time DoB) time0(datebeg) exit(time .) scale(31557600000)
```

```
      id: IndividualId
      failure event: delivery == 1
obs. time interval: (datebeg, EventDate]
exit on or before: time .
  t for analysis: (time-origin)/3.16e+10
      origin: time DoB
      if exp: residence==1 & Sex==2
```

```
147203 total obs.
70956  ignored at outset because of -if <exp>-
```

```
76247 obs. remaining, representing
13706 subjects
5387  failures in multiple failure-per-subject data
55473.24 total analysis time at risk, at risk from t = 0
      earliest observed entry t = 0
      last observed exit t = 103.8564
```

```
. stcox edu1 edu2 edu3 eduA nv1 nv2 rap1 rap2
```

```
      failure _d: delivery == 1
      analysis time _t: (EventDate-origin)/3.16e+10
      origin: time DoB
exit on or before: time .
      id: IndividualId
```

```
Iteration 0: log likelihood = -36115.647
Iteration 1: log likelihood = -36080.738
Iteration 2: log likelihood = -36080.155
Iteration 3: log likelihood = -36080.154
Refining estimates:
Iteration 0: log likelihood = -36080.154
```

Cox regression -- Breslow method for ties

```
No. of subjects =      13706          Number of obs   =      76247
No. of failures =       5387
Time at risk    = 55473.24171
Log likelihood  = -36080.154          LR chi2(8)      =      70.99
                                          Prob > chi2    =      0.0000
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
edu1	.8059281	.0431767	-4.03	0.000	.7255947	.8951554
edu2	.8448703	.0571654	-2.49	0.013	.7399396	.9646812
edu3	.5861782	.065082	-4.81	0.000	.4715445	.7286797
eduA	.8705961	.0650341	-1.86	0.064	.7520235	1.007864
nv1	.9513724	.034423	-1.38	0.168	.8862412	1.02129
nv2	.9144783	.0375167	-2.18	0.029	.8438254	.9910468
rap1	1.084903	.0624084	1.42	0.157	.9692285	1.214384
rap2	1.151972	.0622682	2.62	0.009	1.036171	1.280715

```
. end of do-file
```

- SORTIE MODELE DE COX avec interaction entre les variables instruction (EDU) ET niveau de vie (NV)

```
. stcox edu1 edu2 edu3 eduA nv1 nv2 edu1nv1 edu1nv2 edu2nv1 edu2nv2 edu3nv1 edu3nv2
```

```
      failure _d:  delivery == 1
      analysis time _t:  (EventDate-origin)/3.16e+10
      origin:  time DoB
      exit on or before:  time .
      id:  IndividualId
```

```
Iteration 0:  log likelihood = -36115.647
Iteration 1:  log likelihood = -36080.378
Iteration 2:  log likelihood = -36079.211
Iteration 3:  log likelihood = -36079.204
Iteration 4:  log likelihood = -36079.204
Refining estimates:
Iteration 0:  log likelihood = -36079.204
```

```
Cox regression -- Breslow method for ties
```

```
No. of subjects =          13706          Number of obs   =          76247
No. of failures =           5387
Time at risk   =  55473.24171
Log likelihood = -36079.204          LR chi2(12)      =          72.89
                                          Prob > chi2     =          0.0000
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
edu1	.8605236	.0555526	-2.33	0.020	.7582493 .976593
edu2	.8930052	.0832514	-1.21	0.225	.7438747 1.072033
edu3	.6776439	.1031793	-2.56	0.011	.5028009 .9132864
eduA	.8580434	.0640627	-2.05	0.040	.7412375 .993256
nv1	.9576189	.0372266	-1.11	0.265	.8873665 1.033433
nv2	.9771726	.0433731	-0.52	0.603	.8957556 1.06599
edu1nv1	.8936427	.1251232	-0.80	0.422	.6791772 1.175831
edu1nv2	.7273016	.110378	-2.10	0.036	.5401732 .9792555
edu2nv1	1.063113	.1723234	0.38	0.706	.7737587 1.460675
edu2nv2	.7160847	.1223715	-1.95	0.051	.5122746 1.000981
edu3nv1	.842603	.268069	-0.54	0.590	.451667 1.57191
edu3nv2	.6721978	.1671456	-1.60	0.110	.4128965 1.094342

```
. end of do-file
```

ANNEXE D : PRESENTATION DE L'IRD AU SENEGAL

LES MOYENS AU SERVICE DE LA RECHERCHE (p. 43, La représentation au Sénégal, IRD, rapport 2011).

Organigramme de la Représentation de l'IRD au Sénégal avec extension de compétences au Cap-Vert, à la Gambie, à la Guinée et à la Mauritanie

Juin 2011



ANNEXE E : PRESENTATION DU SYSTEME DE SUIVI DEMOGRAPHIQUE DE LA ZONE DE NIAKHAR ET DE SES DONNEES

Les données analysées sont issues du système de suivi démographique (SSD) de Niakhar au Sénégal. Ce système de suivi fait partie des dispositifs d'enquêtes exhaustives permettant de suivre l'évolution de la population géographiquement circonscrite, ici celle de la zone de Niakhar.

La zone d'étude de Niakhar s'étend sur un peu plus de 200 km², dans la région sahélienne du bassin arachidier sénégalais à environ 150 km au sud-est de Dakar. Les limites et l'organisation actuelle ont été fixées en 1983. Toutes les informations sur la population (naissances, mariages, fécondité, migrations, décès, épidémies, vaccinations, scolarisations, activités professionnelles) sont obtenues par des enquêtes dans chaque concession. Ces enquêtes sont effectuées par des enquêteurs professionnels spécialisés qui interrogent les habitants chaque trimestre. La zone d'étude est composée de 30 villages. L'ensemble de ces villages constitue la zone d'étude qui est ensuite divisée en cinq sous-zones dirigées chacune par un enquêteur. Ce dispositif permet donc un suivi longitudinal de population par un enregistrement continu des événements vécus par les individus au cours de leur(s) séjour(s) dans la zone d'observation (incluant au minimum les naissances, les décès et les migrations). Il s'agit d'une fenêtre d'observation à la fois géographique et temporelle¹³.

Toutes les informations collectées font l'objet de vérifications successives tout au long du processus de collecte et de traitement informatique. La base de données doit refléter la situation sur le terrain. Elle est utilisée pour l'extraction de fichiers d'analyse et pour la confection de tableaux statistiques résumant la situation démographique et épidémiologique dans la zone d'étude.

La collecte démographique dans la zone de Niakhar a été initiée en décembre 1962 par Pierre Cantrelle (ORSTOM), et est encore dans le système de suivi de nos jours.

¹³ Valérie Delaunay, Adama Marra, et Pierre Levi, « Analyser la fécondité à partir de données de Systèmes de Suivi Démographique, Application au site de Niakhar, Sénégal » (Les collections du CEPED, 2006).

Résumé

Ce présent document est le rapport de mon stage effectué à l'Institut de Recherche pour le Développement (IRD) de Dakar au sein de l'unité de recherche sur les maladies infectieuses et tropicales émergentes (UMR 198 - URMITE).

La mission qui m'a été confiée consistait, d'une part, à étudier l'évolution de la fécondité dans la zone de Niakhar, zone rurale du Sénégal, entre 1983 et 2011 et, d'autre part, à mettre en place une démarche d'analyse permettant l'analyse de ces données issues du système de suivi démographique de cette zone.

Pour parvenir à l'objectif de cette mission plusieurs étapes ont été essentielles.

J'ai dû tout d'abord produire les programmes d'analyse suivant les procédures bien spécifiques de la méthode d'analyse des biographies, mis en œuvre sous le logiciel Stata. Cette méthode permet d'étudier au niveau individuel l'arrivée d'un ou plusieurs phénomènes, mesurée par une durée. Ici l'évènement étudié est la « naissance d'un enfant » dans le parcours de vie des femmes résidentes de la zone de Niakhar.

Après avoir obtenu les indicateurs transversaux de la fécondité, j'ai pu étudier l'évolution de la fécondité de cette zone. Les résultats montrent une tendance à la baisse de la fécondité, avec un indice synthétique de fécondité passant de 8 enfants par femme en 1984 à 6 enfants par femme en 2011. On constate aussi un changement de calendrier dans le sens d'un retard des naissances avec une baisse des taux de fécondité avant 25 ans, qui se traduit par une légère augmentation de l'âge à la maternité.

De plus cette analyse a été approfondie en étudiant certains déterminants de la fécondité. L'étude nous apprend que le niveau de fécondité des femmes baisse au plus le niveau d'instruction augmente. Il en est de même avec le niveau de vie, le niveau de fécondité baisse plus le niveau de vie à tendance à être élevé. A l'inverse au plus le niveau de richesse agro-pastorale est élevée au plus le niveau de fécondité est élevé aussi. La religion, quant à elle, n'est pas un déterminant significatif du niveau de fécondité dans cette étude.

Mots clés : Système de suivi démographique (SSD), fécondité, méthode d'analyse des biographies, zone de Niakhar, indicateurs transversaux de fécondité, analyses des données, femmes, indice synthétique de fécondité (ISF), âge moyen à la maternité (AMM).