

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

IMPACTS DE LA CONSANGUINITÉ MATERNELLE  
SUR LA MORTALITÉ DE LA PROGÉNITURE  
DANS UNE POPULATION ENDOGAME DU QUÉBEC

MÉMOIRE  
PRÉSENTÉ  
COMME EXIGENCE PARTIELLE  
DE LA MAÎTRISE EN BIOLOGIE

PAR  
SYLVIE MAHEU

AOÛT 2001

UNIVERSITÉ DU QUÉBEC À MONTRÉAL  
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.04-2020). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

## REMERCIEMENTS

Je tiens tout d'abord à remercier ma directrice de thèse, le Dr Francine M. Mayer, à qui je suis très redevable pour son assistance constante, ses conseils éclairés, ses encouragements et son humanisme plein d'humour. Mme Mayer, sans votre soutien, ce projet n'aurait jamais vu le jour.

À l'intérieur de l'équipe de recherche du Dr Mayer, des gens travaillent très fort. Un gros merci à Mme Mireille Boisvert, assistante de recherche, dont la précision et l'exactitude du travail sont à la base du présent ouvrage. Merci aussi à M. Éric Labelle, responsable de la gestion des bases de données, pour sa patience devant l'ampleur de mes requêtes informatiques. Des gens de l'extérieur m'ont également très bien conseillée. Je tiens donc à remercier Mme Sylvie de Grosbois, du CINBIOSE, et M. Bertrand Fournier, du SCAD, pour leurs précieux « conseils statistiques ».

Merci à mes collègues de maîtrise, François, Dalinda et Mathieu, avec qui il a été si agréable de travailler. Un merci supplémentaire à toi, Mathieu, qui partages mes joies et mes peines, sans parler de mon stress...

Merci à mes amis qui ont accepté mes fréquentes absences à leurs sorties pendant ma rédaction. Votre amitié m'est chère.

Enfin, merci à mes parents qui m'ont soutenue et encouragée, et ce, à tous les niveaux depuis le début de mes études. Vous êtes mon exemple et mon inspiration.

## TABLE DES MATIÈRES

LISTE DES FIGURES.....	viii
LISTE DES TABLEAUX.....	ix
RÉSUMÉ.....	xi
INTRODUCTION.....	1
Notions de biologie évolutive.....	1
La consanguinité dans une population.....	1
Le maintien du polymorphisme dans une population.....	3
CHAPITRE I	
ÉTAT DES CONNAISSANCES.....	5
1.1 L'apparement et la consanguinité, des facteurs néfastes?.....	5
1.1.1 Tout n'est pas si clair.....	7
1.1.2 La compensation reproductive.....	7
1.1.3 Les conclusions prudentes.....	8
1.2 La consanguinité, un facteur neutre?.....	9
1.3 La consanguinité liée à l'endogamie, un moyen d'éliminer les gènes délétères?.....	10
1.4 Le débat est ouvert!.....	11
1.5 Notions d'histocompatibilité.....	12
1.5.1 Le MHC chez les populations humaines.....	12
1.5.2 Essais sur des populations de souris.....	15
1.5.3 Essais sur des populations humaines.....	16

## CHAPITRE II

HYPOTHÈSES DE TRAVAIL.....	18
2.1 Notions de biologie évolutive appliquées.....	18
2.2 Élaboration des hypothèses.....	19
2.3 Études menées par le laboratoire du Dr Francine M. Mayer sur la consanguinité.....	22
2.4 Implications scientifiques de l'étude.....	23

## CHAPITRE III

POPULATION À L'ÉTUDE.....	25
3.1 La population de l'Île-aux-Coudres.....	25
3.2 Le Registre de la population.....	28

## CHAPITRE IV

MÉTHODOLOGIE.....	29
4.1 Critères de sélection des couples.....	29
4.2 La progéniture totale.....	34
4.2.1 Création d'un fichier par individus.....	34
4.2.2 Les exclus.....	36
4.2.3 Tests comparatifs entre les inclus et les exclus.....	38
4.3 Critique de l'information généalogique ; l'entropie et la variance (Kouladjian, 1986).....	42
4.4 Le coefficient de consanguinité.....	44
4.5 Le coefficient de parenté.....	46
4.6 Consanguinité proche et éloignée.....	47
4.7 Technique d'analyse choisie.....	51
4.7.1 Approche multivariée : la modélisation par régression logistique.....	51
4.7.2 Le rapport de cotes.....	53
4.8 Transformation des variables de mortalité pour répondre à un mode binaire.....	55

4.9	Création des classes d'apparement et de consanguinité pour les modèles de régression logistique.....	57
4.10	Organisation de l'analyse des modèles de régression logistique.....	59
4.11	Choix des variables soumises aux régressions logistiques et tests de corrélations.....	60
4.11.1	Les variables démographiques.....	61
4.11.2	Les variables génétiques.....	63
4.11.3	Tests de symétrie des variables génétiques.....	65
CHAPITRE V		
	ANALYSES ET RÉSULTATS.....	66
5.1	Analyse descriptive de la variance et de l'entropie généalogiques de l'effectif.....	66
5.2	Analyse descriptive des variables sélectionnées pour être soumises aux régressions logistiques.....	67
5.2.1	Distribution des fréquences pour les variables génétiques.....	67
5.2.2	Distribution des variables démographiques.....	69
5.2.3	Tables de contingences pour qualifier la variable « année de naissance ».....	73
5.3	Présentation des modèles de régression logistique.....	75
5.3.1	Classes de mortalités touchant la progéniture totale.....	76
5.3.2	Classes de mortalités touchant la progéniture masculine.....	87
5.3.3	Classes de mortalités touchant la progéniture féminine.....	95
5.4	Analyse de l'ajustement et du pourcentage explicatif des modèles.....	100
5.5	Analyse de la calibration et des capacités de discrimination des modèles.....	103
5.5.1	Tests de Hosmer et Lemeshow : la calibration.....	103
5.5.2	Tables de classification et « ROC curves » : la capacité de discrimination.....	105

CHAPITRE 6	
DISCUSSION.....	107
6.1 Les variables démographiques.....	111
6.1.1 La taille de la famille.....	111
6.1.2 L'intervalle génésique.....	112
6.1.3 L'âge de la mère à la naissance et le rang de naissance de l'enfant.....	113
6.1.4 La différence d'âge entre les conjoints.....	115
6.1.5 L'année de naissance de l'enfant.....	117
6.2 Les variables génétiques.....	119
6.2.1 L'apparentement des conjoints.....	120
6.2.2 La consanguinité maternelle.....	121
6.2.3 Les inclus et les exclus.....	123
CONCLUSION.....	127
APPENDICE A	
LISTE DES NOMS COMPLETS DES VARIABLES ASSOCIÉS À LEUR ACRONYME.....	136
APPENDICE B	
CORRÉLATIONS ENTRE LES VARIABLES UTILISÉES.....	138
APPENDICE C	
DISTRIBUTION DES VARIABLES UTILISÉES.....	143
APPENDICE D	
ANALYSE DE LA MORTALITÉ PAR RAPPORT À L'ANNÉE DE NAISSANCE.....	151

APPENDICE E	
TESTS DE SYMÉTRIE DES VARIABLES GÉNÉTIQUES.....	168
APPENDICE F	
MODÈLES DE RÉGRESSION LOGISTIQUE FAITS AVEC LA SÉLECTION ENTER.....	172
APPENDICE G	
ÉVALUATION DE L'AJUSTEMENT ET DES POURCENTAGES EXPLICATIFS DES MODÈLE DE RÉGRESSION LOGISTIQUE.....	182
APPENDICE H	
ÉVALUATION DE LA CALIBRATION DES MODÈLES DE RÉGRESSION LOGISTIQUE.....	189
APPENDICE I	
ÉVALUATION DE LA CAPACITÉ DE DISCRIMINATION DES MODÈLES DE RÉGRESSION LOGISTIQUE.....	199
APPENDICE J	
DISTRIBUTION DES VARIABLES UTILISÉES POUR LES TESTS COMPARATIFS ENTRE LE GROUPES DES « INCLUS », LE GROUPE DES « EXCLUS » ET LE GROUPE « COUPLES ».....	228
BIBLIOGRAPHIE.....	253

## LISTE DES FIGURES

Figure	Page
4.1 Représentation schématique de l'application des critères de sélection à la base de données informatisée de l'Île-aux-Coudres.....	32
4.2 Formules et exemple de calcul de l'entropie et de la variance à partir de la représentation de l'arbre généalogique binaire théorique de l'individu Z.....	43
4.3 Chaînes de consanguinité reliant l'individu Z aux ancêtres A et B, communs à son père (Y) et à sa mère (X).....	45
4.4 Illustration de la quantité d'information disponible pour le calcul des coefficients de consanguinité (l'individu A OU l'individu B) comparée à la quantité d'information disponible pour le calcul de l'apparentement entre A et B pour chaque génération (pour une profondeur de 3 générations).....	50
4.5 Illustration de la forme générale d'une courbe de fonction logistique et formule associée à la régression logistique.....	53

## LISTE DES TABLEAUX

Tableau	Page
4.1 Représentation des intervalles de valeurs sélectionnées pour former les classes d'apparement et de consanguinité avec les valeurs d'apparement correspondant à un lien social reconnu.....	58
5.1 Distribution de la progéniture selon les classes d'apparement et de consanguinité des parents pour les variables génétiques totales, proches et éloignées.....	68
5.2 Moyennes et écart-types de certaines variables démographiques pour l'ensemble de l'effectif sur toute la période étudiée.....	72
5.3 Modèles de régression logistique effectués par sélection FORWARD touchant la progéniture totale et soumettant les variables génétiques totales.....	77
5.4 Modèles de régression logistique effectués par sélection FORWARD touchant la progéniture totale et soumettant les variables génétiques proches.....	81
5.5 Modèles de régression logistique effectués par sélection FORWARD touchant la progéniture totale et soumettant les variables génétiques éloignées.....	84
5.6 Modèles de régression logistiques effectués par sélection FORWARD touchant la progéniture masculine et soumettant les variables génétiques totales.....	87

5.7	Modèles de régression logistique effectués par sélection FORWARD touchant la progéniture masculine et soumettant les variables génétiques proches.....	89
5.8	Modèles de régression logistique effectués par sélection FORWARD touchant la progéniture masculine et soumettant les variables génétiques éloignées.....	92
5.9	Modèles de régression logistique effectués par sélection FORWARD touchant la progéniture féminine et soumettant les variables génétiques totales.....	95
5.10	Modèles de régression logistique effectués par sélection FORWARD touchant la progéniture féminine et soumettant les variables génétiques proches.....	97
5.11	Modèles de régression logistique effectués par sélection FORWARD touchant la progéniture féminine et soumettant les variables génétiques éloignées.....	99

## Résumé:

Dans les populations pratiquant des traditions endogames depuis plusieurs générations, on constate que la plupart des mariages s'effectuent entre individus apparentés. Cet apparentement est justement dû à la tradition de mariage interne de la population; le manque d'immigration empêche l'apport de nouveaux allèles dans le pool génique et les individus résultants sont consanguins, leurs parents partageant des ancêtres communs. De façon générale, les populations où le coefficient moyen de consanguinité est élevé feraient face à des problèmes de morbidité et de mortalité infantile liées à des problèmes génétiques. De plus, beaucoup de recherches tentent de prouver que cette endogamie se traduit par des problèmes d'histocompatibilité entre la mère et son fœtus. La population québécoise de l'Île-aux-Coudres a vécu, depuis sa colonisation, une tradition endogame et plusieurs études ont porté, entre autres, sur les effets de la consanguinité sur la mortalité infantile. La présente étude propose une approche de modélisation par régression logistique pour caractériser non seulement les effets de la consanguinité de la mère sur la mortalité infantile mais également l'apparentement des parents, la consanguinité paternelle ainsi que plusieurs variables socio-démographiques. Plusieurs contrôles de l'effectif ont été faits afin d'obtenir des modèles satisfaisants permettant d'interpréter l'impact de chaque variable indépendante sur la variable dépendante : la mortalité de la progéniture. Plusieurs tests servant à quantifier l'information généalogique disponible et à caractériser l'effectif ont été appliqués. Afin de bien comprendre l'impact des variables de nature génétique, celles-ci ont été découpées en composantes proches et éloignées. Les modèles obtenus sont classés selon le type de mortalité étudié : mortalité (0 jours), mortalité néonatale précoce (entre 1 et 6 jours), mortalité infantile, mortalité entre 1 et 4 ans et mortalité entre 1 et 14 ans. Ils ont également été séparés par types de variables génétiques : totales, proches ou éloignées. Les modèles sont aussi présentés séparément pour la progéniture totale, les filles seulement et, finalement, les garçons. Les variables socio-démographiques ont été prises en compte afin de comprendre l'impact qui leur est propre sur la mortalité de la progéniture pour comprendre l'impact et renforcer la compréhension des effets des variables génétiques en les mettant dans un contexte socio-démographique. Ainsi, l'apparentement des parents et la consanguinité de la mère ont eu des impacts notables sur le destin de la progéniture. La consanguinité paternelle est demeurée sans effet dans tous les modèles. L'apparentement manifeste un effet délétère sur la progéniture féminine pour la première année de vie, ce qui reflète les risques d'homozygotie pour un allèle délétère. Pour les garçons, les effets de l'apparentement se manifestent pour les mortalités entre 1 et 4 ans ainsi que 1 et 14 ans et dénotent cette fois-ci un effet protecteur. L'interprétation de ces derniers résultats est sociale puisqu'elle réfère au soutien offert par une famille élargie par l'apparentement entre les parents. La consanguinité de la mère, surtout éloignée, manifeste une forte protection contre la mortalité des filles ce qui reflèterait l'épuration des allèles délétères au fil des générations. Les relations mère-fœtus sont aussi mises en cause vu les indices de l'impact du MHC sur le rapport de masculinité inversé suggérant un surplus de mortalité des fœtus masculins.

## INTRODUCTION

### Notions de biologie évolutive

Les populations sont les sous-unités d'une espèce et elles constituent pour celle-ci autant d'unités reproductives. Chez l'espèce humaine, une sous-unité reproductrice désigne une population dont les individus ont plus de chances de se reproduire entre eux qu'avec les individus d'une autre population et ce, pour des raisons géographiques (distance, insularité) ou culturelles (religion, traditions de mariages différentes dans les groupes humains, etc.). Ainsi, pour chaque population correspond un pool génique particulier. On définit le pool génique comme "l'ensemble des gènes dans les gamètes reproductibles d'une population, c'est-à-dire la totalité de l'information génétique contenue dans une population" (Beaudry, 1985). À l'intérieur d'une population, les unions sont dites panmictiques si elles sont "faites au hasard". Cependant, si la population pratique une tradition de mariage particulière en évitant ou en favorisant des unions entre certains types d'individus selon leur rang social, leur religion ou leur lien familial, alors l'équilibre panmictique de la population est rompu et cela pourrait avoir des conséquences sur le pool génique de la population. Ainsi, une population qui évite un type de mariage particulier de sorte que celui-ci arrive moins fréquemment que sa probabilité en panmixie, est dite exogame (Griffits et al., 1997). Une population endogame désigne une population où les mariages ne s'effectuent qu'entre individus appartenant à la population. Bien qu'elle puisse être panmictique, le fait de garder toujours le même pool génique sans ajouts de l'extérieur accumule au fil des générations des liens multiples entre les individus de la population. De cette manière, les individus finissent par avoir plusieurs ancêtres communs et être apparentés de façon éloignée par des liens nombreux. Après plusieurs générations de mariages entre individus apparentés de près ou de loin, une certaine homogénéisation du pool génique s'effectue au sein de la population (effet de consanguinité de dérive) (Futuyma, 1986). D'ailleurs, dans un de ses articles, Bittles (1994) aborde ce sujet et affirme que dans les recherches actuelles, il y a une sous-estimation du coefficient moyen de

consanguinité d'une population endogame où le choix des conjoints est restreint en raison de la petitesse et/ou de l'isolement de la communauté et ce, en présence ou non d'une tradition de mariage entre apparentés. À l'intérieur de la population, tous les individus finissent par être apparentés et ont donc une probabilité plus élevée d'avoir, pour un même gène, deux copies d'allèles identiques issus d'un même ancêtre.

### **La consanguinité dans une population**

La consanguinité moyenne d'une population est liée à son degré d'endogamie, à la taille de ses effectifs, à son ancienneté (nombre de générations d'individus ayant constitué cette population), à son histoire démographique (reproduction différentielle) et à sa tradition de choix de conjoint. On peut mesurer à l'aide du coefficient de consanguinité individuel proposé par Wright (1922). Ce coefficient de consanguinité « f » est donc la probabilité pour que les deux gènes qu'un individu possède en un locus quelconque soient identiques, c'est-à-dire soient la copie d'un même gène ancêtre. Le coefficient de consanguinité d'un individu particulier correspond au coefficient d'apparement de ses parents puisque celui-ci égale la probabilité pour qu'un gène pris au hasard chez A (père) soit identique à un gène pris au même locus chez B (mère) (Malécot, 1948). Le coefficient de consanguinité moyen de la population se calcule comme étant la somme des quotients de tous les individus de la population divisée par le nombre d'individus dans la population à un moment donné. Jakobi et Jacquard (1971) le définissent comme étant la probabilité pour que les deux gènes possédés en un locus quelconque par un individu choisi au hasard dans la population soient identiques. Ce « f » théorique dépend toujours de l'information dont on dispose pour le calculer.

Au niveau de la population, la consanguinité produit un surplus d'homozygotes. Ce phénomène a pour effet d'exposer à la sélection naturelle une proportion plus grande des phénotypes récessifs (Khlal et Khoury, 1991). La proportion de gènes délétères ou létaux à l'état homozygote augmente elle aussi et les individus affublés de cette double

dose ont moins de chances de survivre et/ou de se reproduire, ce qui élimine une partie de ces allèles du pool génique. Toutefois, une certaine proportion de ces allèles reste dans la population sous forme hétérozygote. La conséquence de cette consanguinité est la baisse du polymorphisme i.e. l'homogénéisation du pool génique de la population à mesure que les gènes les plus désavantageux sont éliminés et, avec eux, des gènes liés peuvent aussi être purgés (Futuyma, 1986). Le polymorphisme peut se définir comme étant la coexistence de deux ou plusieurs allèles codant pour un même gène au sein d'une population (Beaudry, 1985). La perte de variabilité résulte des interactions entre l'endogamie, la consanguinité et la sélection naturelle. Cet appauvrissement du pool génique se nomme dépression consanguine et s'observe chez des populations endogames de petits effectifs se retrouvant en état d'isolation reproductive lorsque les caractères phénotypiques les plus avantageux sont ceux des homozygotes dominants ou des hétérozygotes (Futuyma, 1986).

#### **Le maintien du polymorphisme dans une population:**

Plusieurs théories sur le maintien du polymorphisme et de la variabilité génétique ont été élaborées. Ainsi, des allèles également avantageux, les mutations, le flux génique, les avantages hétérozygotiques ou homozygotiques, agissant de pair avec les pressions de sélection, contribuent à conserver un certain polymorphisme propre à chaque population. Des pressions environnementales jumelées aux réactions de la population exposée à ces changements (variation de la fécondité et/ou de la mortalité, etc) peuvent également changer l'équilibre du polymorphisme et ainsi faire varier les fréquences alléliques du pool jusqu'à un nouvel équilibre (Futuyma, 1986). De même, les croisements panmictiques dans une population favorisent le maintien du polymorphisme. Une tradition d'évitement des mariages entre individus apparentés proches, même dans une population endogame, atténuerait la vitesse d'homogénéisation. Au contraire, un choix préférentiel du conjoint parmi les apparentés proches accélérerait le processus. De plus, les populations de petits effectifs sont plus sensibles à tous ces facteurs modifiant le rythme d'homogénéisation. Il est donc

important d'avoir une bonne connaissance des caractéristiques démographiques de la population à l'étude. Bien comprendre sa dynamique sociale permet une meilleure interprétation de son niveau d'endogamie et de ses coefficients moyens de consanguinité et d'apparentement.

Un autre mécanisme qui contribue au maintien du polymorphisme est la compensation reproductive. Bien que ce phénomène soit difficile à mettre en évidence selon Bittles (1994), on parle de compensation reproductive lorsqu'un couple remplace un enfant homozygote récessif mort en bas âge par un autre enfant hétérozygote ou homozygote dominant. De cette manière, la compensation reproductive contribuerait à un certain ralentissement de l'homogénéisation du pool génique.

Le fait de maintenir le polymorphisme à l'intérieur d'une population entraîne nécessairement la présence d'allèles délétères dans le pool génique. L'ensemble de ces allèles constitue le fardeau génétique de la population. Wallace (1989) a défini le fardeau génétique comme "le prix à payer pour maintenir la diversité génétique".

## **CHAPITRE I**

### **ÉTAT DES CONNAISSANCES**

Plusieurs recherches ont déjà été faites sur diverses populations dans le but principal de mesurer l'impact de l'apparentement biologique sur la fécondité des couples, la fertilité et la mortalité infantile. Les études effectuées ont principalement porté sur des sous-groupes relativement importants dans une population plus grande ou sur des isolats. Les isolats, populations de petit effectif où on ne retrouve pas (ou peu) d'immigration, constituent des "laboratoires" idéaux pour ce genre de recherche englobant à la fois démographie et génétique.

La plupart des recherches arrivent à la conclusion que l'apparentement des parents entraîne un risque plus élevé de mortalité infantile et de tares génétiques chez les enfants. Le risque augmenterait selon l'importance du lien biologique entre les deux conjoints puisqu'il augmente la probabilité de retrouver chez leur progéniture des gènes létaux à l'état homozygote. C'est donc de cette manière que certains auteurs (Bittles, 1994; Sutter, 1968) expliquent que leurs résultats montrent un taux de mortalité infantile significativement plus élevé pour la progéniture de couples apparentés que pour les couples témoins dont l'apparentement est nul.

#### **1.1 L'apparentement et la consanguinité, des facteurs néfastes?**

C'est aussi le cas de William J. Shull et ses associés qui ont écrit une série d'articles (Shull et al., 1970a, 1970b; Shull et Neel, 1972) sur les effets de l'apparentement et de la consanguinité sur les fausses couches, la mortalité pré-reproductive, la fécondité et la compensation reproductive dans la population japonaise d'Hirado. L'apparentement, dans cette étude, correspond au lien biologique unissant les deux conjoints et ce lien est déclaré s'il est supérieur ou égal à l'apparentement de cousins de second degré. La consanguinité réfère à la consanguinité respective de chaque conjoint. Afin d'établir un statut de

consanguinité chez un individu, Shull a remonté une génération pour voir si les parents de cet individu étaient apparentés. Shull a étudié en parallèle des sous-populations de fermiers et de non-fermiers et a établi pour les couples un « score socio-économique » (SES) à partir de plusieurs variables démographiques, soit le niveau d'éducation et le métier de chaque conjoint, les dépenses allouées pour la nourriture par mois par personne, l'espace vital de chaque personne dans la maison (household density) et les revenus annuels selon les rapports d'impôt. Pour les fermiers, des données concernant les caractéristiques de leurs terres ont aussi été incluses. Ce score, brut ou pondéré selon qu'il tienne compte proportionnellement ou non de l'importance de chaque variable englobée, avait pour but de simplifier l'approche démographique de l'étude et de mettre en évidence les variables d'ordre génétique.

Les résultats des recherches de Shull mettent plusieurs points en évidence. Notamment, il a été prouvé que l'apparentement des couples et la consanguinité (lorsqu'un ou les deux membre(s) du couple est(sont) issu(s) d'un mariage entre apparentés) sont distribués différemment dans les deux sous-populations (non-fermiers et fermiers). Parmi les non-fermiers, l'augmentation du SES manifestait une corrélation négative avec l'apparentement et la consanguinité de chacun des membres de ce couple; au contraire, chez les familles fermières, l'augmentation du statut social allait de pair avec l'augmentation de l'apparentement et de la consanguinité. Selon l'auteur, l'explication de ce phénomène serait que les familles de fermiers à l'aise auraient tout à gagner de l'organisation d'un mariage entre apparentés car, de cette manière, les terres restent dans la famille (du côté paternel). De plus, le lien entre la consanguinité paternelle et le statut social des fermiers tend d'ailleurs à confirmer cette hypothèse.

### 1.1.1 Tout n'est pas si clair...

Certains des résultats obtenus par Shull ne sont pas explicables. En effet, dans les familles non-fermières, la consanguinité de la mère apparaît aussi responsable de la mortalité de ses enfants que son apparentement à son conjoint. Cependant, chez les familles fermières, la consanguinité maternelle a un effet protecteur et semble diminuer les risques de mortalité de la progéniture (Shull et al., 1970a). Ces résultats ne sont pas significatifs, mais l'auteur s'y attarde tout de même et souligne que ce qu'il observe peut être dû au hasard mais aussi à des mécanismes biologiques encore non-identifiés. Les résultats sont cependant très clairs en ce qui concerne l'augmentation de mortalité de la progéniture due à l'apparentement des parents. Le lien avec la consanguinité maternelle est plus difficile à interpréter.

### 1.1.2 La compensation reproductive

Sans considération particulière pour l'activité professionnelle, Shull a aussi démontré la présence de compensation reproductive chez les couples bouddhistes apparentés. Cette compensation reproductive est caractérisée par une augmentation de la mortalité pré-reproductive avec la consanguinité de l'enfant mais aussi par une augmentation des grossesses et des naissances avec l'apparentement des parents. Cette fécondité un peu plus élevée est suffisante pour ramener le nombre d'enfants vivants jusqu'à l'âge adulte à un effectif comparable à celui des enfants issus de couples non-apparentés qui arrivent à l'âge adulte. Ce phénomène de compensation ralentirait l'élimination des gènes récessifs délétères dans la population (Shull et al., 1970b) mais aussi, par la même occasion, les effets de la dépression consanguine dans une population.

### 1.1.3 Les conclusions prudentes

Dans le cadre de son étude sur une population du Pakistan, Bittles explique les effets de la consanguinité et des mariages entre apparentés d'une façon globale en rappelant les principales motivations des diverses sociétés qui ont favorisé ce type de mariage. Il explique entre autres que les traditions de mariages entre apparentés pratiquées depuis longtemps dans plusieurs pays orientaux s'appuient sur 1) des raisons économiques (possession de terres familiales), 2) des négociations pré-nuptiales plus faciles vu les intérêts communs aux deux partis et, (aussi surprenant que cela puisse paraître) 3) des raisons de santé puisqu'on savait à quoi s'attendre de ce nouveau conjoint; on ne craignait pas de voir apparaître une maladie inconnue. En ce qui concerne son étude sur une population du Pakistan, Bittles (1994) a trouvé que la mortalité postnatale était significativement plus élevée chez les enfants consanguins.

Globalement, Bittles demeure prudent face à ses conclusions sur l'effet de l'apparentement ainsi que de la consanguinité des parents sur la mortalité infantile et il amène plusieurs arguments pour appuyer sa position. Premièrement, il affirme que, même en l'absence de consanguinité proche, la concentration d'un gène délétère peut subir des variations de fréquences dues à son histoire évolutive. Pour cette raison et aussi parce que la plupart des études sur les effets de la consanguinité se font dans des régions moins développées (ce qui amène une difficulté dans le diagnostic de la mort), Bittles souligne la difficulté de mesurer l'impact d'une cause de mortalité et d'en déterminer la nature, génétique ou non).

Dans une étude réalisée en 1989, Shami, Schmitt et Bittles avaient également remarqué une fécondité différentielle favorisant les couples apparentés mais malgré ces résultats tendant à démontrer la présence de compensation reproductive, aucune étude plus poussée n'a été effectuée en ce sens. Bittles mentionne qu'il serait nécessaire de vérifier si cette fécondité différentielle est le

fruit d'une plus grande compatibilité sociale ou biologique. Finalement, Bittles (1994) affirme que dans les sociétés où l'on pratique une tradition de mariages entre apparentés, la consanguinité d'un individu ainsi que l'apparentement d'un couple sont sous-estimés lorsqu'on les calcule sur seulement une, deux ou trois générations. Le niveau d'homozygotie du pool génique de telles populations est donc aussi sous-estimé. Vu le manque de profondeur génétique, Bittles n'a toutefois utilisé qu'une seule génération de profondeur pour calculer la consanguinité de la progéniture (apparentement des parents) et les unions étaient dites apparentées si elles liaient des individus cousins du second degré ou plus proches.

D'autres études menées au Pakistan et appuyées, entre autres, sur des régressions logistiques (Grant et Bittles, 1997) démontrent les effets associés de l'apparentement des parents et de plusieurs variables démographiques et socio-économiques significatives dans les modèles sur la mortalité infantile (telles que le lieu de résidence, la qualité de vie, l'éducation maternelle, l'âge de la mère à la naissance, les intervalles entre les naissances, l'ordre de naissance et le sexe de chaque enfant). Dans ces régressions, on inclut ou non les facteurs génétiques afin de pouvoir établir une comparaison entre les degrés de signification des modèles. On met en évidence la valeur significative de la consanguinité par rapport à l'augmentation de la mortalité infantile postnéonatale (2-12 mois) et avant 5 ans. Les intervalles entre les naissances ainsi que l'éducation de la mère sont cependant les deux variables les plus influentes sur la mortalité infantile.

## **1.2 La consanguinité, un facteur neutre?**

Selon les populations ou les sous-groupes étudiés, d'autres auteurs n'ont trouvé aucun impact significatif de la consanguinité sur la mortalité infantile. Ainsi, les études menées par Ross, Edmond et De Braekeleer (De Braekeleer et Ross, 1991; Edmond et De Braekeleer, 1993a, 1993b) sur la population du Saguenay-Lac-Saint-

Jean (SLSJ) n'ont pas trouvé de mortalité différentielle entre les enfants consanguins (issus de couples apparentés) et les non-consanguins. De Brakeleer et al. ont utilisé les données des fournies par les dispenses pour leurs analyses. Leurs résultats indiquaient que ce n'est pas le haut niveau de consanguinité moyenne de la population qui est responsable du taux anormal de maladies autosomales récessives puisque le coefficient de consanguinité calculé au SLSJ est similaire à celui de plusieurs pays européens à la même époque (1842-1971). En fait, la consanguinité et l'apparentement ne seraient, selon leurs résultats, que des facteurs mineurs. L'apparentement ne serait responsable que de 1% de la mortalité pré-reproductive (Edmond et De Brakeleer, 1993). Ils n'ont pas non plus relevé de différences significatives dans la fécondité des couples apparentés ou non.

### **1.3 La consanguinité liée à l'endogamie, un moyen d'éliminer les gènes délétères?**

Un type différent de résultats a été obtenu dans d'autres études sur l'effet de l'apparentement biologique et de la consanguinité dans des populations endogames. Dronamraju (1963) qui a étudié une population indienne à tradition endogame, a observé que la plupart des maladies liées à un problème génétique n'étaient pas plus fréquentes dans les familles présentant un fort taux d'endogamie que pour la moyenne de la population. Il a émis l'hypothèse que, dans cette région de l'Inde, de nombreux gènes délétères récessifs auraient pu être éliminés par simple sélection naturelle au fil des générations à cause de l'endogamie qui persiste depuis des siècles. Les personnes porteuses de gènes délétères à double dose ne vivent pas jusqu'à l'âge de reproduction, donc ces gènes sont progressivement éliminés du pool génique de la population.

#### 1.4 Le débat est ouvert!

La controverse quant aux effets de la consanguinité sur la descendance est encore très actuelle car personne n'a pu apporter l'argument final pour expliquer ces résultats si différents. Il est vrai que la dynamique de la génétique des populations est plutôt complexe et que l'évolution biologique varie selon les populations. De plus, plusieurs facteurs peuvent influencer la mortalité infantile et ainsi biaiser les résultats obtenus. Parmi ces facteurs, nous retrouvons des variables démographiques telles que l'âge de la mère au mariage, l'âge de la mère lors des naissances, les intervalles entre les naissances, l'écart d'âge entre les parents, la gémellité, la situation socio-économique de la population à l'étude etc. Chaque auteur a fait sa propre sélection des facteurs qu'il juge comme influents pour la mortalité infantile et les résultats doivent toujours être interprétés en fonction des variables et des méthodes utilisées.

Un bon exemple de recherche basée sur des variables démographiques est le travail d'Ann W. Brittain (1992) qui étudie la population endogame de Saint-Barthélémy dans les Antilles entre les années 1878 et 1976. Ses résultats démontrent qu'une grossesse multiple (jumeaux, triplés etc.) fait encourir plus de risques de mortalité à la mère et aux enfants et que pour les naissances de rang élevé, la mortalité de l'enfant précédent multiplie par 1.65 les risques de mortalité d'un enfant. Bien qu'elle n'utilise que des variables démographiques dans ses analyses, l'auteure émet l'hypothèse que ces risques amplifiés de mortalité ont des causes sociales et/ou génétiques. L'âge de la mère influence aussi de façon significative le facteur de risque de mortalité infantile. Brittain souligne aussi qu'une autre grossesse dans la première année de naissance d'un enfant favorisait la survie de ce dernier. Ce résultat paraît surprenant mais il peut être attribuable à la bonne santé de la mère qui influence les chances de survie de l'enfant (elle doit être en bonne santé pour assumer une deuxième grossesse). En opposition, on a observé au Québec que les intervalles entre les naissances avaient une influence

sur le poids de l'enfant à la naissance. En effet, lorsque l'intervalle entre un accouchement et une grossesse est inférieur à une année, le nombre de bébés de petit poids augmente à l'issue de la deuxième grossesse. Ces enfants sont plus fragiles et donc plus exposés à mourir en bas âge (Lepage et al., 1989).

Brittain a aussi mis en évidence que l'arrivée d'une présence médicale permanente dans l'île en 1934 a eu un effet significativement négatif sur la probabilité de mourir avant l'âge d'un an. Malgré le détail de sa démarche et le nombre des variables explicatives de son modèle, Brittain déplore le manque de capacité prédictive de ses régressions car elle n'est jamais supérieure à 8%. Rappelons toutefois que Brittain, vu la nature de sa base de données, n'utilise pas de variables d'ordre génétique ou biologique dans ses analyses.

## **1.5 Notions d'histocompatibilité**

### **1.5.1 Le MHC chez les populations humaines**

Parallèlement à ces études, d'autres chercheurs travaillent sur un plan beaucoup plus appliqué pour comprendre les mécanismes biologiques qui régissent la survie d'un fœtus ainsi que ses interactions avec sa mère sur le plan immunitaire.

Les réponses immunitaires et les interactions entre cellules humaines sont principalement contrôlées par le complexe principal d'histocompatibilité (en anglais: major histocompatibility complex ou MHC). Ce complexe se résume en un groupe de gènes situés sur le chromosome 6. Les gènes encodés par cette région correspondent aux antigènes leucocytaires humains (human leukocytes antigens, ou HLA) et se traduisent par des glycoprotéines à la surface des cellules qui enclenchent les réactions immunitaires.

Le MHC est caractérisé par un très haut taux de polymorphisme pour chacun des gènes qui le composent. Il existe donc pour chaque locus, un grand nombre d'allèles si bien que dans une population, aucun allèle ne dépasse la fréquence de 30% et que la grande majorité des individus sont hétérozygotes pour chaque locus. De plus, le taux de recombinaison par crossing-over lors de la méiose (formation des gamètes) est très bas dans cette région. Ceci a pour conséquence que les allèles du MHC sont habituellement transmis d'un bloc à la génération filiale sans qu'il y ait eu recombinaison entre les brins D'ADN originaux. De génération en génération, des séquences d'allèles sont donc maintenues de façon caractéristique pour former une espèce d'unité appelée haplotype (Alberts et Ober, 1993). Puisque les gènes du MHC sont transmis en séquences de gènes dépendants, un individu homozygote pour le premier locus l'est probablement aussi pour le deuxième, le troisième et ainsi de suite.

Le MHC se divise en trois régions qui codent pour les classes de protéines I, II et III. Chacune de ces classes a une fonction immunitaire distincte. La classe II est celle qui entraîne les plus fortes réactions immunitaires lors des transplantations d'organes. La classe I est la plus répandue et se retrouve à la surface de presque toutes les cellules nucléées (HLA-A, HLA-B et HLA-C). Il existe une autre protéine HLA de classe I, HLA-G, qui est la seule à être exprimée dans les tissus foetaux, à l'interface mère-foetus. On ne retrouve pas cette protéine ailleurs dans l'organisme et, contrairement à toutes les autres protéines HLA, HLA-G présente très peu de polymorphisme. C'est pour cette raison que l'on suppose (Ober, 1998) que les HLA-G limitent la prolifération de lymphocytes T dans la région utérine, empêchant ainsi le système immunitaire maternel d'attaquer le foetus.

Ober (1998) cite à l'appui des exemples de populations afro-américaines chez qui on observe une mutation du gène HLA-G: l'allèle nul. Cet allèle inhibe l'expression de HLA-G à l'interface mère-foetus chez les homozygotes. L'absence de HLA-G permet le passage de lymphocytes T de la mère vers les tissus foetaux. Cet allèle

nul a pu être maintenu dans le pool en raison de fortes pressions de maladies infectieuses omniprésentes dans l'histoire de ce type de population (Ober et Aldrich, 1997). Le passage de cellules immunitaires maternelles vers les tissus foetaux est, dans ce cas, bénéfique pour prévenir le fort risque d'infections intra-utérines (Ober, 1998). Le caractère conservateur des HLA-G en opposition au caractère très polymorphique du MHC en général démontre que les contraintes de sélection pour les HLA-G ont été différentes de celles exercées sur le reste du MHC.

Plusieurs théories ont tenté d'expliquer le maintien du polymorphisme du MHC. Toutes ces théories s'appuient sur la sélection naturelle mais le mécanisme par lequel s'exerce cette sélection reste encore incertain. Cependant, la déficience d'individus homozygotes pour ces loci renforce l'hypothèse de sélection favorisant les hétérozygotes et donc le polymorphisme. Alberts et Ober (1993) résument les principaux modes de sélection proposés pour expliquer le phénomène. Tout d'abord, plusieurs, tels Bodmer (1972), Hugues et Nei (1989), Nei et Hugues (1991), ont soutenu que le maintien de la variabilité était dû à la présence d'agents pathogènes. En présence d'un pathogène, le fait d'être hétérozygote pour ces loci présente un net avantage puisque l'individu possède deux fois plus d'antigènes différents qu'un homozygote. Sa capacité de défense est donc plus grande et le pathogène a plus de difficulté à s'adapter.

D'autres études proposent une sélection due à un agent non-pathogène. Dans ces cas, la variabilité du MHC serait maintenue par les relations mère-foetus, les gènes du MHC liés à des gènes létaux ainsi que par les effets néfastes de la formation de couples dont les membres sont trop semblables (Alberts et Ober, 1993). Lors d'une grossesse, des lymphocytes foetaux passent la barrière placentaire et se retrouvent dans la circulation de la mère (Ober, 1998). Ces lymphocytes expriment les HLA hérités par le père et ainsi entraînent une réaction immunitaire (qui semble s'amplifier d'une grossesse à l'autre) chez la mère. Cette réaction immunitaire et la reconnaissance d'HLA paternels par le système de la mère entraînent la production de

facteurs bloquant les anticorps libérés contre les HLA paternels et empêchent ainsi le rejet de l'embryon. Plusieurs études semblent prouver que le partage d'HLA entre le père et la mère (et donc la production d'embryons homozygotes trop semblables à la mère) entraînerait une sous-réactivité de la mère aux lymphocytes foetaux. On dit à ce moment que la mère et le fœtus sont compatibles. La production de facteurs inhibant les anticorps de la mère contre son fœtus ne se produirait pas normalement et ces femmes auraient significativement plus de fausses couches que les femmes des couples ne partageant pas d'HLA (Beer et al., 1981). D'ailleurs, selon cette même étude, plus les couples partagent de HLA, plus leurs risques de perdre des enfants avant terme est élevé. Les auteurs s'entendent donc pour dire qu'une réaction immunitaire de la mère contre des HLA foetaux hérités par le père est non seulement normale mais bien souhaitable pour le bon déroulement d'une grossesse et la tolérance du fœtus par la mère (Alberts et Ober, 1993; Beer et al., 1981; Ober, 1998). Les fœtus mâles et femelles ne semblent pas réagir de la même manière face à la compatibilité avec la mère. En effet, lorsque les parents partagent des antigènes, parce qu'ils sont apparentés, on observe une majorité de naissance féminine. Pourtant, le rapport de masculinité normal dans les populations humaines se caractérise par une majorité de naissances masculines (Astolfi et al., 1990; Adinolfi, Polani et Crolla, 1985).

#### 1.5.2 Essais sur des populations de souris:

Des recherches ont été effectuées sur des souris. Après avoir croisé deux lignées homozygotes (à deux loci) pour obtenir une F1 (génération filiale) hétérozygote aux loci contrôlés, le croisement d'individus mâles de la F1 avec des femelles de la lignée parentale ainsi que le croisement de femelles F1 avec des mâles de la lignée parentale ont amené des résultats intéressants. Premièrement, le croisement de femelles homozygotes avec des mâles apparentés (allèles en commun) produit des portées avec un surplus significatif d'hétérozygotes (le rapport attendu entre homozygotes et hétérozygotes dans la portées était 1:1 et le résultat obtenu démontrait un surplus significatif de rejetons hétérozygotes). Le surplus d'hétérozygotes augmente aussi

avec le rang de la portée. Chez les femelles homozygotes, il semble également y avoir une sous-natalité de rejetons mâles (Macklin, 1951). Les auteurs associent ce fait à la possible existence d'un gène lié au chromosome Y qui influencerait la production des antigènes HLA. Ceci pourrait donc avoir un effet multiplicatif et délétère sur la progéniture mâle en entraînant le rejet de ces foetus trop compatibles. Chez les femelles hétérozygotes, aucun effet significatif sur les proportions d'homozygotes versus d'hétérozygotes ainsi que mâle versus femelle n'a été noté lorsque celles-ci étaient croisées avec des mâles apparentés (Hull, 1969).

### 1.5.3 Essais sur des populations humaines :

Les Hutterites constituent une population de descendance européenne reconnue et très étudiée en raison de sa forte endogamie. Comme il a déjà été mentionné, le MHC possède habituellement une très grande variabilité dans les populations à tradition exogame. Cependant, chez les Hutterites, le nombre d'haplotypes pour les HLA est très limité. Pour cette raison, on y retrouve une plus grande proportion de couples qui portent le même haplotype que dans une population exogame (Ober et al., 1992; Ober, 1995). Par conséquent, le nombre de grossesses où les mères sont compatibles avec leur foetus est plus grand. Les résultats d'une étude prospective démontrent que les couples portant le même haplotype pour les HLA ont significativement plus de pertes de foetus que les couples portant des haplotypes différents. Pour un couple, le fait d'avoir le même haplotype signifie que la progéniture a de fortes chances d'être compatible avec la mère (donc à risque). Il est également démontré que le risque de perte foetale est accru par la compatibilité pour le locus des HLA-B. Toutefois, les résultats ne permettent pas encore de dire si ce risque est lié directement à ce locus, à un gène lié ou à l'ensemble de l'haplotype. L'interprétation repose sur l'allégation suivante: s'il est désavantageux pour le foetus d'être compatible avec sa mère, nous devrions retrouver une sous-représentation d'haplotypes compatibles dans la progéniture vivante et une sur-représentation d'haplotypes incompatibles. À cause de la petitesse de l'échantillon, les différences ne peuvent être considérées comme

significatives. Toutefois, un résultat semble suggérer que la reconnaissance du non-soi se fait par le fœtus et non par la mère. En effet, chez des femmes hétérozygotes, les grossesses sont regroupées en trois catégories: fœtus compatible homozygote, fœtus compatible hétérozygote (identique à la mère) et fœtus incompatible et hétérozygote (l'allèle donnée par le père est différent de ceux que possède la mère; le fœtus est donc différent de sa mère). La mère peut théoriquement considérer les deux premières catégories comme "soi" et compatibles puisque ces fœtus n'expriment que des antigènes identiques aux siens. On aurait pu prévoir la sous-représentation de ces deux catégories et la sur-représentation de la troisième. Ober et ses collaborateurs ont cependant constaté que seuls les fœtus hétérozygotes et identiques à la mère étaient sous-représentés. Les enfants homozygotes compatibles et hétérozygotes incompatibles représentent au contraire une plus grande proportion de la progéniture. On amène l'hypothèse que le fœtus homozygote quoique compatible peut discerner les HLA spécifiques à sa mère et de cette manière poursuivre son développement (Ober et al., 1998). Il s'agirait donc d'une reconnaissance bilatérale.

## CHAPITRE II

### HYPOTHÈSE DE TRAVAIL

#### 2.1 Notions de biologie évolutive appliquées

Comme il a été mentionné auparavant, dans une population endogame, un apparentement entre les individus s'établit au fil des générations. Cet apparentement est issu des nombreux liens unissant les familles les unes aux autres à l'intérieur de la population. Vu l'apparentement moyen élevé entre deux individus pris au hasard dans une telle population, il est normal de voir un coefficient de consanguinité moyen de la population plus élevé que dans une population panmictique. En étant apparentés et peut-être eux-mêmes consanguins, les couples ont plus de chances d'avoir dans leur progéniture des enfants homozygotes pour des gènes hérités par descendance d'un ancêtre commun à leurs deux parents. Certains de ces gènes font partie du fardeau génétique de la population et sont récessifs et létaux à double dose. Le surplus d'homozygotes engendré par le fort niveau de consanguinité de la population expose donc à la sélection une plus forte proportion de ces gènes récessifs désavantageux et les individus homozygotes ainsi défavorisés ne peuvent contribuer aux générations suivantes. Ces gènes trop délétères voient donc leurs proportions diminuer dans la population puisque seulement les individus hétérozygotes pour ces loci sont viables. Dans ces circonstances, on voit apparaître une dépression dite consanguine du pool génique de la population lorsque ces gènes commencent à être éliminés. Cette élimination n'est pas constante. Lorsqu'un allèle récessif devient rare (caché sous forme presque exclusivement hétérozygote) son élimination devient plus lente (Futuyma, 1986). Ceci correspond au "fardeau de substitution" ou au "coût de remplacer un allèle par un autre". Dans un de ses ouvrages, Haldane (1990) a calculé que pour remplacer complètement un allèle par un autre, le nombre de morts requis est d'environ cinq à quinze fois l'effectif de la population et que le processus complet devrait s'effectuer sur environ trois cents générations pour éviter que la population ne soit décimée.

De plus, dans un de leurs articles, Shull et Neel (1972) ont démontré que la compensation reproductive ralentit aussi le remplacement allélique puisque les couples qui perdent un enfant homozygote récessif létal (phénotype "aa" dont la proportion dans la population serait normalement de  $q^2$  si les individus ne mouraient pas et où  $q$ =fréquence de l'allèle récessif) remplacent cet enfant par un homozygote dominant (phénotype "AA" dont la proportion dans la population est de  $p^2$  et où  $p$ =fréquence de l'allèle dominant) ou hétérozygote (phénotype "Aa" dont la proportion dans la population est de  $2pq$ ). De cette manière, au lieu d'éliminer les gènes létaux à un taux estimé à  $q^2$ , le taux en présence de compensation reproductive serait de  $2/3q^2$  puisque les  $2/3$  des enfants remplaçant un homozygote récessif mort seront hétérozygotes et donc porteurs de ce gène létal pour la prochaine génération. C'est donc en raison de la lenteur évolutive du pool génique ainsi que des mutations nouvelles qui se fixent dans les populations que le polymorphisme se maintient au fil des générations.

## 2.2 Élaboration des hypothèses

C'est au fil des études touchant la biologie évolutive des populations endogames humaines qu'on a pu observer les effets de l'apparement des conjoints sur le destin de leur progéniture (mortalité infantile, anomalies). Dans environ 50% des cas, les résultats des recherches indiquaient un lien entre la mortalité infantile et l'apparement des conjoints. En ce qui concerne l'autre 50% des études, aucun lien significatif entre l'apparement des conjoints et la mortalité de leur progéniture n'a pu être démontré. Les méthodes utilisées pour les analyses, la profondeur généalogique dont disposaient les chercheurs ainsi que les caractéristiques intrinsèques des populations étudiées ont certainement un impact sur les résultats obtenus ce qui nous empêche de pouvoir établir une comparaison directe entre leurs conclusions respectives. De plus, l'objet principal de la majorité de ces recherches était la plupart du temps l'apparement des conjoints, estimé sur peu de générations et sans un regard très approfondi sur l'histoire généalogique de ceux-ci. L'apparement réel des

conjoints ainsi que leur consanguinité respective s'en trouvaient donc sous-évalués sinon ignorés.

En effet, les études visant à dégager les causes de mortalité infantile dans des populations endogames sont axées sur l'apparentement des parents et ne traitent de consanguinité que de façon superficielle.

De plus, pour plusieurs auteurs, certains résultats demeurent ambigus. L'effet protecteur de la consanguinité maternelle pour la survie de la progéniture mentionné par Shull et al. (1970a) sans qu'il puisse discuter le phénomène en est un bon exemple. Dans ces cas, les auteurs affirment qu'une investigation plus poussée est nécessaire pour donner une explication logique de ces cas contradictoires.

À la lumière de toutes ces informations, cette étude permettra peut-être de répondre à l'interrogation de Shull sur l'effet protecteur de la consanguinité de la mère sur la survie de sa progéniture. En effet, étant donnée la grande profondeur généalogique des registres de populations dont dispose le laboratoire du Dr Mayer au sein duquel a été mené le présent travail, il sera possible de bien estimer l'apparentement, la consanguinité proche et la consanguinité éloignée (définies dans la section 4.6) des individus sélectionnés dans la population retenue pour l'étude.

**L'hypothèse principale s'énonce donc comme suit:**

**Dans une population d'effectif restreint et pratiquant l'endogamie depuis plusieurs générations, les mères consanguines (donc fortement homozygotes) sont le résultat d'une forte exposition à la sélection génique au fil des générations. Si ces femmes ont pu survivre et se reproduire, elles sont porteuses d'allèles avantageux sélectionnés et à l'état homozygote. Leur génôme épuré leur confère un avantage reproductif qui peut compenser les effets néfastes (avortements et fausses couches) encourus par le partage, avec un conjoint apparenté, de plusieurs**

antigènes liés à l'histocompatibilité. La transmission à la progéniture de ces gènes sélectionnés diminue, en présence d'un conjoint apparenté, le risque de retrouver chez les descendants des gènes récessifs létaux à l'état homozygote. Par conséquent, cette progéniture est plus viable in utero et plus apte à se rendre à l'âge adulte pour contribuer aux générations suivantes.

Une hypothèse secondaire est que le rapport de masculinité anormal observé dans la population retenue est lié à la survie préférentielle des hétérozygotes sur les homozygotes. Les gènes codant pour les antigènes HLA sont reconnus pour être hautement polymorphiques. Le partage de ces gènes chez les conjoints apparentés reflète donc aussi la probabilité de partager plusieurs allèles par descendance d'un ancêtre commun et donc d'avoir une progéniture hautement homozygote. Vu leurs chromosomes sexuels disparates, les garçons ont moins de chances que les filles d'exprimer des gènes à l'état hétérozygote. Le fait que le fœtus mâle possède un génome trop semblable au génome maternel ainsi que le passage de lymphocytes fœtaux exprimant des antigènes HLA paternels semblables à ceux de la mère peut avoir un effet multiplicatif dans la non-reconnaissance de la grossesse par sous-réaction immunitaire de la mère. Cette sous-réactivité peut aussi être due à un lien entre le MHC et le chromosome Y (comme il a été proposé chez les souris). Donc, chez les mères homozygotes, une plus grande proportion de garçons est perdue dans les premiers moments de la grossesse en raison de la combinaison de plusieurs facteurs qui empêchent la reconnaissance immunitaire du fœtus et donc son maintien. Il est possible que ce soit pour cette raison qu'il naît moins de garçons que de filles.

L'outil principal proposé pour vérifier ces hypothèses dans la présente est la modélisation par régression logistique dont les principes seront expliqués dans la section 4.7.1. À l'aide de cette méthode, il sera possible d'évaluer l'impact des variables génétiques souhaitées et de tenter d'interpréter ces résultats de modélisation dans le contexte socio-démographique de l'Île-aux-Coudres entre 1800 et 1973. La variable dont on tente de cerner les causes de modulations est bien sûr la mortalité de

la progéniture. Dans la section 4.7.2, on présente les différents types de mortalité qui ont été sélectionnés pour vérifier les deux hypothèses citées plus haut. Parmi ces types de mortalité, on retrouve, entre autres, la mortinatalité (enfant mort né) (Péron et Strohmenger). C'est principalement par l'interprétation des modèles obtenus grâce à cette variable que nous espérons répondre à l'hypothèse secondaire concernant le rapport de masculinité et le MHC. En effet, ne disposant pas de données sur le MHC et les relations mère-foetus de la population à l'étude, c'est cette variable qui nous donnera le reflet le plus fidèle de la relation entre la mère et son foetus. Afin de voir l'impact du sexe sur cette relation, des modèles divisés selon le sexe seront présentés. Les détails de la démarche démographique et statistique mise en oeuvre pour tirer des conclusions adéquates sont présentées en détail dans la section méthodologie (chapitre IV).

### **2.3 Études menées par le laboratoire du Dr Francine M. Mayer sur la consanguinité**

Comme il a été mentionné précédemment, la plupart des études menées sur la consanguinité et l'apparentement basent leurs analyses sur une connaissance généalogique peu profonde (quelques générations). Leurs résultats concernant la consanguinité ne constituent alors qu'une estimation des effets de la consanguinité proche. La consanguinité proche d'un individu est le résultat d'un apparentement de ses deux parents par un ou des ancêtres peu éloignés dans la généalogie ascendante de l'individu. Le laboratoire de Francine M. Mayer s'intéresse depuis longtemps aux effets de l'apparentement et de la consanguinité dans les populations humaines endogames. Grâce à une connaissance des généalogies qui peut remonter jusqu'à quinze générations pour les individus de la banque de données de l'Île-aux-Coudres, et des généalogies semblables pour la population de Saint Barthélemy, une île antillaise, l'apparentement biologique de deux individus peut facilement être calculé à partir de toutes les chaînes de leur généalogie les reliant à un ancêtre commun. Ainsi, on peut connaître la consanguinité totale de chaque individu et, de plus, on peut

caractériser cette consanguinité en la divisant en consanguinité proche et en consanguinité éloignée selon les besoins de l'étude.

Plusieurs études menées au sein du laboratoire du Dr Francine M. Mayer ont porté directement sur la consanguinité. Ainsi, des travaux qui mettent en évidence l'impact de la consanguinité des mères sur la mortalité infantile serviront de base à mes hypothèses. Ceux-ci ont été effectués par Katia Allain en 1996 (sur la population de Saint Barthélémy ) ainsi que par Mireille Boisvert sur l'Île-aux-Coudres (1992, Boisvert et Mayer en 1994 ainsi que des données en voie de publication). Ces dernières faisaient suite à ses travaux sur l'impact de l'apparentement des conjoints sur la mortalité infantile. Ils consistaient en une simple analyse statistique tendant à mettre en relation la mortalité infantile et la consanguinité maternelle au sein des couples dont les conjoints étaient apparentés versus non-apparentés.

Par la suite, une étude préalable sur les facteurs influençant la mortalité infantile à l'Île-aux-Coudres (Maheu, 1997) a été effectuée. La consanguinité maternelle est ressortie comme significativement protectrice contre la mortalité de la progéniture féminine. D'autres études menées dans une optique très différente aideront à amener une explication plus biologique des résultats observés. Il s'agit des recherches menées par d'autres chercheurs sur l'histocompatibilité et les relations mère-foetus dont il a été question dans l'état des connaissances.

#### **2.4 Implications scientifiques de l'étude**

Afin d'aller plus loin dans la démarche de Madame Boisvert, cette étude servira à caractériser l'impact différentiel ou non de la consanguinité maternelle proche et de la consanguinité maternelle éloignée sur la mortalité infantile. Les résultats et les conclusions qui suivront cette étude pourront alors servir de pistes pour d'autres travaux dont les champs d'actions incluront un volet de génétique appliquée et de biologie

moléculaire. Ceux-ci pourront alors préciser les mécanismes biologiques que nous ne pouvons vérifier ici étant donné l'approche évolutive et le contexte théorique de la présente étude.

Certaines populations endogames actuelles ont des problèmes similaires à ceux observés chez les Hutterites. Parmi celles-ci, on retrouve certaines populations caucasiennes, taiwannaises, brésiliennes (isolat amérindien) et les Kel Kummer (Gill, 1997). D'autres études telles que celles réalisées sur les Hutterites (Ober, 1995, Ober et al., 1998) feront probablement suite aux recherches portant sur des populations de rongeurs (Hull, 1969) afin de bien comprendre les mécanismes qui régissent les échanges mère-foetus pendant la grossesse.

## CHAPITRE III

### POPULATION À L'ÉTUDE

#### 3.1 La population de l'Île-aux-Coudres

La population de l'Île-aux-Coudres a été le sujet de plusieurs études qui ont montré son caractère très endogame et consanguin. En effet, Pierre Philippe (1969), a évalué que les coefficients moyens de consanguinité et d'apparentement étaient respectivement de 0.015 et 0.012. En comparaison, les estimations de Philippe à partir des dispenses religieuses de l'archidiocèse de Québec pour la période 1955-65, révélaient seulement un coefficient de consanguinité de 0.0009 (Laberge 1967). De plus, Philippe et Gomila (1972) ont démontré qu'entre les années 1939 et 1949, 90% des mariages étaient entre conjoints apparentés par des liens éloignés. Cette forte endogamie observée à l'Île-aux-Coudres est le résultat d'une organisation sociale appuyé sur l'entraide entre parents. Les pères de famille cèdent leurs terres et la maison à leur fils aîné en priorité et le réseau social de partage des tâches inclut les voisins qui sont souvent des parents. On voit donc naître des réseaux de soutien familiaux très complexes. Pour les travaux qui demandent une main-d'oeuvre nombreuse, tous les parents participent en collaboration selon leurs capacités (les hommes font de la pêche, du déboisement et les femmes et les jeunes enfants, du jardinage) sous les ordres d'un aîné commun (le père ou le grand-père). Les terres se font cependant de plus en plus rares puisque la population insulaire augmente. À cause de la surpopulation, un mouvement d'émigration s'entame en 1790 et se poursuit jusqu'en 1972 (Le Querrec, 1981). Ce mouvement, principalement unilatéral, n'est toutefois pas suffisant et les ressources continuent de diminuer. Plusieurs hommes se retrouvent donc sans terres et sont appelés « emplacitaires » (n'ont pas de terres arables).

Les premières concessions de terres à l'île eurent lieu en 1728, date où dix titres furent cédés (Martin, 1957). Vu la tradition de légation de terres de père en fils, il est normal de constater une tradition de mariage entre apparentés du côté paternel afin de préserver les avoirs de la famille (Le Querrec, 1981). On voyait donc souvent un jeune homme marier la fille de la cousine de son père. Étant donné que la colonisation de l'Île-aux-Coudres s'est faite en majeure partie par des gens originaires de Charlevoix, il est raisonnable de présumer que le même type de mariage se pratiquait à l'Île-aux-Coudres. On remarque toutefois que les individus évitent de se marier avec un apparenté proche, tel un cousin germain (Philippe, 1969).

En 1954, on dénombrait 257 foyers à l'Île-aux-Coudres et seulement 14 de ces couples n'étaient pas constitués de deux personnes originaires de l'Île, ce qui appuie le fait que la migration s'est faite à sens unique depuis 1871 et qui explique la forte endogamie de la population. En fait, il est remarquable de constater qu'en 1954, le nombre de patronymes de la population n'a pas beaucoup augmenté depuis 1871; ainsi, huit des patronymes présents en 1871 et représentant 75% de la population à cette époque étaient toujours présents en 1954 et comptaient alors pour 80% de cette population (Martin, 1957).

Martin a mis en évidence le taux de mortalité infantile très élevé observé jusqu'à 1930 à l'Île-aux-Coudres. Il attribue cette mortalité à divers facteurs tels que la consanguinité et l'absence de médecin résident. En effet, une diminution de la mortalité a été observée après l'établissement d'un médecin en 1930 (Martin, 1957). Cependant, encore aujourd'hui, avec un bon accès à des soins médicaux, la mortalité infantile dans Charlevoix ouest (en admettant que l'Île-aux-Coudres corresponde au modèle de Charlevoix ouest) est remarquablement plus élevée que la moyenne québécoise et il en est de même pour les taux d'hospitalisation d'enfants atteints de certaines maladies à caractère héréditaire se manifestant dans cette région dans des proportions anormalement élevées. On précise même

dans un rapport de santé (Rochette, Drouin et Tremblay, Rapport du DSC Enfant Jésus, 1986) que la mortalité entre 0 et 4 ans est 1,8 fois plus élevée dans Charlevoix ouest que dans le reste du Québec et qu'on y retrouve des taux anormalement élevés de maladies congénitales. Pendant la période périnatale (0 à 27 jours), les fillettes de Charlevoix ouest seraient significativement plus hospitalisées que dans le reste de la Province. Finalement, on observe aussi une proportion plus grande de maladies endocriniennes, métaboliques et respiratoires chez les jeunes garçons de Charlevoix ouest que pour l'ensemble du Québec.

Certaines maladies génétiques sont également reconnues pour avoir une prévalence plus élevée que la normale dans la région de Charlevoix et du Saguenay. On retrouve entre autres la tyrosinémie héréditaire qui apparaît dans les six premiers mois de vie et cause le décès infantile si un régime strict n'est pas suivi, le rachitisme vitamino-dépendant (autosomal, récessif) qui apparaît entre 1 et 2 ans, l'atrésie intestinale et l'acidose métabolique qui apparaissent dans la première année et finalement, le syndrome de diarrhée sécrétoire et fibrose hépatique congénitale qui apparaît entre 6 et 20 mois. Toutes ces maladies peuvent être la cause de décès dans la ou les premières années de vie de l'enfant atteint (De Brakeleer, 1991).

Dans la majorité des populations humaines, le rapport de masculinité à la naissance (qui correspond au nombre moyen de naissances masculines pour une naissance féminine) est supérieur à 1, ce qui indique qu'il naît plus de garçons que de filles. L'équilibre entre les quantités d'hommes et de femmes se fait par la suite car on observe également plus de mortalité masculine que de mortalité féminine (Vandeschrick, 1994). À l'Île-aux-Coudres, le rapport de masculinité est inversé, c'est-à-dire qu'il y naît plus de filles que de garçons. On y observe cependant plus de mortalité féminine en bas âge et ce, surtout pour la progéniture des couples apparentés (Boisvert, 1992). Étant donné que 90% des mariages à l'Île-aux-

Coudres sont entre apparentés, on peut étendre cette statistique à la majorité de la population.

### 3.2 Le Registre de la population

Le registre de population utilisé pour cette recherche contient les actes d'état civil jumelés entre eux contenus dans les registres religieux des paroisses de l'Île-aux-Coudres (1741 à 1973) et de Baie-St-Paul (1720 à 1740). Ces actes d'état civil ont aussi été jumelés aux données du recensement de 1967 effectué par Pierre Philippe ainsi qu'à certaines informations provenant des recensements nominatifs canadiens (pour la deuxième partie du XIXe siècle). Le Registre de population permet d'établir les biographies des individus ainsi que les fiches de familles et les arbres généalogiques ascendants et descendants de tous les individus de la population pour la période étudiée. Il rend possible l'étude d'une multitude de variables d'ordre démographique caractérisant entre autres l'histoire féconde du couple et le destin de sa progéniture. Grâce à sa profondeur, ce même Registre permet le calcul de variables génétiques telles que l'apparentement entre les conjoints, le coefficient de consanguinité de chacun d'eux ainsi que le coefficient de consanguinité de leur progéniture.

Afin d'en assurer une gestion rapide, le Registre de la population de l'Île-aux-Coudres est administré par le logiciel ANALYPOP, conçu par M. Éric Labelle, technicien informatique du laboratoire de Francine M. Mayer. Le logiciel comprend des modules d'analyse généalogique et démographique qui permettent la sélection de couples selon les critères désirés et le calcul des variables nécessaires à l'étude.

## CHAPITRE IV

### MÉTHODOLOGIE

#### 4.1 Critères de sélection des couples

Puisque le but premier de cette étude est de tenter de modéliser les différents types de mortalité chez la progéniture des couples de l'Île-aux-Coudres et de déterminer l'impact des différents types de consanguinité maternelle (totale, proche et éloignée) sur cette mortalité, la sélection première des sujets de la banque de données s'est faite en utilisant le couple comme unité statistique. Cette méthode a donc permis d'inclure la progéniture totale de tous les couples répondant aux critères énumérés ci-bas.

Voici la liste des critères de sélection des couples pour l'étude à venir. Il est important de spécifier que ces critères sont très semblables à ceux appliqués par Mireille Boisvert lors de son étude effectuée en 1992 .

*1)Le couple doit avoir des enfants (si on veut pouvoir étudier la mortalité de la progéniture).*

*2)La période choisie pour l'étude (1800 à 1940) est la même que celle préalablement utilisée par Boisvert (1992) et ce, pour les mêmes raisons: pouvoir profiter d'une profondeur généalogique assez importante pour pouvoir calculer des coefficients de consanguinité et des degrés d'apparentement s'il y en a. Parfois, on pourrait avoir un coefficient nul à cause d'une profondeur généalogique insuffisante ne permettant pas de voir une consanguinité plus éloignée).*

*3)L'âge des parents étant potentiellement un facteur de risque pour la mortalité infantile, nous excluons les couples dont on ne connaît pas la date de naissance d'un ou des deux parents.*

4) Pour ne pas biaiser les résultats par perte d'informations, les couples qui ont eu des enfants à l'extérieur de l'île (donc ayant émigré) sont exclus.

5) Afin de ne pas compliquer les calculs, nous excluons les femmes ayant contracté plus d'une union car le risque de perdre un premier enfant ne serait pas le même pour le premier enfant issu d'un deuxième mariage s'il y a eu des enfants lors du premier mariage.

6) Le critère suivant sert à contrôler la perte d'information sur des naissances due à la migration. Son application a été suggérée par Henry (1958). Il permet de ne garder que les couples qui n'ont pas migré hors de l'île en se basant sur les intervalles marqués par les naissances puisque la population à l'étude n'utilisait à l'époque aucun moyen contraceptif et était donc soumise à un régime de fertilité naturelle (Philippe, 1973). Le critère de Henry se divise en trois catégories:

L'intervalle protogénésique (intervalle entre le mariage et la première naissance). Les couples dont l'intervalle protogénésique dépassait 48 mois ont été exclus.

L'intervalle intergénéésique (intervalle de temps entre deux naissances). Pour les femmes âgées de moins de 35 ans au milieu de l'intervalle, on excluait les couples dont l'intervalle dépassait 48 mois et pour les femmes de 35 ans ou plus au milieu de l'intervalle, on excluait les couples quand l'intervalle dépassait 84 mois.

L'intervalle pausigénéésique (intervalle de temps entre la dernière naissance et les 45 ans de la mère (âge théorique de la ménopause) ou la fin de l'union avant les 45 ans de la femme (mortalité d'un conjoint). Tous les couples dont l'intervalle pausigénéésique excédait 84 mois ont été exclus.

Cette série de critères a été appliquée sur une banque totalisant 2000 couples pour réduire cet effectif à 308 couples répondant aux préalables à l'étude (voir la figure 4.1 pour les étapes de la sélection).

Il est à noter que les couples ayant eu une ou des naissances gémellaires ont été exclus car l'étude de ces naissances constitue un autre sujet d'étude. De plus, les naissances gémellaires pourraient nuire à l'interprétation des résultats car on ne connaît pas l'interaction de ces deux phénomènes biologiques (naissances gémellaires et impact de la consanguinité sur la mortalité infantile). On sait cependant que la gémellité augmente le risque de mortalité de la mère et de sa progéniture (Bittles, 1992) et peut aussi être due à des facteurs génétiques spécifiques qui, le cas échéant, pourraient compliquer les analyses.

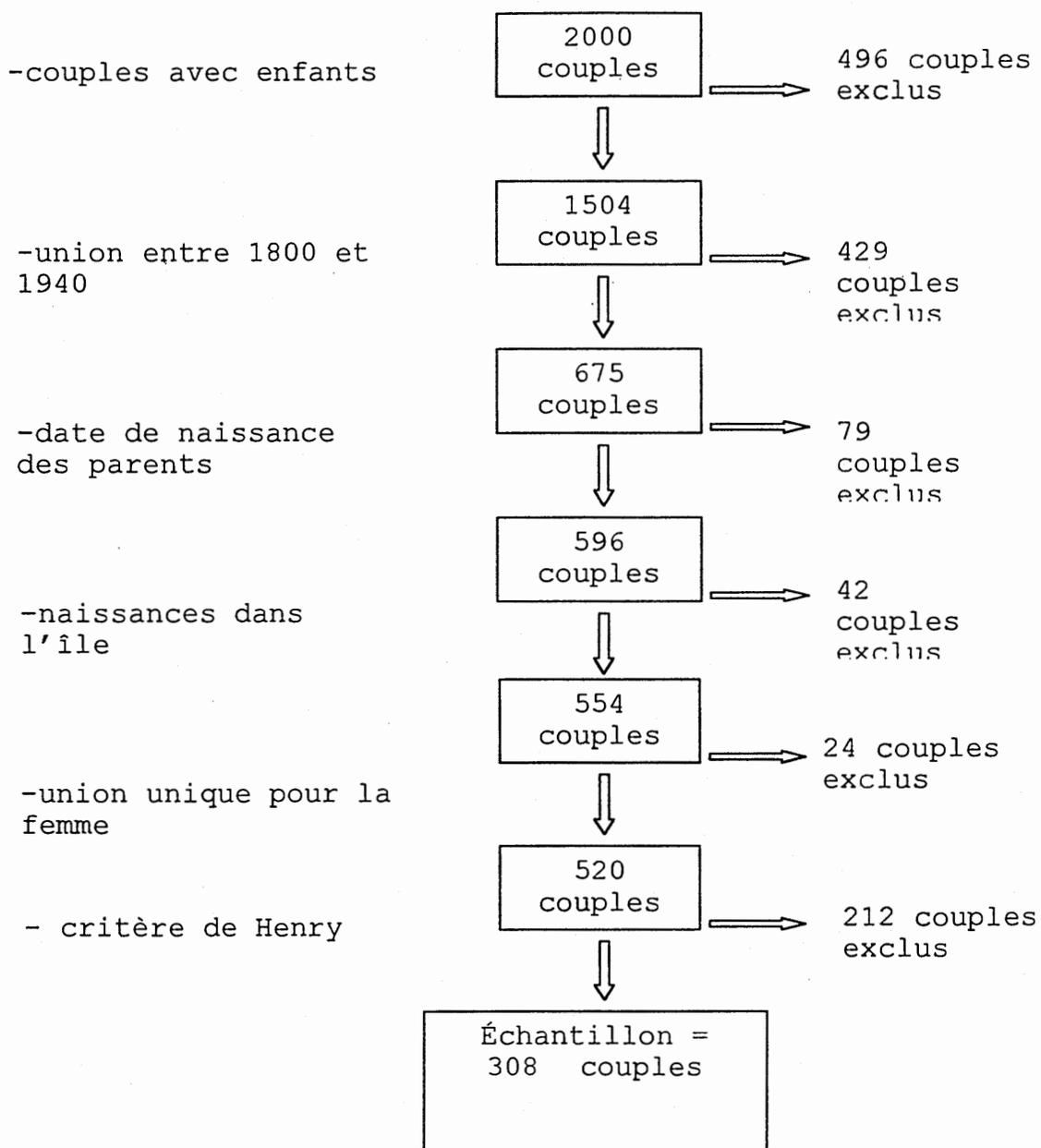


Figure 4.1 : Représentation schématique de l'application des critères de sélection à la base de données informatisée de l'île-aux-Coudres.

Comme il a été mentionné plus haut, les critères d'exclusion applicables à cette étude sont sensiblement les mêmes que ceux appliqués par Boisvert dans ses recherches sur la mortalité infantile à l'Île (1992, Boisvert et Mayer, 1994). Toutefois, Boisvert a appliqué les critères de Henry de façon plus large en proposant une hypothèse de sous-fécondité appuyée sur la vérification dans les recensements de la présence des couples dans l'intervalle entre les naissances. Cette hypothèse permet d'obtenir un échantillon de plus grand effectif mais, bien que sa méthode n'ait entraîné aucun biais majeur, j'appliquerai les critères de Henry de manière stricte pour ne pas alourdir la démarche.

Pour chacun des couples sélectionnés, nous disposons de plusieurs informations. Les variables qui seront retenues sont:

variables démographiques:

- le nombre d'enfants de chaque sexe
- le nombre de décès d'enfants séparé par sexe et par groupes d'âges (la mortalité infantile a été codée selon que l'enfant est mort le premier jour, entre 1 et 6 jours, 7-27 jours, 0-365 jours 1 à 4 ans et 1 à 14 ans)
- l'intervalle intergénésiq ue moyen
- l'intervalle protogénésiq ue ou l'âge de la mère à la première naissance (ces deux variables contiennent des informations similaires)
- l'intervalle pausigénésiq ue ou l'âge de la mère à la dernière naissance (ces deux variables contiennent des informations similaires)
- l'âge du père et l'âge de la mère au mariage
- la date de mariage du couple (car la période de mariage pourrait entraîner des différences de conditions de vie et donc un biais potentiel au moment de l'interprétation des résultats)

### variables généalogiques liées aux variables génétiques :

- l'entropie généalogique (permet d'évaluer pour un individu si on dispose d'une profondeur généalogique suffisante) (Kouladjian, 1986)
- la variance généalogique (permet d'évaluer si l'information généalogique dont on dispose est comparable du côté paternel et maternel) (Kouladjian, 1986)

### variables génétiques:

- l'apparentement total du couple
- la consanguinité totale de chaque membre du couple

Chacune de ces variables sera l'objet d'une subdivision en ses composantes proche et éloignée dans la section 4.6.

## **4.2 La progéniture totale**

### 4.2.1 Création d'un fichier par individus

Une fois les couples sélectionnés, c'est au destin de leur progéniture que nous nous intéressons. L'effectif qui comptait à la base 308 couples (donc 308 lignes) a été manipulé afin d'utiliser la progéniture comme unité d'analyse. Les couples sélectionnés précédemment étant également les parents des individus à l'étude, les termes « couple » ou « parents » seront utilisés pour les désigner tout au long de l'étude. Pour ce faire, un deuxième fichier, dont l'unité est la progéniture, a été uni au premier en utilisant la variable « numéro d'union » des parents ( commune aux deux fichiers) pour apparier chaque enfant au bon couple de géniteurs. Puisque chaque couple possède habituellement plus d'un enfant, les lignes du fichier par couples sont multipliées proportionnellement par le nombre d'enfants de chaque couple. Nous obtenons finalement un effectif total de 2551 enfants provenant de

308 différents couples de géniteurs. Les valeurs d'apparentement et de consanguinité des parents ainsi que toutes les variables d'ordre démographique se rapportant initialement aux parents ont donc été l'objet d'une multiplication dont le facteur est égal au nombre d'enfants de chaque couple.

En d'autres mots, chaque variable propre aux parents à été recopiée pour chacun de leurs enfants. Ainsi, à partir du fichier initial de 308 lignes, un nouveau fichier possédant 2551 lignes a été généré. Chacune de ces lignes représente un individu à qui on peut attribuer toutes les variables caractéristiques à ses parents en plus de toutes les variables qui lui sont propres (présentées plus loin dans la présente section).

Le fait que tous les enfants issus d'un même couple possèdent les mêmes valeurs pour plusieurs variables telles que le nombre d'enfants au sein de la famille, les valeurs de consanguinité paternelle et maternelle, les valeurs d'apparentement, etc. ne constitue pas un problème puisque la méthode statistique choisie utilise ces variables par rapport à l'individu lui-même et les évalue comme facteurs de risque pour la mortalité. Comme il a été mentionné plus haut, chacun des enfants d'une même famille possède également des variables qui lui sont propres. Ces variables proviennent du fichier ayant la progéniture pour unité de base. Elles permettent de mettre en valeur les risques de mortalité différents à l'intérieur d'une même famille.

Voici la liste de ces variables se rapportant à l'individu (l'enfant):

- sexe de l'enfant
- rang de naissance de l'enfant
- âge au décès de l'enfant (en jours)
- année de naissance de l'enfant
- âge de la mère à la naissance de l'enfant (en années)
- intervalle génésique : s'il s'agit du premier enfant, la variable correspond au nombre de jours entre le mariage des parents et la naissance de l'enfant; si le rang de naissance de l'individu est différent de un, la variable correspond au nombre de jours séparant la naissance de l'individu de celle de son aîné.

#### 4.2.2 Les exclus

La variable « âge au décès de l'enfant » a également entraîné une modification supplémentaire de l'effectif. En effet, sur les 2551 individus, seulement 1009 ont une date de décès connue. Quatre possibilités se présentent:

- la date de décès des autres individus a été perdue (ou manquante),
- ces personnes ont émigré et sont décédées ailleurs,
- elles sont encore vivantes ailleurs,
- finalement, ces personnes sont encore en vie lors de la constitution du Registre (après tout, la date de naissance la plus récente remonte aux années 60).

Ces 1542 individus sont donc exclus de tous les calculs par le logiciel lui-même qui les rejette pour des raisons de « données manquantes ».

Il est très important de spécifier que le retrait de ces 1542 individus n'altère en rien l'intégrité de l'ensemble des variables des 1009 individus restants. Il est vrai que par cette soustraction, des familles ont été tronquées ou éliminées de l'étude. Ce problème sera débattu dans les paragraphes suivants. Cependant, l'information

disponible se rapportant aux individus restants n'est aucunement changée et demeure fiable pour illustrer la situation de l'individu au sein de sa véritable famille. Prenons un exemple concret pour illustrer cette affirmation. Si une famille donnée possède trois enfants, trois « lignes-individus » seront générées pour créer le fichier « par individus » de 2551 lignes. Par la suite, si l'aîné (rang de naissance = 1) de cette famille n'a pas de date de décès connue, il sera éliminé de l'étude. Cependant, pour les deux individus conservés dans le fichier de 1009 lignes, on lira toujours respectivement « 2 » et « 3 » pour la variable « rang de naissance » (le rang « 1 » est disparu). Il s'agit de particularités qui les caractérisent dans leur situation familiale. De plus, pour la variable « nombre d'enfants », on verra toujours, pour les deux individus, apparaître le chiffre « 3 », reflet de leur véritable famille, ici tronquée pour les besoins de l'étude. Il est important que ces caractéristiques soient conservées dans leur intégrité puisque l'outil statistique utilisé, la régression logistique (présentée dans la section 4.7), utilisera chacune des 1009 « lignes-individu » comme une unité afin d'estimer les facteurs de risque de mortalité de la progéniture à partir des variables qui la caractérisent.

En ce qui concerne les individus éliminés de l'étude parce qu'ils n'ont pas de date de décès, il importe de s'interroger. En effet, que la date de décès soit manquante ou encore que les individus soient en vie, il demeure toutefois que cette exclusion entraîne un biais. Il est fort possible que ces dates de décès manquantes reflètent une émigration de familles complètes ou d'individus. Ainsi, des individus nés à l'Île-aux-Coudres peuvent être partis et décédés par la suite, privant l'étude de cas de survie aux périodes de mortalités couvertes par l'étude. Des cas de mortalité dans les périodes étudiées sont aussi perdus mais dans une moindre mesure puisque pour être perdus leur départ doit être antérieur à leur décès en bas âge. Ceux qui ne sont pas encore morts sans toutefois avoir émigré ont réussi à passer le cap des années d'expositions étudiées. Ils auraient pu être considérés eux aussi comme des survivants à ces périodes mais leur absence de date de décès (ils sont vivants au moment de la constitution du Registre) nous force à les exclure. En résumé,

une grande quantité des personnes ayant probablement survécu pendant les périodes étudiées sont exclues de la recherche. Ces personnes sont donc sous-représentées par rapports à celles qui sont mortes et ce, probablement pour l'extrémité la plus récente de la période étudiée.

Il a été décidé de maintenir le rejet de ces individus malgré ce problème potentiel de sous-représentation. En fait, ce biais peut être considéré comme un moindre mal puisque la réinsertion de ces individus risquerait de compter dans l'étude des familles complètes d'émigrants; ce phénomène viendrait donc compromettre la fiabilité des critères de sélection (plus particulièrement les critères de Henry) appliqués lors de la sélection initiale des couples. Afin de connaître un peu plus le biais entraîné par le rejet de ces individus (au nombre de 1542), il a été convenu de procéder à des analyses comparatives entre ce groupe (les exclus) et les 1009 individus conservés (les inclus). De plus, des 308 familles initiales, il a été constaté que la progéniture complète de 41 d'entre elles avait été exclue pour « données manquantes » concernant la date de décès. Ces 41 familles (générant à elles seules 257 lignes, donc 257 enfants) ont été l'objet d'un autre regroupement (appelé « les couples » dans les analyses) puisque, si une famille au complet est éliminée, il y a fort à parier qu'elle ait émigré ou, sinon, que tous les enfants la constituant sont encore en vie. Les analyses, afin de tenter de distinguer toutes différences entre ces trois groupes, sont présentées dans la section suivante (4.2.3).

#### 4.2.3 Tests comparatifs entre les inclus et les exclus

Afin de bien mesurer la nature des différences entre les trois groupes mentionnés dans la section 4.2.2, des analyses descriptives accompagnées de tests statistiques ont été menés afin de caractériser certaines variables communes aux trois groupes. Par la même occasion, la distribution de chacune de ces variables a été vérifiée par un test de normalité qui a révélé pour toutes les

variables, et ce, à travers chaque groupe, que les distributions étaient très significativement différentes de la distribution normale ( $0,001 > \text{prob.}$  dans tous les cas). Un test de chi-carré a été fait pour comparer les proportions d'hommes et de femmes dans trois groupes et des tests de comparaison de moyennes non-paramétriques (Wilcoxon/Kruskal-Wallis) ont été effectués pour comparer le nombre d'enfants moyen par famille, la date de mariage moyenne des parents ainsi que pour les valeurs totales, proches et éloignées de l'apparement des parents et de la consanguinité des enfants (dont la nature sera expliquée en détail dans la section 4.6) des trois groupes. Les figures correspondant aux différents tests effectués (distributions des variables, chi-carrés et comparaisons de moyennes sont présentés en annexe J).

Les résultats de ces tests seront analysés dans la présente section plutôt que dans la section « Analyses et résultats » (chapitre 5) puisqu'ils ne concernent pas seulement l'effectif à l'étude mais aussi tous les individus qui en ont été exclus en raison de l'absence de date de décès. Cependant, les conséquences et les interprétations de ces résultats seront mentionnées dans la discussion et dans la conclusion chaque fois qu'elles pourront apporter une compréhension supplémentaire des modèles ou encore éclairer le biais de sélection entraîné par une telle exclusion qui a pourtant été essentielle à l'intégrité de l'étude.

Avant d'aller plus loin, rappelons que les trois groupes en jeu sont : les inclus, (les 1009 individus qui possèdent une date de décès et font partie de l'étude), les exclus (tous les individus exclus pour absence de date de décès, au nombre de 1542), et finalement, un groupe appelé « couple » rassemblant en fait un sous-groupe des exclus au nombre de 257. Ce dernier groupe englobe en fait tous les individus dont la totalité de la famille a été rejetée de l'étude pour données manquantes. Puisque tous les frères et soeurs sont sans date de décès, on peut se demander si ce n'est pas une émigration familiale qui est en cause. Pour cette

raison, on croit que ces individus ont peut-être des caractéristiques particulières et méritent d'être observés séparément.

Le chi-carré effectué sur les proportions d'hommes et de femmes révèle que ces proportions sont différentes d'un groupe à l'autre de façon très hautement significative (voir fig.J.13). Si on regarde les proportions obtenues aux figures J.1 c, J.2 c et J.3 c, on peut voir qu'effectivement, les rapports de sexe diffèrent d'un groupe à l'autre. Le seul groupe à posséder des individus de sexe non-identifié est le groupe des inclus. Les proportions observées chez les inclus sont donc de 0,48563 femme pour 0,48365 homme et 0,03072 individu de sexe inconnu. Chez les exclus, on voit 0,50454 femme pour 0,49546 homme. À part pour les personnes de sexe inconnu, on peut croire que ces deux rapports se ressemblent un peu et présentent des quantités équivalentes d'hommes et de femmes. Cependant, si on regarde le groupe « couples », on voit que le rapport homme-femme est de 0,52918 pour 0,47082. Ce rapport marque un net déséquilibre entre les sexes, la majorité étant en faveur des hommes.

Si on s'attarde maintenant au nombre d'enfants du couple (voir la fig. J.14), on voit que la comparaison de moyennes entre les trois groupes nous révèle que la disparité qui règne entre les groupes est très hautement significative (test de Wilcoxon/Kruskal-Wallis dont  $\text{prob} < 0,0001$ ). Si on désire voir quels groupes diffèrent entre eux, il faut se référer au test de Tuckey qui compare les groupes paire par paire. Toute valeur positive obtenue en comparant deux paires révèle que les moyennes sont différentes. Toutes les paires formées entre groupes différents donnent des valeurs positives, donc, les moyennes d'enfants par familles des trois groupes sont significativement différentes. Si on regarde les moyennes respectives du nombre d'enfants par famille des trois groupes (fig. J.1 a, J.2 a et J.3 a) on voit que les inclus ont en moyenne des familles de 9,531 enfants avec un écart-type de 2,973. La moyenne chez les exclus est plus élevée (10,237 avec

écart-type de 3,179) et, au contraire, celle du groupe « couples » est plus basse (8,8132 avec écart-type de 3,6223).

La date de mariage des parents a aussi été l'objet des mêmes analyses. On voit à la figure J.15 que les dates de mariages moyennes des trois groupes sont différentes de façon très hautement significative encore une fois. De plus, en s'attardant au test de Tuckey, on voit que toutes les paires formées sont différentes entre elles. Aux figures J.1 b, J.2 b et J.3 b, on constate effectivement que les dates moyennes de mariages pour les parents des individus de chacun des groupes sont très différentes. Pour les inclus dans les analyses, la date moyenne de mariage des parents est 1869.292 (écart-type de 34.104). Chez les exclus, cette date moyenne est plus récente (1889.963, écart-type de 37,248) et finalement, chez le groupe « couples », elle est encore plus proche dans le temps (1907,051, écart-type de 37,414).

Les valeurs d'apparentement et de consanguinités (proches totales et éloignées) ont aussi été soumises à des tests de comparaison de moyennes (voir les figures J.16 à J.24). Les seuls résultats ne décelant pas de différence entre les trois groupes sont ceux qui concernent la consanguinité totale de la mère (prob=0,0593), la consanguinité éloignée du père (prob=0,4664) et la consanguinité éloignée de la mère (prob=0,2727). Pour les deux premiers groupes énoncés, le test de Tuckey ne discerne aucune différence entre aucune des paires formées. Pour la consanguinité maternelle éloignée, le test de Tuckey révèle que la paire « couples » et « exclus » n'a pas de différence significative entre les moyennes. Si on observe maintenant les figures J.4 à J.12, on peut observer une tendance générale : pour les moyennes d'apparentement et de consanguinités, le groupe des inclus présente généralement une valeur relativement plus faible que les deux autres groupes, la valeur la plus forte étant habituellement attribuée au groupe « couples ».

### 4.3 Critique de l'information généalogique ; l'entropie et la variance (Kouladjian, 1986)

Afin de pouvoir critiquer l'information généalogique dont on dispose, il existe deux mesures quantitatives: l'entropie et la variance généalogiques.

L'entropie ( $S_B$ ) nous renseigne sur la profondeur de l'arbre généalogique étudié (nombre de générations à notre disposition) et la variance ( $V$ ) permet de voir s'il y a déséquilibre de l'information dans l'arbre généalogique binaire  $B$ . L'entropie donne donc une idée de la quantité d'information ascendante disponible tandis que la variance caractérise la complétude de cette information.

Les formules de l'entropie ( $S_B$ ) et de la variance ( $V$ ) s'énoncent respectivement comme suit :

$$S_B = \sum (N_i / 2^{N_i})$$

$N_i$  correspond à la génération de l'ancêtre fondateur  $i$ .

NOTE : Ici, un fondateur correspond à tout individu qui marque la fin de l'information généalogique disponible dans le Registre (individu dont on ne connaît pas les parents).

$$V = \sum P_i (\log^2 P_i) - \sum S_B^2$$

$P_i$  correspond à la probabilité de l'origine du gène provenant du fondateur  $i$  dans l'arbre binaire  $B$  ( $P_i = 1/2$  à la  $N_i$ ).

Le détail de développement mathématique de ces formules est présenté dans Kouladjian, 1986.

À la lumière de ces informations, on constate que plus on peut remonter de générations, plus l'entropie sera élevée. De même, plus l'information est déséquilibrée entre la branche maternelle et la branche paternelle, plus la valeur de la variance augmentera. Idéalement, on souhaite une entropie maximale associée à une variance minimale.

Un exemple du calcul de l'entropie et de la variance illustré par l'arbre généalogique binaire théorique de l'individu Z est présenté ci-dessous (figure 4.2). Les individus 1, 2, 3, 4, 5 et 6 sont des ancêtres fondateurs car ils constituent la fin de l'information généalogique.

1) Formules et exemple de calcul pour l'entropie et la variance (Kouladjian, 1986)

$$S_B = \sum (N_i / 2^{N_i})$$

$$V = \sum (P_i (\log_2 P_i)^2 - S_B^2)$$

$$S_B = 4 (3/2^3) + 2 (2/2^2)$$

$$V = ((4 \cdot 1/8 (\log_2 1/8)^2 + 2 \cdot 1/4 (\log_2 1/4)^2) - (2,5)^2)$$

$$S_B = 2,5$$

$$V = (4,5 + 2) - 6,25$$

$$V = 0,25$$

2) Représentation de l'arbre généalogique binaire théorique de l'individu Z associé à l'exemple de calcul

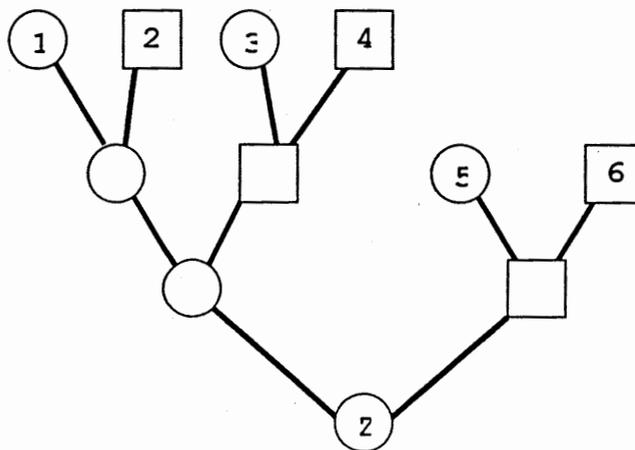


Figure 4.2 : Formules et exemple de calcul de l'entropie et de la variance à partir de la représentation de l'arbre généalogique binaire théorique de l'individu Z.

#### 4.4 Le coefficient de consanguinité

Le coefficient de consanguinité correspond à « la probabilité pour que les deux gènes, possédés par un individu en un locus donné, soient identiques, c'est-à-dire la copie d'un même gène ancêtre » (Jacquard, 1974).

Cette définition exprime la possibilité statistique qu'un individu soit homozygote pour un gène donné et que les deux copies identiques (les allèles) de ce gène que lui ont transmis son père et sa mère originent du même ancêtre.

Le coefficient de consanguinité total d'un individu se mesure selon la formule:

$$f_z = \sum (0.5)^{n_i + p_i - 1} (1 + f_{c_i}) \quad (\text{Wright, 1922})$$

$f_z$  = coefficient de consanguinité de l'individu à l'étude

$f_{c_i}$  = coefficient de consanguinité de l'ancêtre  $i$  commun aux parents de l'individu à l'étude

$n_i$  et  $p_i$  = nombre de générations séparant l'ancêtre  $i$ , commun aux parents de l'individu, en passant respectivement par la mère et par le père

Pour chaque personne (possédant toujours deux allèles, semblables ou non, pour un même gène) la probabilité de transmission d'un de ces deux allèles en particulier est de  $\frac{1}{2}$  ou 0,5 à chaque saut de génération (maillons qui forment la chaîne l'unissant à l'ancêtre  $I$  en passant par  $X$  ( $n_i$ ) ou  $Y$  ( $p_i$ )). Le coefficient de consanguinité d'un individu peut donc varier en fonction du nombre d'ancêtres communs à ses parents ainsi qu'au nombre et à la longueur des chaînes de parenté le reliant à ces ancêtres.

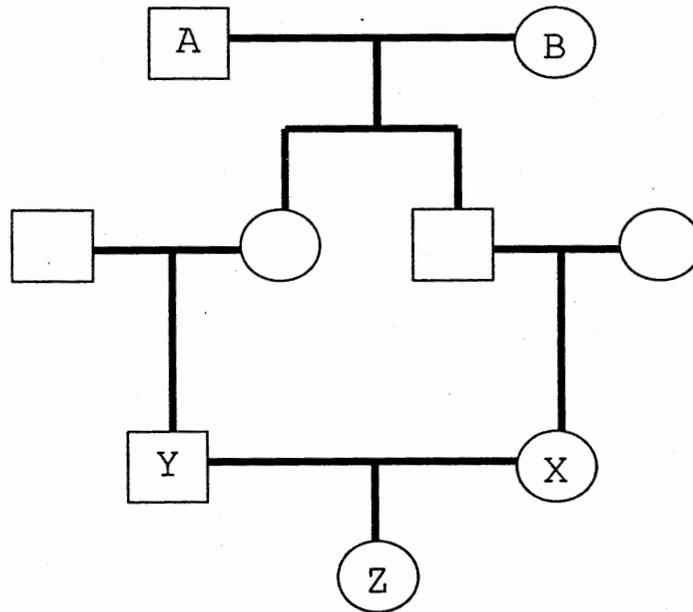


Figure 4.3 : Chaînes de consanguinité reliant l'individu Z aux ancêtres A et B, communs à son père (Y) et à sa mère (X).

Dans la figure 4.3 on présente les ascendances de l'individu Z sur trois générations de profondeur (sans considérer égo). Calculé à partir de cette information, le coefficient de consanguinité de Z correspond à la probabilité que l'ancêtre A ait pu transmettre au père, Y, et à la mère, X, de l'individu Z des copies identiques d'un même gène additionnée à la probabilité que l'ancêtre B ait fait de même. La formule tient également compte du coefficient de consanguinité de chacun de ces ancêtres communs ( $f_{ci}$ ). En effet, si l'ancêtre  $i$  commun aux parents X et Y est consanguin, le coefficient de consanguinité de l'individu Z se voit augmenté d'autant.

#### 4.5 Le coefficient de parenté

Quand nous étudions la consanguinité, c'est à un individu en particulier que nous nous attardons. Le coefficient de parenté, quant à lui, réfère au lien généalogique entre deux individus. En fait, on le définit comme la probabilité qu'un allèle pris au hasard en un locus donné pour un individu soit identique à l'allèle pris chez un autre individu, au même locus.

La formule du coefficient de parenté à été développée par Malécot (1948) et s'énonce comme suit :

$$\Phi_{AB} = \sum (0.5)^{n_i + p_i + 1} (1 + f_{ci}) \quad (\text{Malécot 1948})$$

Si l'on revient à l'exemple précédent, on peut calculer ce coefficient entre les individus X et Y, parents de Z. Le coefficient de parenté entre X et Y correspond à la somme des chaînes unissant X et Y à l'ancêtre A et à l'ancêtre B. Le coefficient de parenté tient également compte, si possible, du coefficient de consanguinité de chaque ancêtre commun à X et Y puisqu'une éventuelle consanguinité d'un ou plusieurs ancêtres communs trahirait la présence d'autres chaînes de parenté caractérisant la consanguinité de l'ancêtre commun.

Il est à noter que le coefficient de consanguinité de l'individu Z correspond exactement au coefficient de parenté de ses parents. Il s'agit, mathématiquement, de la même mesure mais appliquée différemment. On peut donc dire qu'un enfant hérite non seulement d'une partie du bagage génétique de chacun de ses parents mais qu'en plus, il porte en lui la même probabilité d'homozygotie que la probabilité de ses parents ont de posséder des allèles communs.

#### 4.6 Consanguinité proche et éloignée

Il est possible de séparer la consanguinité proche de la consanguinité éloignée afin de tenter de déceler un impact présumé différentiel de ces deux variables dans le cadre de l'étude. Dans la vie courante, on juge une consanguinité proche comme tout lien de parenté que se reconnaissent les individus entre eux. Cette définition varie selon les sociétés et donc, pour les besoins de l'étude, il sera nécessaire de fixer des limites arbitraires entre la consanguinité proche et la consanguinité éloignée.

Pour pouvoir fixer des limites, il faut se référer aux chaînes de consanguinité. Les cycles de consanguinité correspondent aux chaînes reliant un individu à chacun des ancêtres communs à ses parents. La chaîne part toujours de l'individu dont on veut mesurer la consanguinité et remonte les générations jusqu'à l'ancêtre en passant du côté maternel et recommence ensuite du côté paternel. Un lien 3-3 indique donc qu'il faut remonter 3 générations à partir de l'individu en passant par le côté maternel et 3 générations en passant par le côté paternel pour remonter jusqu'à l'ancêtre commun. Le lien entre la mère et l'ancêtre versus le lien entre le père et l'ancêtre peut très bien être inégal (4-5, 2-3, etc.).

Dans son étude menée sur la population de l'île de Saint Barthélémy et réalisée à l'intérieur du laboratoire du Dr Mayer, Gaël Plassart (1997) a dû établir la frontière entre consanguinité proche et éloignée. Dans le cadre de son travail, tout lien supérieur à un lien 5-5 a été calculé comme consanguinité éloignée (les gènes ancêtres issus de cette consanguinité ont été considérés comme le résultat de l'épuration et de la sélection progressive d'une génération à l'autre). Cette séparation arbitraire a mis en valeur le fait que dans une population qui accumule des liens de consanguinité (endogamie) sur plusieurs générations, on peut tenter de mesurer les effets des deux types de consanguinités (proche et éloignée avec une division 5-5). Des tests de symétrie ont prouvé que les distributions des

individus classés selon la consanguinité proche ou la consanguinité éloignée était différentes; il en est de même pour les distributions entre les consanguinités proches et totales. Cependant, les distributions des individus selon la consanguinité totale ou éloignée étaient identiques, ce qui veut dire que pour cette population, le coefficient de consanguinité total d'un individu reflète en majorité la consanguinité éloignée. Étant donné une histoire endogame comparable (quoique moins prononcée) la présente étude utilisera les mêmes frontières pour séparer la consanguinité proche de la consanguinité éloignée. Il est important de souligner que cette délimitation demeure une division arbitraire même si elle a permis de montrer une différence notable dans les travaux de Plassart. Ainsi, pour chaque enfant de l'effectif, trois valeurs d'apparentement, trois valeurs de la consanguinité de leur mère et trois valeurs de la consanguinité de leur père ont été calculées (totale, proche et éloignée).

Tout d'abord, le fichier initial dont l'unité était le couple possédait des variables génétiques « totales » (c.f. section 4.1 : Critères de sélection des couples). Ces valeurs d'apparentement entre les conjoints et de consanguinité de chacun des individus correspondent à la somme des liens (les rattachant aux ancêtres communs) accumulés au fil des générations. Chaque coefficient ainsi obtenu porte donc potentiellement une part de liens proches et une autre de liens éloignés.

La consanguinité proche de chaque parent ainsi que leur apparentement proche ont été obtenus directement à partir du Registre de population informatisé en établissant les limites de recherche d'ancêtres communs à la profondeur généalogique voulue. Pour obtenir les valeurs éloignées, il suffit alors de soustraire la valeur proche de la valeur totale, les formules de consanguinité et d'apparentement correspondant en fait à une sommation des liens entraînés par chaque ancêtre commun.

La consanguinité proche utilisée dans le cadre de nos analyses considère donc toutes les chaînes de consanguinité de valeur 5-5 ou inférieure (5-4, 4-4, 4-3, etc.) conformément aux analyses de Plassart. Pour l'apparentement proche, ce sont les chaînes de valeurs 4-4 ou inférieures qui constituent la variable. Pourquoi cette différence d'une génération? Afin de conserver, pour les valeurs de consanguinité et d'apparentement, une quantité d'information comparable.

Pour bien visualiser cette affirmation, voici un schéma simplifié (figure 4.4) de l'information dont on dispose théoriquement sur la généalogie d'un individu lorsque l'on calcule son coefficient de consanguinité en comparaison avec celle disponible lorsqu'on s'attarde au calcul du coefficient de parenté entre deux individus (apparentement). Si l'information généalogique est bien complète, on a, pour chaque génération, deux fois plus d'individus quand on observe l'apparentement que lorsqu'on s'intéresse à la consanguinité d'une seule personne. C'est donc dire qu'on a deux fois plus de chances de trouver des ancêtres communs! Pour pallier à ce biais, on mesure l'apparentement proche en coupant (dans les calculs) une génération par rapport à la consanguinité. De cette manière, notre information proche se base sur les mêmes « probabilités » de trouver des ancêtres communs car on dispose ainsi du même nombre d'individus « ancêtres potentiels ».

Dans la figure 4.4, qui présente une profondeur de 3 générations (sans tenir compte d'égo), on voit que la richesse de l'information double théoriquement à chaque génération puisque chaque individu possède deux parents qui possèdent également deux parents, etc. Cette affirmation est applicable à l'individu A aussi bien qu'à B. De cette façon, on possède, à la troisième génération uniquement, 8 ancêtres pour A comme pour B. Si on regarde l'apparentement entre A et B pour cette troisième génération, nous avons donc un potentiel de 16 ancêtres disponibles pour rechercher un lien génétique entre eux.

Si l'on veut pouvoir disposer d'une quantité comparable d'ancêtres, on peut donc garder l'information jusqu'à la deuxième génération qui nous fournit, pour l'apparement de A et B, une possibilité de 8 ancêtres

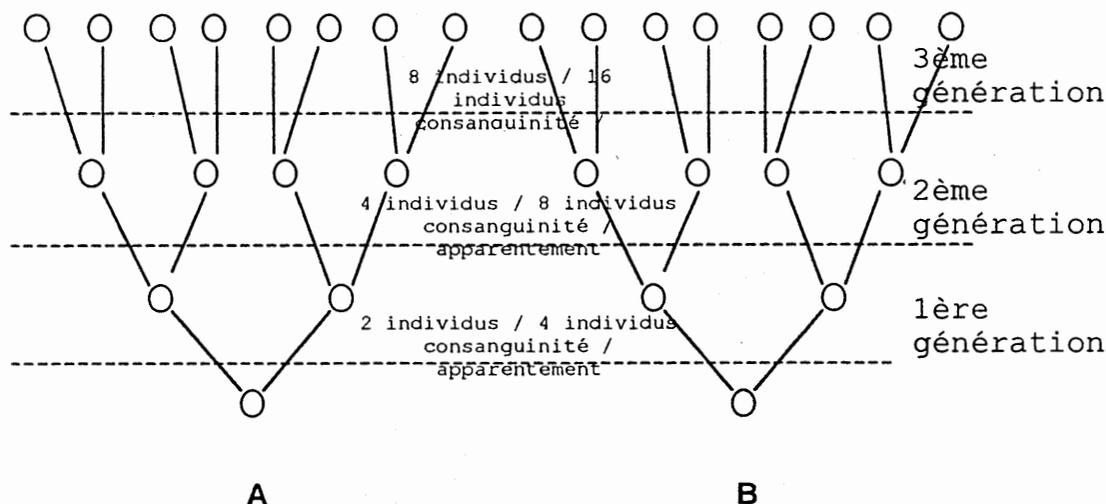


Figure 4.4 : Illustration de la quantité d'information disponible pour le calcul des coefficients de consanguinité (l'individu A OU l'individu B) comparée à la quantité d'information disponible pour le calcul de l'apparement entre A et B pour chaque génération (pour une profondeur de 3 générations).

Pour chaque type de mortalité, nous disposerons donc de trois groupes de variables génétiques :

- l'apparement total entre les parents ainsi que les consanguinités maternelle et paternelle totales.
- l'apparement proche (liens jusqu'à 4-4) entre les parents ainsi que les consanguinités maternelle et paternelle proches (liens jusqu'à 5-5).
- l'apparement éloigné (liens plus profonds que 4-4) entre les parents ainsi que les consanguinités maternelle et paternelle éloignées (liens plus profonds que 5-5).

Les variables génétiques totales et proches ont été calculées directement à partir du Registre de population. Afin d'obtenir les variables génétiques éloignées, il suffit

de soustraire la valeur proche de la valeur totale. Chaque régression effectuée a donc été faite en soumettant au modèle un des trois groupes de variables génétiques (totales, proches ou éloignées) et l'analyse des résultats favorise l'établissement d'un parallèle entre chacun des types de régression pour voir si un impact différentiel est détectable.

#### 4.7 Technique d'analyse choisie

##### 4.7.1 Approche multivariée: la modélisation par la régression logistique

La régression logistique est un modèle statistique qui permet d'obtenir la probabilité d'occurrence d'une variable binaire dépendante (Y) en fonction de plusieurs variables explicatives indépendantes ( $X_1$  à  $X_n$ ).

Le modèle logistique s'énonce comme suit :

$$\text{logit } [p(Y)] = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + \dots + a_nX_n$$

Le terme « logit [p(Y)] » se traduit par «  $\ln[p(Y)/1-(Y)]$  » ainsi la probabilité « p(Y) » dans le modèle s'écrit de la façon suivante :

$$p(Y) = \frac{1}{1 + e^{-(a_0 + a_1X_1 + a_2X_2 + a_3X_3 + \dots + a_nX_n)}}$$

Y=variable dépendante

p(Y)= probabilité de la variable dépendante

$a_0$ =ordonnée à l'origine

$a_1$  à  $a_n$ =coefficients des variables indépendantes

$X_1$  à  $X_n$ =variables indépendantes

La fonction obtenue est une courbe de forme sigmoïde allant de moins l'infini à plus l'infini sur l'axe des X (cet axe correspond à l'effet d'ensemble de toutes les variables explicatives du modèle et peut être comparé à une "dose d'exposition") et de 0 à 1 sur l'axe des Y, nous donnant ainsi la probabilité d'occurrence de la variable Y (risque) en fonction de la "dose d'exposition". Plus le paramètre en X est petit, plus le risque est faible et plus ce même paramètre X est élevé, plus le risque est grand. Cette technique s'applique très bien aux données et nous avons donc la mortalité, la mortalité néonatale précoce, la mortalité infantile, la mortalité entre 1 et 4 ans ainsi que la mortalité pré-reproductive (1 à 14 ans) comme variables dépendantes (Y) pour les différents modèles de régression (Hosmer et Lemeshow, 1989; Kleinbaum, Kupper et Muller, 1987). La figure 4.5 illustre la forme générale adoptée par une courbe de régression logistique.

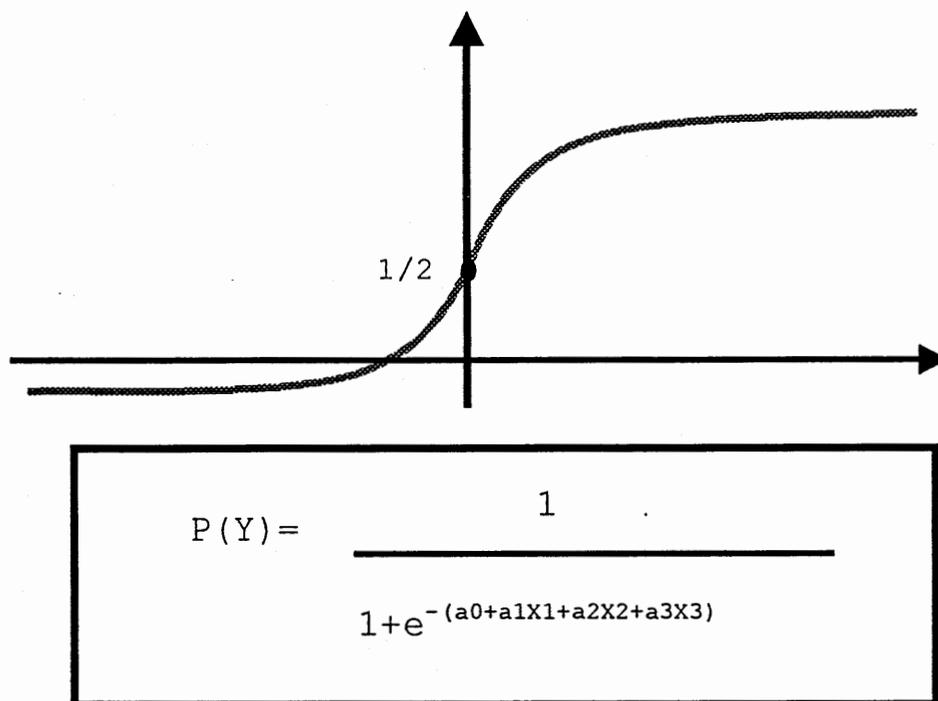


Figure 4.5 : Illustration de la forme générale d'une courbe de fonction logistique et formule associée à la régression logistique.

#### 4.7.2 Le rapport de cotes :

Dans toutes les régressions linéaires, il est très simple d'interpréter les coefficients de régression. Cependant, pour la régression logistique, la formule complexe utilisant la fonction logarithmique impose un réarrangement de l'équation et un nouveau mode d'interprétation pour assurer la compréhension de l'impact de chaque variable dépendante. L'outil choisi pour interpréter cet impact est le rapport de cotes obtenu pour chaque variable.

Dans un tableau de contingences à deux entrées qui servirait à entrer la survie ou le décès des individus, le rapport de cotes s'exprime ainsi :

$$\text{rapport de cotes} = \frac{a/b}{c/d}$$

	présence de décès	absence de décès
exposé	a	b
non-exposé	c	d

a = fréquence des individus exposés qui sont décédés

b = fréquence des individus exposés qui ne sont pas décédés

c = fréquence des individus non-exposés qui sont décédés

d = fréquence des individus non-exposés qui ne sont pas décédés

Ici, le rapport de cotes compare donc le risque de décès des individus exposés (a/b) au risque de décès des individus non-exposés (c/d).

Dans la régression logistique,  $P(Y)$  nous donne la probabilité d'occurrence d'un événement (dans notre cas, la mortalité). Si cette probabilité était de 100%, alors,  $P(Y)=1$ , si au contraire, elle est de 0%,  $P(Y)=0$ .

Une fois cette probabilité trouvée, on peut étudier l'impact de chaque variable en se fiant à son rapport de cotes. Celui-ci indique le facteur par lequel il faut multiplier le risque de mortalité à chaque fois que la variable étudiée augmente d'une unité. Si la variable est exprimée en jours, l'augmentation d'une unité correspond à l'ajout d'un jour; s'il s'agit d'années, alors l'unité considérée est l'année.

Si le rapport de cote est supérieur à 1, l'augmentation a donc un effet nocif sur le risque de mortalité puisque son augmentation d'une unité augmente également le

risque de mortalité. Dans le cas contraire, si le rapport de cote de la variable est inférieur à 1, son augmentation d'une unité diminue le risque de mortalité (SPSS Inc., 1997).

Les modèles ont été faits à l'aide du logiciel SPSS et des sélections de type ENTER (toutes les variables choisies font partie du modèle indépendamment de leur seuil de signification) et FORWARD (ajout de variables selon leur niveau de signification) ont été appliquées.

#### **4.8 Transformation des variables de mortalité pour répondre à un mode binaire**

Afin de mener à bien les modélisations du risque de mortalité chez la progéniture à l'aide de la régression logistique, il a fallu effectuer quelques transformations des variables dépendantes. En effet, comme il est expliqué dans la section précédente, la régression logistique présente un modèle explicatif pour l'occurrence ou la non-occurrence d'un événement. La variable dépendante est donc binaire et les variations observées au sein des variables indépendantes du modèle modulent le risque d'occurrence (« 1 ») ou de non-occurrence (« 0 ») de la variable dépendante.

Ici, l'événement qui sera modélisé est la mort d'un enfant à l'intérieur d'une période déterminée. Cinq types de mortalité ont été choisis pour former des variables binaires : la mortinatalité (0 jour, enfant mort-né), la mortalité néonatale précoce (mortalité entre le premier et le sixième jour de vie), la mortalité infantile (mortalité dans la première année de vie), la mortalité entre 1 et 4 ans et la mortalité entre 1 et 14 ans. Les trois premières catégories de mortalité sont reconnues par Statistiques Canada comme des composantes démographiques de la mortalité foeto-infantile (Péron et Strohmenger, 1985). La signification démographique des deux derniers types de mortalité correspond respectivement à la période de susceptibilité aux maladies infantiles et à la mortalité dite pré-reproductive dont les

critères d'âge peuvent varier selon les études et le type de données (Liazoghli, 2000).

Tous ces types de mortalité sont issus d'une variable spécifique à la progéniture : « aged », ou plutôt, l'âge de l'enfant à son décès. L'unité de cette variable est la journée; l'information obtenue par « aged » est donc le nombre de jours de vie de l'individu. Une année compte en moyenne 365,25 jours (4 années totalisent 1462 jours et 14 années, 5113 jours). À l'aide du logiciel SPSS, des variables binaires ont donc été déduites à partir de la variable continue « âge au décès ».

On peut donc exprimer ainsi les cinq nouvelles variables :

- Mortalité infantile binaire (mortinfb) :

mortinfb = 1 si aged < ou = à 365,25

mortinfb = 0 si aged > 365,25

- Mortinatalité binaire (mort0jb) :

mort0jb = 1 si aged = 0

mort0jb = 0 si aged est différent de 0

- Mortalité néonatale précoce binaire (mort06b) :

mort06b = 1 si aged < 7

mort06b = 0 si aged > ou = à 7

- Mortalité entre 1 et 4 ans binaire (mort4b) :

mort4b = 1 si 365,25 < aged < 1462

mort4b = 0 si 365,25 > ou = aged et si 1462 < ou = aged

- Mortalité pré-reproductive binaire (mort14b) :

mort14b = 1 si  $365,25 < \text{aged} < 5113$

mort14b = 0 si  $365,25 > \text{ou} = \text{aged}$  et si  $5113 < \text{ou} = \text{aged}$

Ces cinq nouvelles variables binaires sont celles qui seront utilisées comme variables dépendantes pour toutes les régressions logistiques de cet ouvrage.

#### **4.9 Création des classes d'apparentement et de consanguinité pour les modèles de régression logistique**

La consanguinité et l'apparentement sont des variables qui oscillent entre 0 et 1 avec des variations de l'ordre de la sixième décimale. Les rapports de cotes utilisés en régression logistique se basent sur la variation de plus ou moins une unité pour établir l'apport positif (protecteur) ou négatif (néfaste) de cette variable indépendante. Cependant, le rapport de cote mesure le passage de 0 à 1, non de 0,0000 à 0,0001. Afin de pouvoir utiliser les variables d'apparentement et de consanguinité dans les modèles de régression logistique, il a fallu les séparer par classes pour que les rapports de cotes obtenus soient sensibles au passage d'une classe à l'autre et, donc, interprétables.

Le tableau 4.1 présente les classes d'apparentement et de consanguinité créées pour permettre d'intégrer ces variables dans les modèles de régression logistique.

Dans ce tableau sont également présentées les principales valeurs correspondant à des liens sociaux reconnus. Puisque les effets de l'évolution des coefficients d'apparentement et de consanguinité sont mal connus, il a été décidé de créer des classes arbitraires équivalentes pour toutes les variables d'apparentement et de consanguinité (totales, proches et éloignées). Les divisions ont cependant été faites de manière à ce que chacune ne puisse contenir plus d'une valeur socialement reconnue.

Tableau 4.1: Représentation des intervalles de valeurs sélectionnées pour former les classes d'apparentement et de consanguinité avec les valeurs d'apparentement correspondant à un lien social reconnu.

Classes d'apparentement ou de consanguinité	valeurs d'apparentement ou de consanguinité correspondantes	Valeurs correspondant à un lien social
1	valeur nulle	-----
2	[0,000 000 000 1 à 0,001]	-----
3	[0,001 000 000 1 à 0,01]	-----
4	[0,010 000 000 1 à 0,03]	cousins issus de germains (0,015625)
5	[0,030 000 000 1 à 0,04]	cousins de consanguinités inégales (0,03125)
6	[0,040 000 000 1 à 0,06]	cousins germains (0,0625)
7	[0,060 000 000 1 à 0,07]	-----
8	[0,070 000 000 1 à 0,08]	-----
9	[0,080 000 000 1 à 0,1]	-----
10	[0,100 000 000 1 à 0,2]	oncle-nièce (ou tante-neveu) (0,125)

Par la suite, une distribution de l'effectif restant (les 1009 enfants) selon les classes d'apparentement et de consanguinité a été faite. Cette distribution est représentée dans le tableau 5.1 de la section 5.2.1.

#### **4.10 Organisation de l'analyse des modèles de régression logistique**

Afin de tenter de caractériser le rôle de la consanguinité maternelle face à la mortalité de la progéniture, plusieurs régressions logistiques ont été effectuées. Chacun de ces modèles présente des variables socio-démographiques ainsi que les variables d'apparentement et de consanguinité totales, proches ou éloignées (variables génétiques).

La mortalité infantile (0-1 an), les mortalités périnatales (mortalité, 0 jour et mortalité néonatale précoce, 1-6 jours), la mortalité 1-4 ans et la mortalité pré-reproductive (1-14 ans) ont été utilisées comme variables dépendantes afin de mesurer l'impact potentiel de la consanguinité maternelle sur la vie ou la mort de la progéniture et ce, en présence de multiples variables pouvant également avoir un poids significatif dans le modèle.

Dans le but de faciliter l'analyse, les résultats seront présentés dans la section 5.4 selon l'ordre suivant :

- 5.3.1 Classes de mortalités touchant la progéniture totale
- 5.3.2 Classes de mortalités touchant la progéniture masculine
- 5.3.3 Classes de mortalités touchant la progéniture féminine

Chacun de ces modèles est fait en trois volets de consanguinité et d'apparentement (total, proche et éloigné), ce qui donne lieu aux trois sections suivantes :

- 5.3.1.1 modèle effectué avec les variables génétiques totales
- 5.3.1.2 modèle effectué avec les variables génétiques proches
- 5.3.1.3 modèle effectué avec les variables génétiques éloignées

Les trois catégories citées ci-haut concernent la progéniture totale. Les sections 5.3.2.1, 5.3.2.2 et 5.3.2.3 sont semblables mais visent les enfants de sexe masculin et les sections 5.3.3.1, 5.3.3.2 et 5.3.3.3 concernent la progéniture féminine.

La seule différence entre ces trois types de modèles réside donc dans les variables génétiques qui leurs sont soumises. Si aucune de ces variables génétiques n'atteint le seuil de signification inférieur ou égal à  $\text{sig}=0,05$ , alors les trois modèles seront identiques. Dans ce cas, un seul sera analysé.

Enfin, chacun de ces 9 tableaux sera divisé selon les cinq classes de mortalité établies dans la section précédente; chacune de ces classes de mortalité est l'objet d'une régression logistique distincte :

- A) Mortalité infantile (0-1 an)
- B) Mortinatalité (0 jour)
- C) Mortalité néonatale précoce (1-6 jour(s))
- D) Mortalité entre 1 et 4 ans
- E) Mortalité entre 1 et 14 ans (pré-reproductive)

#### **4.11 Choix des variables soumises aux régressions logistiques et tests de corrélations**

Comme il a été mentionné au début du présent chapitre, le Registre de population de l'Île- aux-Coudres donne accès à une multitude de variables démographiques et biologiques. La liste complète des variables disponibles pour l'étude associées à leur acronyme est fournie en appendice dans le tableau A.1.

#### 4.11.1 Les variables démographiques

Les variables démographiques qui ont été sélectionnées pour la modélisation sont :

- le nombre d'enfants du couple (nbenf)
- la date de mariage des conjoints (datem)
- le rang de naissance de chaque enfant (rangnai)
- l'âge de la mère à la naissance (agemn)
- l'année de naissance de l'enfant (year)
- la différence d'âge entre les deux conjoints (difconj)
- l'intervalle (en jours) qui sépare la naissance de l'enfant de celle de l'enfant qui le précède dans le rang des naissances (interidi)

Parmi ces variables, une sélection doit être faite afin d'éviter un surplus de colinéarité, plusieurs d'entre elles étant fortement corrélées. La régression logistique n'est pas particulièrement sensible à la corrélation mais il est inutile de surcharger un modèle et ses interprétations avec des variables qui transmettent trop d'informations similaires.

Comme il a déjà été mentionné dans le contexte théorique (Boisvert,1992), l'Île-aux-Coudres présente une population relativement homogène sur le plan économique (Boisvert, 1992; Martin,1957) c'est pourquoi un indice de situation économique tel que le SES ou score socio-économique (Shull, 1970) n'est ni disponible, ni utile dans la population à l'étude. Parmi les variables citées, la seule pouvant refléter une certaine situation économique est la différence d'âge entre les conjoints. En effet, plus la différence est marquée, plus on pourrait s'attendre à ce que l'homme, généralement plus vieux que son épouse, ait une situation stable et soit apte à faire vivre la famille. Cette variable porte donc une information unique et utile. Pour cette raison, elle sera retenue pour la modélisation.

En ce qui concerne le nombre d'enfants du couple, la date de mariage des conjoints, le rang de naissance de chaque enfant, l'âge de la mère à la naissance et l'année de naissance de l'enfant, un test de corrélation a été effectué afin de déterminer les liens qui existent entre ces variables et celles qui devraient faire partie du modèle. Le but n'étant pas d'éliminer à tout prix des variables, seulement celles qui présentent une corrélation très forte entre elles seront l'objet d'une élimination sélective.

L'âge du père au mariage n'a pas été retenu puisque cette information transporte de façon évidente une information redondante si on la compare avec la variable « différence d'âge entre les conjoints » et « l'âge de la mère à la naissance de chaque enfant ». La figure B.1 présentée en appendice démontre les corrélations existant entre certaines variables démographiques.

Comme prévu, les variables sont très corrélées. Tout d'abord, le nombre d'enfants du couple a sa plus haute corrélation avec le rang de naissance (0,4079). Cette corrélation s'explique du fait que plus le nombre d'enfants est grand au sein d'une famille, plus le rang des naissances ira jusqu'à une haute valeur. Par exemple, pour une famille de trois enfants, le rang de naissance maximal est de trois tandis que pour une famille de douze enfants, ce maximum passe également à douze.

Toutefois, le nombre d'enfants d'une famille et le rang de naissance portent des informations qui leur sont particulières : le premier reflète la charge des parents, donc l'attention donnée à chaque enfant et le second, la situation « sociale » et « biologique » de l'enfant au sein de la famille. Pour ces raisons, aucune de ces variables ne sera éliminée. Les autres variables corrélées au nombre d'enfants de la famille le sont de façon beaucoup moins marquée.

Une autre corrélation importante se dénote entre la date de mariage des parents et la date de naissance de chaque enfant (0,9827). L'information importante dans

ces deux variables reflète l'époque, l'étude étant menée sur une tranche de plus de cent ans. Il est normal que la date de mariage des parents soit intimement liée aux dates de naissance de chacun de leurs enfants puisque le mariage est, en temps normal, le point de départ d'une famille et vise, surtout au 19<sup>e</sup> siècle, à avoir une progéniture nombreuse pour aider aux activités liées à l'économie des terres entre autres (Le Querrec, 1981). Il a donc été jugé nécessaire, pour ne pas alourdir le modèle, de ne garder qu'une seule de ces deux variables. La date de naissance est propre à chaque enfant et reflète probablement plus adéquatement la situation à laquelle font face l'enfant et sa famille au moment de la naissance. Pour cette raison, la date de mariage des parents a été retirée des variables à entrer dans le modèle.

On observe également une corrélation élevée entre le rang de naissance et l'âge de la mère à la naissance (0,8037). Cette corrélation est très normale et prévisible. En effet il est normal qu'une mère soit plus âgée à la naissance de son deuxième enfant qu'à la naissance de son premier!

Enfin, le rang de naissance et l'année de naissance de chaque enfant sont également corrélés (0,2368). Comme pour le nombre d'enfants de la famille et le rang de naissance, cette corrélation est très prévisible. Il est normal que le rang de naissance augmente au fil des années. En effet, la date de naissance d'un enfant premier-né est toujours antérieure à celle de son cadet. Toutefois, l'information véhiculée par ces deux variables est très distincte car elles expriment respectivement l'époque et la situation socio-biologique au sein de la famille. Pour cette raison, ces deux variables seront gardées pour les analyses.

#### 4.11.2 Les variables génétiques

L'apparentement et la consanguinité (totaux, proches et éloignés), variables d'ordre génétique, seront incluses dans le modèle afin de voir l'impact biologique (s'il y en

a un) de l'endogamie perpétuée depuis des générations au sein de la population à l'étude. Cependant, ces variables sont calculées à l'aide de données généalogiques dont la précision dépend directement de la profondeur de l'information généalogique disponible. Aussi, les variables d'entropie et de variance (définies dans la section 4.3) ne peuvent être incluses dans le modèle, vu leur manque de signification propre. Elles ont tout de même un impact sur les valeurs d'apparentement et de consanguinité. Des tests de corrélation ont été effectués afin de mesurer le lien qui unit ces deux groupes de variables; les résultats sont présentés en appendice à la figure B.2.

Tel que prévu, les coefficients d'apparentement des membres du couple ainsi que les coefficients de consanguinité du père et de la mère sont corrélés plus ou moins fortement avec l'entropie et la variance. Une augmentation de la profondeur des informations généalogiques, donc de l'entropie, est corrélée avec une augmentation des trois coefficients avec des indices respectifs de 0,3018, 0,2669 et 0,1413. Au contraire, on dénote une corrélation négative entre la variance et les trois coefficients. Ces corrélations négatives sont moins fortes que celles obtenues pour l'entropie et sont respectivement  $-0,1225$ ,  $-0,0186$  et  $-0,0200$ . Une diminution de la variance (ce qui reflète un arbre plus équilibré) entraîne donc habituellement une augmentation des trois coefficients (d'apparentements et de consanguinités). En effet, s'il manque peu d'information généalogique, on a plus de chances de trouver des ancêtres communs.

#### 4.11.3 Tests de symétrie des variables génétiques

Des tests de symétrie ont été effectués afin de bien pouvoir discerner les variables génétiques totales des variables génétiques proches et des variables génétiques éloignées. Les résultats de ces tests sont présentés en appendice (tableaux E.1 à E.9). Selon ces tests de symétrie, la distribution des enfants selon la consanguinité maternelle totale est comparable à leur distribution selon la consanguinité

maternelle proche et est aussi comparable à celle de la distribution selon la consanguinité maternelle éloignée. Toutefois les distributions des enfants selon les consanguinités maternelles proche et éloignée sont significativement différentes (Stokes, Davis et Kosh, 1995). Les mêmes phénomènes sont observables pour la consanguinité paternelle (totale, proche et éloignée) et pour l'apparement des parents (total, proche et éloigné). Ceci nous porte à croire que les individus à l'étude (et la population de l'Île-aux-Coudres) portent tous une part de liens proches et une part de liens éloignés. Lorsqu'on regarde la distribution totale des enfants selon une variable génétique, on voit donc que cette distribution est à la fois le reflet de ses liens proches et de ses liens éloignés. Ce phénomène reflète bien le caractère endogame de la population. Toutefois, la majorité des enfants classés selon des liens proches se distribuent différemment de la majorité de ceux-ci lorsque classés selon des liens éloignés dans la population.

## CHAPITRE V

### ANALYSES ET RÉSULTATS

#### 5.1 Analyse descriptive de la variance et de l'entropie généalogiques de l'effectif

Maintenant que les corrélations entre les variables génétiques et généalogiques sont démontrées, il serait important de connaître la puissance de l'entropie et de la variance de l'effectif avant de commencer les analyses. La figure C.1, présentée en appendice, illustre la distribution des variables « entropie » et « variance ».

Ainsi, en ce qui concerne notre échantillon, l'entropie moyenne est de 4,347 générations. Cette information doit être comprise comme suit : l'échantillon ayant été sélectionné par « couple », l'entropie moyenne pour chaque couple sélectionné pour l'étude est de 4,347. Cette moyenne ne tient pas compte de la génération de la progéniture de chaque couple. Il serait donc théoriquement possible d'ajouter 1 à l'entropie moyenne. Cependant, tous les calculs de consanguinité se rapportent à la génération parentale pour en mesurer l'impact sur la mortalité des enfants. Il est donc inutile de remoduler les valeurs d'entropie puisqu'elles sont justes dans le cadre de nos calculs. La génération filiale n'est finalement là que pour donner le reflet de l'impact de l'apparentement sur la progéniture et n'est pas incluse dans les générations compilées.

Si l'on s'attarde aux quartiles, on réalise qu'à l'intérieur de notre échantillon, 75% des données fournissent une entropie supérieure ou égale à 3,50 générations à partir de la génération parentale. En comparaison à d'autres études (Bittles, 1994; Shull et al., 1970a, 1970b) qui n'utilisaient jamais plus de trois générations, cette entropie permet de donner une information très précise vu sa profondeur. Il est aussi également important de mentionner que la plupart de ces études se fiaient seulement sur les déclarations des individus faisant partie de l'effectif et incluaient

la génération filiale dans leurs calculs de consanguinité et d'apparentement, ce qui n'est pas notre cas.

La variance moyenne est de 1,180. Cette valeur est une mesure statistique très arbitraire, elle ne nous donne pas d'information directe sur les arbres généalogiques des individus qui composent l'effectif si ce n'est que le déséquilibre entre le côté maternel et paternel n'est pas très élevé, une entropie de zéro indiquant l'absence de déséquilibre.

## **5.2 Analyse descriptive des variables sélectionnées pour être soumises aux régressions logistiques**

### **5.2.1 Distribution des fréquences pour les variables génétiques**

La distribution des fréquences pour les variables génétiques totales, proches et éloignées (apparentement entre les parents, consanguinités maternelle et paternelle) selon les classes créées déjà mentionnées est présentée dans le tableau 5.1.

Tableau 5.1 : Distribution de la progéniture selon les classes d'apparentement et de consanguinité des parents pour les variables génétiques totales, proches et éloignées.

classes des variables	Effectifs pour chaque variable génétique								
	coef p	cngh	cngf	app44	cngh5 5	cngf5 5	cél1	chél 1	cfél1
1	231	411	420	642	821	814	327	479	491
2	17	3	2	0	0	0	35	7	2
3	477	351	375	207	94	121	563	383	380
4	231	224	170	111	81	48	82	133	123
5	39	7	23	39	5	12	2	2	9
6	11	12	19	7	7	14	0	5	4
7	3	1	0	3	1	0	0	0	0
8	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0
total	100 9	100 9	100 9	1009	1009	1009	100 9	100 9	100 9

Les classes les plus élevées sont vides mais ont quand même été créées puisque la classe 10 correspond à un lien social connu, celui de tante-neveu ou encore oncle-nièce. Comme il a été mentionné dans la section 3.1, l'apparentement moyen à l'Île-aux-Coudres est de 0,012 et la consanguinité moyenne est 0,015 (selon l'étude de Philippe, 1969), cependant, dans l'effectif sélectionné pour la présente étude, ces moyennes sont respectivement de 0,009 et 0,007. Cette différence marquée peut être attribuable à la très sévère sélection de l'effectif. Une partie de la population a donc été exclue des calculs. Il semblerait qu'une partie de la population exclue possédait un nombre élevé de liens d'apparentement et de consanguinité. Les résultats de modélisation obtenus s'appliquent donc à l'effectif

lui-même et leur élargissement à l'ensemble de la population demande de considérer toutes les exclusions qui ont été faites lors de l'application des critères d'exclusion. En effet, le passage de 2000 à 308 couples (et par la suite à 267 lorsque les 41 familles complètes sans date de décès sont exclues) permet de mieux contrôler certaines variables et d'orienter la recherche, mais il entraîne également une perte importante d'information. Cette différence entre les coefficients d'apparentement et de consanguinité entre la population totale et l'effectif en sont la preuve. Ainsi, nous travaillons avec un effectif moins apparenté et consanguin que la population dont il est issu. Ceci constitue un biais de sélection dont il faudra tenir compte lors de l'interprétation des résultats.

#### 5.2.2 Distribution des variables démographiques

Les autres variables soumises au modèle sont : le nombre d'enfants du couple, le rang de naissance de chaque enfant, l'année de naissance de l'enfant, l'âge de la mère à la naissance, l'intervalle génésique avec l'aîné et la différence d'âge entre les conjoints (parents de l'enfant).

Tout d'abord, le nombre d'enfants par couple varie entre 1 et 19. En moyenne, le nombre d'enfants est de 9,531 avec un écart-type de 2,73. Pour les garçons, le nombre moyen d'enfants de sexe masculin par famille est de 4,591, écart-type de 2,074. Le nombre de garçons par famille varie de zéro à douze. Chez les filles de l'effectif, le nombre moyen par famille est de 4,823 avec un écart-type de 2,043. Le nombre de filles par famille varie entre zéro et onze. Il est à noter que 23 familles n'ont aucun garçon et seulement 10 familles n'ont aucune fille. De plus, si l'on regarde la proportion de garçons par famille, on voit un rapport de masculinité de  $4,591/4,823 = 0,952$ . Dans la majorité des populations, le rapport de masculinité rapport est évalué à 1,05 (il naît plus de garçons) (Boisvert, 1992). Ce rapport est inversé à l'Île-aux-Coudres puisqu'il naît moins de garçons que de filles. La distribution de ces variables est présentée en appendice, figure C.2 a et b.

Rappelons que les variables concernant le nombre d'enfants au sein de la famille se rapportent au couple. Etant donné que parmi les exclus, 41 familles sont soustraites au complet, toute information les concernant est elle aussi absente des analyses sur les 1009 individus de l'effectif.

Le rapport de masculinité inversé ne tient pas compte des enfants dont le sexe n'est pas spécifié (X) dans les registres et qui représente environ 3% des 1009 individus de l'effectif. Le nombre moyen d'enfants au sexe non-identifié par famille est de 0,118 avec un écart-type de 0,400 (voir la figure C.6). En effet, si la variable observée est le sexe de chaque individu plutôt que la variable « nombre de filles » ou « nombre de garçons » de la famille, on voit que, dans l'effectif, les proportions de garçons et de filles (hommes et femmes) se ressemblent (voir la figure J.2). Ce résultat étrange sera discuté dans le chapitre VI, section 6.3. Afin d'obtenir plus d'information sur ces individus au sexe inconnu, la figure C-7 représente l'âge moyen au décès des filles (11 668,3 jours), des garçons (12 075,5 jours) et de ces individus sans sexe (0 jours). Ces individus sont donc des morts-nés.

La distribution de la variable rang de naissance de l'enfant a également été faite. Cette variable a une valeur moyenne de 5,093 avec un écart-type de 3,132. Il est normal de voir une telle moyenne puisque le nombre moyen d'enfants par famille est d'environ une dizaine; le rang de naissance moyen correspond à peu près à la moyenne entre 1 et 10. La distribution de cette variable est présentée en appendice à la figure C.4.

En ce qui concerne les variables « différence d'âge entre les conjoints », « âge de la mère à la naissance de l'enfant » et « intervalle génésique entre l'enfant et son aîné immédiat » leur distribution est également présentée en appendice à la figure C.5 a et b. On constate qu'en moyenne, la différence d'âge entre les parents est de 4,543 ans. L'écart-type est très grand, 3,627. En regardant la distribution, on voit que la moyenne a dû être augmentée par la présence d'une exception où la

différence d'âge entre les époux était de plus de vingt ans. C'est vraisemblablement pour cette raison que l'écart-type est si élevé. L'âge moyen pour mettre au monde un enfant dans l'effectif est de 30,855 ans pour la mère. L'écart-type est de 6,628 ans. Il est intéressant de remarquer que l'âge le moins élevé est de 17 ans (moins de 0,5% des naissances) et l'âge le plus élevé est de 49 ans. L'intervalle génésique est présenté en jours. Ainsi, pour l'effectif sélectionné, l'intervalle génésique moyen est de 724,764 jours soit presque exactement deux ans ( $724,764/365,25=1,98$ ). L'écart-type est de 285,155 jours soit environ 9 mois. Compte tenu des critères de sélection appliqués sur les couples (critères de Henry), cela ne surprend pas.

La variable « année de naissance » a été regroupée par tranches (classes) de 10 ans pour sa description. La figure D.1 en appendice présente donc la distribution des années de naissance par tranches de 10 ans. L'effectif sélectionné initialement s'étendait de 1800 à 1940 pour l'année de mariage des parents. Il est donc normal de voir des naissances de 1800 jusqu'à 1940 et, en plus, des naissances plus récentes que 1940 (puisqu'elles arrivent après le mariage). Dans l'effectif, les tranches d'années pour les naissances s'étendent de 1800 à 1959. Selon la distribution de fréquences présentée en appendice (figure D.1), on peut voir que les décennies où l'on observe le plus de naissances pour notre effectif sont situées au milieu de la distribution, soit à partir de 1860-1869 (99 naissances) jusqu'à 1900-1909 (97 naissances). En fait, la somme des naissances de ces périodes ( et des périodes comprises dans l'intervalle) est de 488 individus, ce qui représente près de la moitié de l'effectif de la progéniture conservée pour les analyses. Rappelons également que l'année de mariage moyenne pour les parents des individus faisant partie de l'effectif est 1869,292 avec un écart-type de 34,104 (voir la figure J.2 b). Il est donc normal que la majorité des naissances s'inscrive peu après la date de mariage moyenne des géniteurs de l'effectif.

Le tableau 5.2 présente un résumé des résultats observés sur la distribution des variables « différence d'âge entre les conjoints », « âge de la mère à la naissance de l'enfant », « intervalle génésique entre l'enfant et son aîné immédiat », « rang de naissance » ainsi que « nombre d'enfants total, masculins, féminins et de sexe inconnu dans la famille » dans les modèles de régressions logistiques qui suivront.

Tableau 5.2 : Moyennes et écart-types de certaines variables démographiques pour l'ensemble de l'effectif (1009 individus) sur toute la période étudiée.

Variable	moyenne	écart-type
différence d'âge entre les conjoints, en années (difconj)	4,543	3,627
âge de la mère à la naissance de l'enfant, en années (agemn)	30,855	6,628
intervalle génésique entre la naissance et celle de l'aîné immédiat, en jours (interidi)	724,764	285,155
rang de naissance de l'enfant (rangnai)	5,093	3,132
nombre d'enfants (nbenf)	9,531	2,973
nombre d'enfants de sexe masculin (nbenfm)	4,591	2,074
nombre d'enfants de sexe féminin (nbenff)	4,823	2,043
nombre d'enfants de sexe inconnu (nbenfx)	0,118	0,400

### 5.2.3 Tables de contingences pour qualifier la variable année de naissance

Finalement, l'année de naissance de l'enfant a été l'objet d'une analyse un peu plus poussée. Afin de bien voir la distribution de la proportion de mortalité de la progéniture (pour chaque catégorie de mortalité), des tables de contingences ont été faites pour voir si les proportions de mortalité (par rapport aux naissances) étaient comparables d'une classe d'années à l'autre. Ces tables de contingences ont été faites pour la mortinatalité, la mortalité néonatale précoce, la mortalité infantile, la mortalité entre 1 et 4 ans et la mortalité pré-reproductive. De plus, elles ont été répétées pour la progéniture totale, masculine et féminine. Les tables de contingences sont présentées en appendice dans les figures D.2 à D.16. Dans ces figures, on présente les classes d'années horizontalement. À la verticale, on voit les différents types de mortalité et leurs issues possibles (mort ou survie). La catégorie « survie » est toujours celle de la ligne qui commence par « 0 ». Pour la mortinatalité, la mort de l'individu correspond à la ligne « 1 ». Ce même événement (la mort) est représenté respectivement par « 2 », « 3 », « 4 » et « 5 » pour la mortalité néonatale précoce, la mortalité infantile, la mortalité entre 1 et 4 ans et la mortalité pré-reproductive.

Pour toutes les tables de contingences effectuées, il faut se fier au test de Pearson pour voir s'il existe une différence dans les proportions de mortalités de la progéniture d'une tranche d'années à l'autre. Ce test sert à évaluer la distribution des variables évaluées (Sall et Lehman, 1996). En général, les tests de contingence révèlent qu'il existe une différence significative entre les proportions de mortalité d'une classe d'année à l'autre. Les seuls tableaux pour lesquels aucune différence n'a été notée sont ceux qui étudient la mortinatalité, la mortalité néonatale, la mortalité entre 1 et 4 ans ainsi que la mortalité pré-reproductive féminine et ceux qui concernent la mortalité entre 1 et 4 ans et la mortalité pré-reproductive masculine. Dans ces cas, aucune différence significative entre les classes d'années ne peut être dénotée.

Quand le test de Pearson est significatif, il indique que le rapport n'est pas équivalent d'une classe à l'autre sans spécifier une tendance particulière (on ne sait pas quelles classes diffèrent). Cependant, par simple observation on peut déceler une légère tendance à l'augmentation au fil des décennies et ce particulièrement pour les catégories de mortalité où l'on voit plus de mortalité comme pour la mortalité infantile ou entre 1 et 4 ans. Toutefois, ce ne sont que des observations de tendances.

Par exemple, dans la figure D.4, illustrant la distribution de la proportion de mortalité infantile pour chaque tranche de 10 ans, on peut voir qu'entre 1820 et 1829, sur les 63 naissances, 14 sont morts dans la première année. En comparaison, entre 1920 et 1929, l'effectif des naissances est comparable, 64 individus, et les mortalités sont de 21 individus, une augmentation de 50%! De même, entre 1860 et 1869, 99 naissances et seulement 9 décès tandis qu'entre 1900 et 1909, 97 naissances et 33 décès.

En plus de ce phénomène qui semble indiquer de façon générale une augmentation du nombre de mortalités au fil des années, rappelons que l'effectif sélectionné, le groupe des « inclus », est formé par les individus dont la date de mariage moyenne (1869,292) des parents est la plus ancienne comparativement aux individus des deux autres groupes (les « exclus » et les « couples ») qui ont été retirés de l'étude car ils n'avaient pas de date de décès. Malgré la nécessité de cette exclusion, il est possible que la soustraction ait entraîné une surestimation des mortalités pour les tranches d'années récentes. L'ensemble de ces observations pourra peut-être orienter l'interprétation des résultats en ce qui concerne l'année de naissance pour les individus nés dans chaque tranche de temps.

### 5.3 Présentation des modèles de régression logistique

Pour chaque modèle présenté, les valeurs de rapports de cotes (facteur de risque multiplicatif) et le niveau de signification de chaque variable sont fournis dans le tableau correspondant. Il est important de spécifier qu'une variable doit avoir un seuil de signification inférieur ou égal à  $\text{sig}=0,05$  pour être retenue par la sélection FORWARD.

Les mêmes modèles de régression logistique ont aussi été soumis à la méthode « ENTER ». Ce mode de sélection considère toutes les variables comme faisant partie du modèle, indépendamment de leur seuil de signification. C'est alors au chercheur de choisir son propre seuil de signification et de faire ses analyses conséquemment. Pour simplifier les analyses, c'est toujours le seuil de 0,05 (très courant dans les analyses statistiques) qui sera respecté dans cette étude.

En s'attardant à ces modèles, on voit tout d'abord une forte similitude entre eux et les premiers, formés avec la sélection FORWARD. Cependant, les corrélations entre toutes les variables présélectionnées font varier quelque peu les seuils de signification. Ce mode d'analyse a pour inconvénient de surestimer l'impact du modèle puisqu'il considère toutes les variables, même les moins significatives, comme ayant un impact sur le type de mortalité étudié. Ce problème rend les résultats moins fiables. Dans le cadre de notre recherche, le mode de sélection ENTER possède toutefois une qualité indéniable, celle de tout montrer au lieu de sélectionner seulement les variables dont le seuil atteint ou dépasse 0,05. De cette manière, on peut voir si une variable a un impact important même si celui-ci n'est pas statistiquement significatif. Cet avantage nous rapproche un peu plus de la réalité où toutes les variables, même la moins importante, jouent un rôle dans le destin de mort ou de vie d'un enfant.

Comme il a été mentionné, les principales interactions entre les variables et la colinéarité trop forte ont été évitées dans les modèles par une pré-sélection. Toutefois, il demeure évident que les variables indépendantes sélectionnées présentent encore des interactions. La régression logistique se prête quand même bien à cette analyse vu sa résistance à la colinéarité (Kleinbaum, Kupper et Muller, 1987).

L'objet de cette étude étant en premier lieu l'impact de la consanguinité maternelle (et en second lieu, l'impact de l'apparentement et de la consanguinité paternelle) sur les risques de mortalité de leurs enfants, nous ferons référence aux modèles ENTER (tableaux de l'appendice F) chaque fois qu'un impact différentiel des variables génétiques y sera dénoté.

### 5.3.1 Classes de mortalités touchant la progéniture totale

#### 5.3.1.1 Modèles soumettant les variables génétiques totales.

Le tableau 5.3 englobe les modèles visant la mortalité de la progéniture totale (garçons et filles) et démontre les rapports de cotes obtenus et les seuils de signification obtenus par les variables indépendantes préalablement énumérées sélectionnées selon la méthode FORWARD. Dans ce premier cas, les cinq modèles possèdent les valeurs d'apparentement total et de consanguinités (maternelle et paternelle) totales.

Les rapports de cotes et les seuils de signification seront fournis pour tous les tableaux présentant des modèles de régression logistique (tableaux 5.3, 5.4, 5.5, 5.6, 5.7, 5.8, 5.9, 5.10 et 5.11), toutefois, pour ne pas alourdir les tableaux, cette information ne sera mentionnée que pour le tableau 5.3. Il est également important de préciser que les lettres « NS » signifient que la variable est non significative

dans le modèle et n'en fait pas partie. Ces variables ne sont retranscrites dans les tableaux que par souci d'uniformité d'un tableau à l'autre.

Tableau 5.3 : Modèles de régressions logistiques effectués par sélection FORWARD touchant la progéniture totale et soumettant les variables génétiques totales.

variables dans le modèle(!)	types de mortalité associés à leur numéro de modèle				
	infantile (1)	mortinatalité (10)	néonatale précoce (19)	1-4 ans (28)	pré-reproductive (37)
nbenf	<b>0,947<sup>1</sup></b> <b>0,036<sup>2</sup></b>	<b>0,868</b> <b>0,001</b>	<b>0,806</b> <b>0,002</b>	NS	NS
difconj	<b>0,950</b> <b>0,028</b>	NS	NS	NS	NS
rangnai	NS	NS	NS	NS	<b>1,053</b> <b>0,042</b>
year	<b>1,014</b> <b>0,000</b>	<b>1,019</b> <b>0,000</b>	<b>1,023</b> <b>0,005</b>	<b>1,007</b> <b>0,021</b>	<b>1,008</b> <b>0,001</b>
agemn	<b>1,038</b> <b>0,005</b>	NS	NS	NS	NS
interidi	<b>0,999</b> <b>0,000</b>	NS	NS	NS	NS
xcoefp	NS	NS	NS	NS	NS
xcngh	NS	NS	NS	NS	NS
xcngf	NS	NS	<b>0,664</b> <b>0,044</b>	NS	NS
constante	0,000	0,000	0,005	0,008	0,000

(!): La liste des noms complets des variables associés à leur acronyme est fournie en appendice dans le tableau A.1.

<sup>1</sup> et <sup>2</sup>: rapport de cote et seuil de signification.

La deuxième colonne présente la modélisation de la mortalité infantile. Les variables retenues comme significativement différentes de zéro dans ce modèle sont le nombre d'enfants du couple (nbenf) avec un sig= 0,036 (seuil de signification), la différence d'âge entre les conjoints (difconj) avec sig=0,028, l'année de naissance de l'enfant (year) avec sig=0,000, l'âge de la mère à la

naissance de l'enfant (agemn) avec  $\text{sig}=0,005$  et l'intervalle en jours séparant l'enfant de son aîné immédiat (interidi) avec  $\text{sig}=0,000$ . Les rapports de cote respectifs de ces variables significatives sont: 0,947, 0,950, 1,014, 1,038 et 0,999.

Étant donné le critère d'exclusion des variables dont le coefficient n'est pas significativement différent de zéro (celles dont le seuil de signification est supérieur à 0,05) appliqué par la sélection FORWARD, toutes les variables qui sont acceptées dans les différents modèles sont significativement différentes de zéro. Pour cette raison et afin de ne pas alourdir la lecture, les seuils de significations des variables incluses dans les modèles subséquents ne seront pas mentionnés. Il suffit au lecteur de se référer au tableau correspondant au modèle à l'étude pour vérifier le seuil de signification d'une variable.

Le rapport de cotes sert à donner un indice de l'augmentation du risque de mortalité en fonction de l'augmentation de la variable indépendante. Le risque de mortalité infantile (#1) est augmenté par l'augmentation de l'année de naissance (le rapport de cote multiplie par 1,014 le risque de mortalité infantile pour chaque année ajoutée au calendrier). Ce rapport, ainsi que celui de "interidi" sont presque neutres. En effet, en ce qui concerne "interidi", l'augmentation d'une unité de cette variable multiplie par 0,999 le risque de mortalité infantile indépendamment du sexe de l'enfant. Ces deux derniers rapports de cotes sont si près de 1,0000 (facteur neutre) que leur effet est minime. En ce qui concerne le nombre d'enfants de la famille, l'ajout d'un enfant supplémentaire multiplie par 0,0947 le risque de mortalité pour un enfant donné au sein de cette même famille. Plus la taille de la famille est grande, plus chacun des enfants qui la compose a de bonnes chances de survie. Pour la variable différence d'âge entre les conjoints, chaque unité d'augmentation diminue le risque de mortalité infantile par un facteur multiplicatif de 0,950. C'est donc dire que plus les conjoints sont d'âges disparates, plus la progéniture a de chances de survie lors de sa première année d'existence. Cependant, plus la mère a son enfant à un âge avancé, plus les risques de mortalité infantile de ce dernier

sont grands puisque l'augmentation de l'âge de la mère à la naissance de son enfant multiplie ce risque par 1,038. Ce facteur de risque n'est pas, lui non plus, très élevé mais il demeure toutefois significatif.

Les autres variables entrées dans le modèle (rangnai, xcoefp, xcngf ainsi que xcngf) n'ont pas des coefficients significativement différents de zéro. C'est donc dire que leur présence ne peut pas être considérée comme ayant un impact mesurable sur la survie lors de la première année de vie.

Dans le deuxième et le troisième modèles (#10 et 19), on retrouve également que le nombre d'enfants de la famille a un impact positif sur la survie de l'enfant (r.c.= 0,868 pour la mortinatalité et 0,806 pour la mortalité néonatale précoce. Le nombre d'enfants total du couple porte donc, pour la mortalité périnatale (Péron et Strohmenger, 1985) et la mortalité infantile, une garantie de survie rétrospective puisque qu'à moins que l'enfant dont le risque de mortalité est mesuré soit le dernier né de sa famille, le nombre de ses frères et soeurs n'est pas encore arrêté.

Une autre variable commune aux trois premiers modèles est l'année de naissance de l'enfant. Ses rapports de cotes de 1,019 et 1,023 respectivement pour la mortinatalité et la mortalité néonatale précoce. Tout comme pour la mortalité infantile, cette variable a un impact presque neutre mais est toujours significativement différente de zéro dans le modèle. On peut donc dire qu'elle a un impact minime mais significatif sur les risques de mortalité dans la première année de vie et ce dès la naissance de l'enfant.

Finalement, en ce qui concerne la mortalité néonatale précoce (#19), la consanguinité totale de la mère est significativement très protectrice (r.c. = 0,664) contrairement aux modèles de mortinatalité et de mortalité infantile.

Les modèles de mortalité entre un et quatre ans et de mortalité pré-reproductive (#28 et 37) font passer les facteurs de risque potentiels dans une nouvelle catégorie : l'individu ayant survécu à sa première année de vie, c'est sa survie à la petite enfance et à l'adolescence qui nous intéresse. En ce qui concerne ces deux modèles pour la progéniture totale par rapport à l'apparentement total et à la consanguinité totale respective de chacun des parents, l'année de naissance de l'enfant ressort comme un facteur significativement néfaste. Encore une fois, ce risque est très faible. Les rapports de cotes sont de 1,007 et 1,053 pour la mortalité entre un et quatre ans et la mortalité pré-reproductive, respectivement. Pour la mortalité pré-reproductive, le rang de naissance atteint pour la première fois le seuil de signification lui permettant d'entrer dans le modèle avec  $\text{sig}=0,042$  et met en valeur un risque multiplicatif de 1,053 comme rapport de cotes qui signifie que plus un enfant a d'aînés, plus ses risques de mortalité entre un et quatorze ans augmentent.

Dans les modèles correspondants faits par sélection ENTER, on peut remarquer quelques variations intéressantes. En effet, pour la mortalité infantile, la consanguinité maternelle totale avait, en présence de toutes les autres variables, un  $\text{sig}= 0,0130$  et présentait un  $r.c.= 0,8525$ . Cependant, lorsqu'on a procédé à la sélection FORWARD, l'impact protecteur de cette variable n'a pas atteint un seuil significatif. On ne peut cependant ignorer la contribution significative de la consanguinité maternelle totale en présence de toutes les autres variables d'exposition.

## 5.3.1.2 Modèles soumettant les variables génétiques proches

Tableau 5.4: Modèles de régressions logistiques effectués par sélection FORWARD touchant la progéniture totale et soumettant les variables génétiques proches.

variables dans le modèle(!)	types de mortalité associés à leur numéro de modèle				
	infantile *(2)	mortinatalité *(11)	néonatale précoce (20)	1-4 ans (29)	pré-reproductive (38)
nbenf	<b>0,947</b> <b>0,036</b>	<b>0,868</b> <b>0,001</b>	<b>0,807</b> <b>0,002</b>	NS	NS
difconj	NS	NS	NS	NS	NS
rangnai	NS	NS	NS	NS	NS
year	<b>1,014</b> <b>0,000</b>	<b>1,019</b> <b>0,000</b>	<b>1,022</b> <b>0,007</b>	NS	<b>1,011</b> <b>0,000</b>
agemn	<b>1,038</b> <b>0,005</b>	NS	NS	NS	NS
interidi	<b>0,999</b> <b>0,000</b>	NS	NS	NS	NS
xapp44	NS	NS	NS	NS	<b>0,855</b> <b>0,017</b>
xcngh55	NS	NS	NS	NS	NS
xcngf55	NS	NS	NS	<b>1,253</b> <b>0,003</b>	NS
constante	0,000	0,000	0,005	0,000	0,000

(!): La liste des noms complets des variables associés à leur acronyme est fournie en appendice dans le tableau A.1.

\* Modèle identique à son homologue soumettant les variables génétiques totales.

Ces cinq modèles (#2, 11, 20, 29 et 38) s'apparentent aux cinq précédents. Leurs structures sont identiques mais les variables génétiques totales y sont remplacées par les variables génétiques proches. Les deux premiers modèles (mortalité infantile, #2 et mortinatalité, #11) sont identiques dans ce tableau et le précédent (tableau 5.3) puisque les variables d'ordre génétique n'ont pas atteint le seuil de

signification. Puisqu'ils ont déjà été étudiés, il n'en sera plus question dans cette section.

Pour la mortalité néonatale précoce (#20), on voit que la composante génétique du risque de mortalité est disparue lors du remplacement des valeurs d'apparentement total et de consanguinités totales par les valeurs d'apparentement proche et de consanguinités proches. Le nombre total d'enfants du couple a encore un impact sur la survie de chacun de ces enfants; une augmentation de ce nombre multipliant le risque de mortalité par 0,807. Passé le cap de la première année, le nombre d'enfants au sein de la famille ne semble toutefois plus avoir d'impact significatif sur la mortalité de la progéniture.

On observe encore que la variable « année de naissance » entraîne un facteur significativement néfaste à la survie de la progéniture; les rapports de cotes pour les modèles de mortalité néonatale (#20) et pré-reproductive (#38) sont respectivement 1,022 et 1,011. Comme pour les modèles englobant les consanguinités totales et l'apparentement total, ces rapports de cotes concernant la variable « année de naissance » comme facteur de risque sont très faibles et presque neutres malgré le fait qu'ils sont considérés comme significatifs par le modèle.

On retrouve maintenant un impact intéressant de l'apparentement proche des parents et la consanguinité proche de la mère pour les mortalités entre un et quatre ans et pré-reproductive. Pour la mortalité pré-reproductive, l'augmentation de l'apparentement proche du couple favorise significativement les chances de survie des individus ( $r.c.=0,855$ ) tandis que l'augmentation de la consanguinité proche de la mère augmente les chances de mortalité entre un et quatre ans ( $r.c.=1,253$ ). Il est fort intéressant de noter ici que dans les modèles correspondants faits par sélection ENTER (présentés en appendice), l'apparentement proche entre les parents et la consanguinité maternelle proche présentaient tous deux des seuils de

signification inférieurs à 0,05 pour les risques de mortalité entre un et quatre ans et de mortalité pré-reproductive. En présence de toutes les autres variables d'exposition, ces deux variables de risques ressortaient donc de façon significative dans chacun des deux modèles de mortalité. On peut donc croire que, bien que l'impact de ces variables sur chacun de ces types de mortalité diffère un peu, l'apparentement proche des parents ainsi que la consanguinité maternelle proche affectent tous deux le risque de mortalité entre un et quatre ans et entre un et quatorze ans. L'apparentement proche, pour la mortalité dans ces groupes d'âges, représente toujours une protection (r.c.= 0,8262 et 0,8372) tandis que l'augmentation de la consanguinité maternelle proche dénote un risque accru (r.c.= 1,2371 et 1,1522).

La variable « rang de naissance » de l'enfant n'a pas atteint le seuil de signification dans le modèle de mortalité pré-reproductive mettant en jeu les variables génétiques proches, contrairement au modèle utilisant les variables génétiques totales.

## 5.3.1.3 Modèles soumettant les variables génétiques éloignées

Tableau 5.5 : Modèles de régressions logistiques effectués par sélection FORWARD touchant la progéniture totale et soumettant les variables génétiques éloignées.

variables dans le modèle(!)	types de mortalité associés à leur numéro de modèle				
	infantile *(3)	mortinatalité *(12)	néonatale précoce **(21)	1-4 ans *(30)	pré-reproductive *(39)
nbenf	0,947 0,036	0,868 0,001	0,807 0,002	NS	NS
difconj	NS	NS	NS	NS	NS
rangnai	NS	NS	NS	NS	1,053 0,042
year	1,014 0,000	1,019 0,000	1,022 0,007	1,007 0,021	1,008 0,001
agemn	1,038 0,005	NS	NS	NS	NS
interidi	0,999 0,000	NS	NS	NS	NS
xcél1	NS	NS	NS	NS	NS
xché1	NS	NS	NS	NS	NS
xcfé1	NS	NS	NS	NS	NS
constante	0,000	0,000	0,005	0,008	0,000

(!): La liste des noms complets des variables associés à leur acronyme est fournie en appendice dans le tableau A.1.

\* Modèle identique à son homologue soumettant les variables génétiques totales.

\*\* Modèle identique à son homologue soumettant les variables génétiques proches.

Pour ces modèles, où l'on a intégré les valeurs d'apparentement éloigné et de consanguinité éloignée des parents, on peut voir que les variables significatives, outre celles d'apparentement et de consanguinité, sont sensiblement les mêmes que dans les deux séries précédentes. En fait, les modèles touchant la mortalité infantile, la mortinatalité, la mortalité entre un et quatre ans et la mortalité pré-

reproductive (respectivement #3, 12, 30 et 39) sont exactement les mêmes que les modèles englobant les variables d'apparement total et de consanguinité totale.

En résumé, aucune variable d'apparement ou de consanguinité ne semble avoir un impact significativement mesurable ni sur la mortalité néonatale pour la progéniture totale, ni sur la mortalité infantile de cette même progéniture puisqu'aucune d'entre elles n'est apparue dans l'un ou l'autre de ces modèles. Encore une fois, ces résultats similaires sont prévisibles vu la ressemblance entre les trois séries de modèles ici étudiés.

Le modèle de mortalité néonatale précoce ci-haut (apparement éloigné et consanguinités éloignées, #21) est identique au modèle correspondant pour les variables génétiques proches.

Les deux derniers modèles, « mortalité entre un et quatre ans » et « mortalité pré-reproductive », confrontés à des valeurs d'apparement éloigné et de consanguinités éloignées, n'ont retenu aucune de ces variables génétiques et sont identiques à leurs homologues des modèles de variables génétiques totales (comme il a été mentionné plus haut).

C'est donc l'inclusion de l'une ou l'autre des catégories de variables génétiques totales, proches ou éloignées qui fait la différence entre les paramètres soumis. Par conséquent, les variations observées d'un modèle à l'autre sont dues à l'impact différentiel de ces catégories de variables génétiques. Pour la progéniture totale, on voit donc que, globalement, la consanguinité totale de la mère a un impact protecteur sur la survie de l'enfant pour la période néonatale précoce (c'est-à-dire, entre le premier et le sixième jour de vie. Lorsque cette consanguinité maternelle est scindée en ses composantes proche et éloignée, aucun impact ne peut être détecté de façon significative. Si on s'attarde au modèles par sélection ENTER, on

ne voit pas non plus d'impact s'approchant du seuil de signification pour les composantes proches et éloignées de la consanguinité maternelle.

Si on observe maintenant les autres endroits où l'on dénote la présence de variables d'ordre génétique dans les modèles, on voit que la consanguinité proche de la mère présente une menace à la survie pendant la période de un à quatre ans; de plus, l'apparentement proche des parents semble protecteur pour la période pré-reproductive de l'enfant. La présence de l'apparentement proche dans ce dernier modèle semble avoir empêché l'entrée de la variable « rang de naissance de l'enfant ». Cette variable ne fait donc pas partie du modèle mais elle est incluse dans les deux modèles correspondants de mortalités pré-reproductive auxquels on a respectivement soumis les variables génétiques totales et éloignées. Il est important de rappeler que l'ordre d'entrée des variables peut influencer les seuils de signification atteints par chaque variable restante. Étant donné la nature très complexe des liens unissant l'ensemble de ces variables, il est quasi impossible d'ignorer les corrélations. Il est cependant essentiel de tenir compte de leur existence pour l'interprétation du modèle.

### 5.3.2 Classes de mortalités touchant la progéniture masculine

#### 5.3.2.1 Modèles soumettant les variables génétiques totales

Tableau 5.6 : Modèles de régression logistique effectués par sélection FORWARD touchant la progéniture masculine et soumettant les variables génétiques totales.

variables dans le modèle(!)	types de mortalité associés à leur numéro de modèle				
	infantile (5)	mortinatalité (13)	néonatale précoce (22)	1-4 ans (31)	pré-reproductive (40)
nbenf	NS	<b>0,778</b> <b>0,003</b>	NS	NS	NS
difconj	NS	NS	NS	NS	NS
rangnai	NS	NS	NS	NS	<b>1,079</b> <b>0,046</b>
year	<b>1,015</b> <b>0,000</b>	<b>1,082</b> <b>0,000</b>	NS	<b>1,009</b> <b>0,033</b>	<b>1,009</b> <b>0,017</b>
agemn	NS	NS	NS	NS	NS
interidi	NS	NS	NS	NS	NS
xcoefp	NS	NS	NS	<b>0,790</b> <b>0,048</b>	NS
xcngh	NS	NS	NS	NS	NS
xcngf	NS	<b>0,663</b> <b>0,036</b>	NS	NS	NS
constante	0,000	0,000	NS	0,000	0,007

(!): La liste des noms complets des variables associés à leur acronyme est fournie en appendice dans le tableau A.1.

Chez les garçons, on retrouve moins de variables significatives. Pour la mortalité infantile (#5), seule la variable « year » (année de naissance) ressort de façon significative et présente, encore une fois, un risque accru de mortalité avec l'augmentation des années (r.c.= 1,015). Pour la mortinatalité (#13), on voit un effet similaire de la variable « année de naissance » (r.c.=1,082) et on remarque que la consanguinité maternelle totale a un impact significatif sur la mortalité à la

naissance. En effet, l'augmentation de cette consanguinité semble manifester des répercussions protectrices sur la progéniture ( $r.c.=0,6147$ ). Rappelons que la progéniture totale bénéficiait d'une protection similaire de la consanguinité maternelle mais cette protection se manifestait pour la période de mortalité néonatale précoce (au lieu de celle de mortinatalité).

Le nombre d'enfants de la famille protège les garçons de la mortinatalité ( $r.c.=0,778$ ). Cependant, cette variable qui était significative pour les modèles de mortalité infantile et néonatale précoce chez la progéniture totale ne l'est pas pour la progéniture masculine. Ce phénomène fera l'objet d'une explication dans la discussion.

Le modèle de mortalité néonatale précoce avec les variables génétiques totales (#22) est non-significatif. Même la constante d'équation n'a pas réussi à atteindre un seuil de signification inférieur à 0,05. Il est à noter qu'il en est de même pour les modèles de mortalité néonatale précoce masculine auxquels ont été soumises les variables génétiques proches et éloignées (modèles présentés ultérieurement). Ce résultat signifie qu'aucune variable, génétique ou démographique, n'a un impact significatif sur ce risque de mortalité chez les garçons. Il faudra se tourner sur ce qui se passe du côté de la progéniture féminine pour analyser ce type de mortalité.

En ce qui concerne le modèle de mortalité 1-4 ans (#31), on remarque que pour cette catégorie d'âge, l'augmentation de l'apparement des parents a un effet protecteur sur la survie des jeunes garçons ( $r.c.=0,790$ ). De plus, on observe encore une fois que l'année de naissance amplifie les risques de mortalité avec un rapport de cotes de 1,009.

Finalement, comme pour le modèle de mortalité pré-reproductive de la progéniture totale (#37), l'augmentation du rang de naissance chez les garçons diminue les

chances de survie de ceux-ci entre 1 et 14 ans (r.c.=1,079) et l'année de naissance agit encore une fois dans le même sens (r.c.=1,009) (pour le présent modèle, #40).

### 5.3.2.2 Modèles soumettant les variables génétiques proches.

Tableau 5.7: Modèles de régressions logistiques effectués par sélection FORWARD touchant la progéniture masculine et soumettant les variables génétiques proches.

variables dans le modèle(!)	types de mortalité associés à leur numéro de modèle				
	infantile *(6)	mortinatalité (14)	néonatale précoce (23)	1-4 ans (32)	pré-reproductive (41)
nbenf	NS	<b>0,788</b> <b>0,004</b>	NS	NS	NS
difconj	NS	NS	NS	NS	NS
rangnai	NS	NS	NS	NS	NS
year	<b>1,015</b> <b>0,000</b>	<b>1,078</b> <b>0,000</b>	NS	<b>1,011</b> <b>0,023</b>	<b>1,014</b> <b>0,000</b>
agemn	NS	NS	NS	NS	NS
interidi	NS	NS	NS	NS	NS
xapp44	NS	NS	NS	<b>0,588</b> <b>0,000</b>	<b>0,756</b> <b>0,005</b>
xcngh55	NS	NS	NS	NS	NS
xcngf55	NS	NS	NS	<b>1,343</b> <b>0,019</b>	NS
constante	0,000	0,000	NS	0,013	0,000

(!): La liste des noms complets des variables associés à leur acronyme est fournie en appendice dans le tableau A.1.

\* Modèle identique à son homologue soumettant les variables génétiques totales.

Dans les modèles où l'on a tenté d'introduire les variables d'apparentement et de consanguinités proches(#6, 14, 23, 32 et 41), on constate encore que l'augmentation de l'année de naissance est significativement néfaste pour la

survie des garçons. Le modèle de mortalité infantile masculine présenté ci-haut (#6) est identique à celui présenté dans le tableau précédent (#5).

On peut voir que la variable « année de naissance » est significative pour tous les modèles présentés excepté pour le modèle de mortalité néonatale précoce (#23) qui ne présente aucune variable significative, constante d'équation incluse. Ce modèle est donc non-significatif. Les valeurs de rapports de cotes pour l'année de naissance sont respectivement de 1,078, 1,011 et 1,014 pour la mortinatalité, la mortalité entre un et quatre ans et la mortalité pré-reproductive (#14, 32 et 41).

Le nombre d'enfants de la famille marque un rapport protecteur contre la mortinatalité masculine tout comme son modèle homologue avec les variables génétiques totales. Toutefois, contrairement à ce dernier modèle où la consanguinité maternelle totale avait démontré un rôle protecteur avec un rapport de cotes de 0,663, on ne retrouve aucune variable d'ordre génétique dans le présent modèle. Ceci laisse penser que les valeurs d'apparentement proche et de consanguinités proches n'ont pas d'impact mesurable sur ce type de mortalité masculine.

Encore une fois, dans les modèles de mortalité entre un et quatre ans et de mortalité pré-reproductive, on voit que l'apparentement des parents, plus particulièrement, l'apparentement proche (« xapp44 ») dans ces modèles, présente une tendance protectrice significative avec des rapports de cotes de 0,588 et 0,756 respectivement. Le premier de ces deux rapports de cotes (0,588) est très élevé puisque l'augmentation d'une unité diminue presque de moitié le risque de mortalité! Finalement, la consanguinité maternelle proche semble avoir un impact mesurable et dangereux sur la survie des jeunes garçons entre 1 et 4 ans. Elle présente en effet un rapport de cotes de 1,343.

Les modèles de mortalité entre un et quatre ans et de mortalité pré-reproductive faits avec le mode ENTER présentent les mêmes variables génétiques significatives. Les influences de l'apparentement proche entre les conjoints et la consanguinité maternelle proche semblent donc beaucoup plus solides dans les modèles touchant seulement la mortalité des garçons que ceux visant l'ensemble de la progéniture. En effet, ces variables sont significatives dans les deux types de sélections, indépendamment des autres variables sélectionnées et de leurs possibles interactions. Il est normal que les modèles masculins soient plus fiables que ceux touchant garçons et filles puisque la composante « fille » a donné des modèles non-significatifs pour ces catégories de mortalité.

## 5.3.2.3 Modèles soumettant les variables génétiques éloignées.

Tableau 5.8: Modèles de régressions logistiques effectués par sélection FORWARD touchant la progéniture masculine et soumettant les variables génétiques éloignées.

variables dans le modèle(!)	types de mortalité associés à leur numéro de modèle				
	infantile *(4)	mortinatalité **(15)	néonatale précoce (24)	1-4 ans (33)	pré-reproductive *(42)
nbenf	NS	0,788 0,004	NS	NS	NS
difconj	NS	NS	NS	NS	NS
rangnai	NS	NS	NS	NS	1,079 0,046
year	1,015 0,000	1,078 0,000	NS	1,009 0,051	1,009 0,017
agemn	NS	NS	NS	NS	NS
interidi	NS	NS	NS	NS	NS
xcél1	NS	NS	NS	NS	NS
xché1	NS	NS	NS	NS	NS
xcfé1	NS	NS	NS	NS	NS
constante	0,000	0,000	NS	0,028	0,007

(!): La liste des noms complets des variables associés à leur acronyme est fournie en appendice dans le tableau A.1.

\* Modèle identique à son homologue soumettant les variables génétiques totales.

\*\* Modèle identique à son homologue soumettant les variables génétiques proches.

En ce qui concerne la mortalité infantile masculine, les trois modèles testés (#5, 6 et 4 selon leur ordre de présentation dans cette section) sont similaires et ne présentent aucun impact des variables d'ordre génétiques, quelles qu'elles soient. Les trois modèles sont donc identiques, ne possédant que l'année de naissance comme variable significative.

Pour les modèles de mortalité, ce n'est pas tout à fait le même scénario qui est observé. En effet, la consanguinité maternelle totale (dans le modèle #13) a un impact protecteur mais ses composantes proche et éloignée (dans les modèles #14 et 15 respectivement) prises séparément n'atteignent pas le seuil de signification.

Le modèle de mortalité masculine néonatale précoce (#24) est non-significatif tout comme ses deux modèles homologues de mortalité masculine (#22 et 23). Comme il a déjà été mentionné précédemment, les variables sélectionnées pour cette étude ne permettent vraisemblablement pas de bien mettre en lumière les expositions réelles auxquelles font face les garçons pendant les premiers jours de leur existence.

Le modèle de mortalité masculine entre un et quatre ans (#33), auquel on a soumis les variables génétiques éloignées, est intéressant. Le premier modèle présenté avait sélectionné l'année de naissance comme significative avec en plus l'apparentement total des parents. Le rapport de cotes obtenu par l'apparentement est relativement fort et protecteur ( $r.c.= 0,790$ ). Ce même modèle avec des variables génétiques proches démontre encore l'impact de l'année de naissance mais une fois de plus, c'est l'apparentement proche qui a l'influence la plus marquée sur le risque de mortalité entre un et quatre ans avec un rapport de cotes très fort ( $r.c.= 0,588$ ). Ce dernier résultat laisse supposer que c'est la part d'apparentement proche de l'apparentement total préalablement estimé qui dicte cet effet protecteur. Lorsqu'on regarde le présent modèle où l'on a tenté d'évaluer l'impact des variables génétiques éloignées, on dénote immédiatement l'absence de l'apparentement éloigné dans le modèle. La seule variable significative est l'année de naissance qui présente comme à l'habitude une faible augmentation de risque avec son augmentation ( $r.c.= 1,009$ ). Cette variable atteint à peine le seuil de signification avec  $sig= 0,051$ .

Pour la mortalité pré-reproductive (#42), on observe un modèle identique au premier modèle (avec les variables génétiques totales). Un fait intéressant : les modèles de mortalité pré-reproductive de la progéniture masculine se comportent exactement comme ceux de la mortalité pré-reproductive de la progéniture totale. Les modèles incluent dans leurs variables significatives soit le rang de naissance de l'enfant ( ou du garçon) comme facteur de risque de mortalité, soit l'apparement proche entre les parents comme facteur de protection mais jamais ces deux variables ne se trouvent ensemble dans un modèle.

### 5.3.3 Classes de mortalités touchant la progéniture féminine

#### 5.3.3.1 Modèles soumettant les variables génétiques totales

Tableau 5.9 : Modèles de régression logistique effectués par sélection FORWARD touchant la progéniture féminine et soumettant les variables génétiques totales.

variables dans le modèle(!)	types de mortalité associés à leur numéro de modèle				
	infantile (7)	mortinatalité (16)	néonatale précoce (25)	1-4 ans (34)	pré-reproductive (43)
nbenf	NS	NS	0,829 0,016	NS	NS
difconj	NS	NS	NS	NS	NS
rangnai	NS	NS	NS	NS	NS
year	1,021 0,000	1,026 0,023	1,034 0,003	NS	1,008 0,019
agemn	NS	NS	NS	NS	NS
interidi	NS	NS	NS	NS	NS
xcoefp	1,294 0,010	NS	NS	NS	NS
xcngh	NS	NS	NS	NS	NS
xcngf	0,837 0,047	NS	NS	NS	NS
constante	0,000	0,014	0,002	NS	0,011

(!): La liste des noms complets des variables associés à leur acronyme est fournie en appendice dans le tableau A.1.

Si on s'attarde maintenant au sort des filles, on voit que, pour les modèles contenant les valeurs d'apparement total et de consanguinités totales (#7, 16, 25 et 43), l'année de naissance de la fille compte encore parmi les variables qui passent le seuil de signification pour tous les modèles présentés excepté celui touchant la mortalité entre un et quatre ans(#34). Ce dernier modèle est non-significatif. Chez les filles comme pour les garçons et la descendance totale, plus

l'année de naissance est récente dans l'intervalle sélectionné, plus le risque de mortalité augmente. De plus, le nombre d'enfant de la famille a un impact protecteur sur la survie de la progéniture féminine dans le modèle représentant la période de mortalité néonatale précoce. Ce phénomène avait également été observé pour la progéniture totale (#19).

Dans le modèle touchant la mortalité infantile chez les filles, certaines variables d'ordre génétique ont également inscrit des rapport de cotes significatifs. Il s'agit de l'apparement total du couple et de la consanguinité maternelle totale. Tout d'abord, l'augmentation de l'apparement entre les parents semble entraîner une risque accru de la mortalité infantile (r.c.= 1,294) tandis que l'augmentation de la consanguinité maternelle totale présente un rapport de cotes protecteur (r.c.= 0,837). Aucune de ces deux variables n'avait semblé avoir d'impact significatif sur les modèles homologues de mortalité infantile chez les garçons et pour la progéniture totale( #1 et 5 respectivement).

## 5.3.3.2 Modèles soumettant les variables génétiques proches.

Tableau 5.10 : Modèles de régressions logistiques effectués par sélection FORWARD touchant la progéniture féminine et soumettant les variables génétiques proches.

variables dans le modèle(!)	types de mortalité associés à leur numéro de modèle				
	infantile (8)	mortinatalité *(17)	néonatale précoce *(26)	1-4 ans *(35)	pré-reproductive *(44)
nbenf	NS	NS	<b>0,829</b> <b>0,016</b>	NS	NS
difconj	<b>0,919</b> <b>0,025</b>	NS	NS	NS	NS
rangnai	NS	NS	NS	NS	NS
year	<b>1,021</b> <b>0,000</b>	<b>1,026</b> <b>0,023</b>	<b>1,034</b> <b>0,003</b>	NS	<b>1,008</b> <b>0,019</b>
agemn	NS	NS	NS	NS	NS
interidi	NS	NS	NS	NS	NS
xapp44	NS	NS	NS	NS	NS
xcngh55	NS	NS	NS	NS	NS
xcngf55	NS	NS	NS	NS	NS
constante	0,000	0,014	0,002	NS	0,011

(!): La liste des noms complets des variables associés à leur acronyme est fournie en appendice dans le tableau A.1.

\* Modèle identique à son homologue soumettant les variables génétiques totales.

Les cinq modèles de mortalité féminine contenant des valeurs d'apparentement proche et de consanguinité proche (#8, 17, 26, 35 et 44) présentent un tableau très comparable au précédent en ce qui concerne leurs variables communes. En effet, seul le premier d'entre eux (mortalité infantile, #8) diffère de son modèle correspondant (#7, confronté aux variables génétiques totales). Dans ce modèle où l'on a tenté de mesurer l'impact de l'apparentement proche et des consanguinités parentales respectives proches, on ne dégage finalement aucune

influence significative de ces composantes. Les seules variables dont l'impact est mesurable dans ce modèle de mortalité infantile sont donc l'année avec un rapport de cotes de 1,021 et la différence d'âge entre les conjoints. Cette dernière variable n'était pas présente dans le modèle précédent ni chez les garçons mais elle s'était cependant démarquée dans tous les modèles de mortalité infantile totale (garçons et filles). La différence d'âge entre les parents aurait donc un impact sur la survie des filles mais n'influencerait pas de façon mesurable celle des garçons.

## 5.3.3.3 Modèles soumettant les variables génétiques éloignées.

Tableau 5.11: Modèles de régressions logistiques effectués par sélection FORWARD touchant la progéniture féminine et soumettant les variables génétiques éloignées.

variables dans le modèle(!)	types de mortalité associés à leur numéro de modèle				
	infantile (9)	mortinatalité (18)	néonatale précoce *(27)	1-4 ans *(36)	pré-reproductive *(45)
nbenf	NS	NS	<b>0,829</b> <b>0,016</b>	NS	NS
difconj	NS	NS	NS	NS	NS
rangnai	NS	NS	NS	NS	NS
year	<b>1,021</b> <b>0,000</b>	<b>1,028</b> <b>0,018</b>	<b>1,034</b> <b>0,003</b>	NS	<b>1,008</b> <b>0,019</b>
agemn	NS	NS	NS	NS	NS
interidi	NS	NS	NS	NS	NS
xcél1	<b>1,316</b> <b>0,022</b>	NS	NS	NS	NS
xché1	NS	NS	NS	NS	NS
xcfé1	NS	<b>0,315</b> <b>0,027</b>	NS	NS	NS
constante	0,000	0,014	0,002	NS	0,011

(!): La liste des noms complets des variables associés à leur acronyme est fournie en appendice dans le tableau A.1.

\* Modèle identique à son homologue soumettant les variables génétiques totales.

Le dernier groupe de modèles présentés (#9, 18, 27, 36 et 45) ressemble encore beaucoup aux deux premiers modèles de mortalité féminine en ce qui concerne leurs variables communes. En fait, les modèles de mortalité néonatale précoce, de mortalité entre un et quatre ans (NS) et de mortalité pré-reproductive (#27, 36 et 45) sont tous trois identiques à leurs homologues respectifs dans les groupes de

modèles pour les variables génétiques totales (#25, 34 et 43) et proches (#26, 35 et 44). Le modèle de risque de mortinatalité (#18) est cependant différent de ses deux modèles homologues précédents (#16 et 17) puisque la consanguinité maternelle éloignée semble l'influencer de façon protectrice. On voit donc un rapport de cotes de 0,315 pour ce modèle. Ce rapport de cotes est extrêmement influent car l'augmentation d'une classe de consanguinité maternelle éloignée divise environ par trois le risque de mortinatalité chez les filles. L'année de naissance présente encore un rapport de cotes semblables aux autres modèles avec r.c.= 1,028 pour la mortinatalité.

Un risque similaire se mesure pour la mortalité infantile (#9) avec un rapport de cotes de 1,021 pour l'année de naissance. De plus, on voit que l'apparentement éloigné entre les parents augmente le risque de mortalité infantile chez les filles (r.c.= 1,316).

Les modèles ENTER de mortalité féminine ne présentent pas de variables génétiques supplémentaires. Toutes les variables d'apparentement ou de consanguinité sélectionnées par FORWARD sont significatives dans les modèles ENTER et vice versa. On peut donc dire que les modèles de mortalité féminine ont mieux répondu et de façon plus « stable » que les autres types étudiés et ce, du moins, pour les composantes génétiques des modèles de mortalité entre zéro et un an.

#### **5.4 Analyse de l'ajustement et du pourcentage explicatif des modèles**

En plus du test de Wald qui fournit un indice sur la probabilité que la constante de chacune des variables indépendantes soit égale à zéro, il existe aussi un test global qui permet d'estimer en groupes les paramètres de chaque modèle et donc de voir si un modèle en particulier est valable. Chaque régression logistique fournit en effet les valeurs de  $-2 \text{ Log likelihood}$  et un test de chi carré (« Model ») qui

estime la différence entre le  $-2$  Log likelihood du modèle avec et sans ses variables (constante exclue) ainsi qu'un seuil de signification conséquent de chaque modèle. La mesure du  $-2$  Log likelihood se doit d'être la plus petite possible, une valeur de zéro révélant un ajustement parfait entre les variables et le modèle estimé. Il est cependant rare d'observer un ajustement dont la résultante donnerait une valeur inférieure à 50. À ces données s'ajoutent aussi les indices «  $R^2$  » de Cox et Snell ainsi que de Nagelkerke. Ces indices estiment le pourcentage de la variation expliqué par le modèle. Ils donnent donc une bonne idée de l'influence totale des variables indépendantes sur la variable dépendante. L'indice le plus approprié pour estimer le pourcentage de la variation expliqué est celui de Nagelkerke puisque celui-ci constitue un ajustement du «  $R^2$  » de Cox et Snell. Il permet d'atteindre la valeur de 1 (100%), ce que le premier ne pouvait pas faire (SPSS Inc., 1997). Cependant, les deux indices sont présentés.

Si on regarde les modèles présentés, on peut maintenant voir que tous ceux qui portent les mêmes variables pour une même classe de mortalité et la même progéniture (totale, masculine ou féminine) ont également les mêmes résultats aux tests présentés ci-haut. Ainsi, les modèles 1, 2, et 3 sont identiques. Les valeurs de consanguinité et d'apparentement n'ont eu aucun impact sur la mortalité infantile totale. Il en est de même pour les modèles 4, 5 et 6 (mortalité infantile masculine), les modèles 10, 11 et 12 (mortalité totale), les modèles 25, 26 et 27 (mortalité néonatale féminine) et les modèles 43, 44, et 45 (mortalité pré-reproductive féminine). Les modèles 28 et 30 (mortalité entre 1 et 4 ans pour la progéniture totale avec les valeurs génétiques totales et éloignées respectivement) sont également identiques mais leur homologue dans les valeurs génétiques proches est différent puisqu'il a inclus la consanguinité maternelle proche dans le modèle. Il en va de même pour les modèles 37 et 39 (mortalité pré-reproductive totale avec les valeurs génétiques totales et éloignées respectivement) et les modèles 40 et 42 (mortalité pré-reproductive masculine avec les valeurs génétiques totales et éloignées). Les modèles correspondants qui testaient l'impact des variables

généétiques proches , 38 et 41, ont inclus la variable apparemment proche des parents. Finalement, les modèles 14 et 15 (mortalité masculine avec les variables génétiques proches et éloignées) ainsi que 20 et 21 (mortalité néonatale totale avec les variables génétiques proches et éloignées) sont également semblables. Leur homologue respectif pour les variables génétiques totales présente dans les deux cas la consanguinité maternelle totale comme significative mais les composantes proches et éloignées de cette consanguinité n'ont pas eu le même impact significatif.

En observant le chi-carré de chaque modèle, on peut aussi éliminer d'office certains d'entre eux dont le seuil de signification n'a pas rencontré le critère de 0,05. Les modèles 22, 23 et 24 (mortalité néonatale précoce masculine pour les variables génétiques totales, proches et éloignées), 34, 35 et 36 (mortalité féminine entre 1 et 4 ans pour les variables génétiques totales, proches et éloignées) n'ont donc pas réussi à mesurer un impact significatif parmi toutes les variables soumises. C'est donc une indication que d'autres facteurs non mesurés lors de cette étude influencent fortement ces types de mortalité. Ces facteurs ont probablement une influence sur les autres types de mortalité mais ces autres modèles ont toutefois pu se qualifier significativement grâce à leurs autres variables influentes.

Les indices de Cox et Snell et de Nagelkerke fournissent une indication supplémentaire de la présence d'autres variables influentes sur les diverses catégories de mortalité puisque ces indices servent à exprimer le pourcentage de la variation expliqué par chaque modèle. En se fiant à l'indice de Nagelkerke (plus précis), on voit que le pourcentage de la variation expliqué oscille entre 1 et 36% pour les modèles effectués dans la présente étude. Cette valeur de 36 % est cependant élevée puisque la majorité des modèles montrent des indices qui vont de 5 à 16%.

Le tableaux G.1, G.2 et G.3 présentés en appendice donnent donc les valeurs de  $-2\text{Log likelihood}$ , de chi-carré (Model) avec leur seuil de signification ainsi que les indices de Cox et Snell et de Nagelkerke pour chacun des modèles étudiés.

### **5.5 Analyse de la calibration et des capacités de discrimination des modèles**

Maintenant que chacun des modèles présentés a été observé, il est important de pouvoir critiquer le pouvoir prédictif de chacun de ceux-ci. Pour ce faire, il existe quelques mesures permettant d'évaluer les capacités de discrimination et la calibration d'un modèle. La capacité de discrimination du modèle correspond à sa capacité de distinguer à quel groupe (mort ou non pendant la période d'observation) appartient un individu de l'échantillon en tenant compte des probabilités prédites par le modèle. Afin de donner une mesure du pouvoir de discrimination des modèles, nous possédons les tables de classification de chaque modèle ainsi que les courbes « ROC » qui permettent de tester la sensibilité et la spécificité des modèles. La calibration du modèle consiste plutôt à voir si les probabilités observées et prédites par le modèle sont concordantes. Un test de Hosmer et Lemeshow a été effectué pour vérifier cette calibration.

Les résultats des tests de calibration (Hosmer et Lemeshow) et de discrimination (ROC curves et les tables de classification) sont présentés dans les tableaux et figures des appendices H et I respectivement.

#### **5.5.1 Tests de Hosmer et Lemeshow : la calibration**

En s'attardant aux résultats des tests de Hosmer et Lemeshow, on peut voir que les modèles présentent un problème de calibration. En effet, sauf exception, les résultats des tests de Hosmer et Lemeshow révèlent que les modèles de régression logistique sont non-significatifs. Ce test consiste à diviser en dix classes à peu près égales l'effectif selon une échelle de la probabilité d'occurrence de

l'événement (ici de mortalité) en commençant par le plus faible risque jusqu'au risque le plus élevé. Le test reprend ensuite l'effectif pour le répartir dans ces différentes classes mais cette fois, selon les résultats observés (et non prédits). Un test de chi-carré est ensuite effectué afin de déterminer s'il existe une différence significative entre la répartition des catégories prédites et observées.

Puisque la grande majorité des modèles étudiés présentent un chi-carré non-significatif pour ce test, il est évident que l'augmentation du risque de mortalité n'évolue pas de façon constante. Ce test de calibration élaboré par Hosmer et Lemeshow s'avère être l'un des plus sévère et le fait que l'effectif soumis au test ne soit pas de très grande taille amplifie la sévérité du test. En effet, pour chaque modèle testé, le nombre d'individus dans chaque classe prévue du test n'est pas très élevé. Hosmer et Lemeshow suggèrent d'autres tests pouvant être effectués (Hosmer et Lemeshow, 1989). Ainsi, il existe entre autres un test nommé « Score test » élaboré par Brown qui est plus permissif (1982). Un des problèmes entraînés par des tests plus souples qui ont, par exemples, moins de 10 classes, est de perdre la sensibilité permettant de faire la différence entre les fréquences prédites et les fréquences observées (Hosmer et Lemeshow, 1989). Pour cette raison, il a été préféré de conserver le test original de Hosmer et Lemeshow pour les analyses.

Il est également fort probable que les chi-carrés non-significatifs des modèles soient dus à un problème de calibration de certaines variables. En effet, si le passage d'une classe à l'autre ou d'une valeur à l'autre pour une variable donnée n'entraîne pas toujours la même variation du risque de mortalité (par le rapport de cotes), alors ce risque de mortalité devient très difficile à estimer. Or, la plupart des variables utilisées ont ce genre de problème. Par exemple, l'augmentation de l'âge de la mère à la naissance ne présente pas le même facteur de risque s'il passe de 18 à 19 ans que s'il passe de 40 à 41 (et ce surtout pour un premier enfant). Il en est de même pour les variables génétiques qui ont été classifiées selon l'augmentation des coefficients (tableau 4.1). Ces variables sont des facteurs

biologiques et dynamiques. Il est donc important d'être conscient de ces différentes modulations et de concevoir en quoi elles peuvent influencer la calibration et la capacité de discrimination du modèle. En fait, ces variables peuvent déstabiliser les modèles et rendre plus difficile les interprétations.

#### 5.5.2 Tables de classification et « ROC curves »: la capacité de discrimination

En ce qui concerne la capacité de discrimination du modèle, les tableaux I.1 à I.3 présentés en appendice présentent les tables de classification pour les modèles visant respectivement la progéniture totale, masculine et féminine. De plus, les courbes « ROC » concernant ces différents modèles sont également présentés dans les figures I.1 à I.23. La présence de ces courbes dans le présent travail n'est pas nécessaire à sa complétude et leur interprétation ici détournerait donc l'intérêt principal de la recherche. Pour ces raisons, et pour conserver la fluidité de l'ouvrage, les explications concernant l'interprétation de ces courbes sont également présentées dans l'annexe I, juste avant les courbes.

Les tables de classification présentent un fait troublant qui s'était déjà annoncé lors des tests de calibration : la capacité discriminatoire représentée est très faible pour tous les modèles. En fait, le problème dénoté est le suivant : le modèle n'est pas capable de prédire efficacement les vrais cas (ceux qui meurent pour une catégorie de mortalité donnée). C'est donc la validité intrinsèque des modèles qui est ici mise en doute. Par conséquent, le pouvoir prédictif est également remis en question. Le pourcentage total de correspondance entre les valeurs observées et prédites est habituellement élevé mais si on regarde plus attentivement, on voit que ce résultat est dû au succès de classement des « 0 » (non-morts pendant la période étudiée). En ce qui concerne les événements « 1 », le classement ne correspond habituellement pas du tout. Par exemple, les modèles # 1, 2 et 3 sur la mortalité infantile de la progéniture totale (identiques), montrent un pourcentage total de 76,7% de correspondance. Cependant, si l'on regarde les résultats prédits, on voit

que le modèle prédit la mortalité infantile de  $756+229$  individus (total de 985 individus) et la survie de  $6+18$  individus (total de 24). Les valeurs observées révèlent une toute autre réalité :  $756+6=762$  individus ne meurent pas entre 0 et 1 an mais  $229+18=247$  individus meurent dans cette période. Ce chiffre est beaucoup plus élevé que la valeur prédite. On voit donc que le pourcentage correctement classé pour les « 0 » observés est de 99,2% mais seulement de 7,3% pour les « 1 » observés. On peut conclure qu'il existe une lacune dans la sensibilité des modèles présentés; ce problème devra être pris en considération quand viendra le temps d'analyser les résultats obtenus.

## CHAPITRE VI

### DISCUSSION

De prime abord, les résultats de l'ajustement des différents modèles effectués nous mettent en garde sur les limites de la modélisation. Cependant, les résultats obtenus à l'intérieur de chacun de ces modèles sont non seulement fort intéressants mais en plus, ils sont, pour la plupart, en harmonie avec les hypothèses de travail élaborées dans la présente étude et la plupart des recherches menées sur le sujet. Les coefficients d'apparentement et de consanguinité maternelle (totaux, proches et éloignés) ont été fréquemment considérés comme influents dans le destin de la progéniture tandis que la consanguinité paternelle est restée sans effet dans tous les cas. De plus, plusieurs variables démographiques ont démontré leurs impacts respectifs dans ce même destin de mort ou survie de la progéniture. Fait aussi très significatif, une certaine constance entre les modèles a été remarquée; donc, malgré tous les problèmes d'ajustement des modèles, on peut quand même considérer que les variables significatives ont un impact réel.

Commençons tout d'abord par rappeler les objectifs de cette étude. En premier lieu, l'hypothèse principale de cet ouvrage allègue que la consanguinité féminine éloignée dans cette population à tradition endogame aurait un impact protecteur sur la survie de la progéniture in utero ce qui entraînerait une plus grande chance de mener à terme une grossesse. La progéniture posséderait également de plus grandes chances de survie jusqu'à l'âge de reproduction puisqu'elle hériterait d'un génome épuré d'une partie de ses allèles délétères au fil des générations. En ce qui concerne le rapport de masculinité anormal à l'Île-aux-Coudres, il est impossible de mesurer directement les interactions entre le complexe majeur d'histocompatibilité et le chromosome Y. En effet, nous ne possédons aucune données d'analyses sériques des HLA. Nous avons quand même analysé les modèles de mortalité féminine et les modèles de mortalité masculine de manière à

mettre en valeur une possible interaction différentielle entre la mère et ses foetus, selon leur sexe.

Ce n'est pas la première fois que l'on tente de cerner l'impact de l'apparentement et de la consanguinité sur la mortalité dans une population (Bittles, 1994; Boisvert, 1992, 1994; Grant, 1997; Khlal et Khoury, 1991; Macklin, 1951; Piro, 1992; Shull et al. 1970a, 1970b; Shull et Neel, 1972). Toutefois, afin d'avoir une vision la plus biologique possible de l'impact de l'histoire génétique d'une population sur la mortalité qu'on y observe, il est absolument essentiel de prendre en compte le maximum des variables socio-démographiques qui la caractérisent. De par sa nature, la régression logistique permet cette prise en compte.

Suite à des recherches déjà effectuées sur l'Île-aux-Coudres, l'objectif ultime de la présente étude est donc de tenter de répondre aux hypothèses de départ en cernant par modélisation l'impact de la consanguinité maternelle sur la mortalité de la progéniture à l'Île-aux-Coudres. Dans le cadre de cette modélisation, ici par régression logistique, il est essentiel de considérer en plus des composantes génétiques (consanguinité respective du père et de la mère ainsi que leur apparentement), celles de nature socio-démographique pour assurer une bonne interprétation des résultats obtenus. Dans cette optique, commençons par discuter de l'impact de chacune des variables démographiques préalablement sélectionnées sur l'effectif à l'étude. De plus, dans la section 6.3, nous apporterons des précisions tirées des analyses faites sur les individus exclus de l'étude en comparaison avec l'effectif conservé (section 4.2). De plus, ces explications contribueront à éclaircir l'influence de certaines variables significatives dans les modèles présentés.

## 6.1 Les variables démographiques

Rappelons brièvement que toutes les variables démographiques soumises aux modèles sont les mêmes (dans tous les modèles): le nombre d'enfants du couple, la différence d'âge entre les conjoints (en années), le rang de naissance de l'enfant, l'année de naissance de l'enfant, l'âge de la mère à la naissance de l'enfant (en années) et l'intervalle génésique (en jours).

### 6.1.1 La taille de la famille

Le nombre d'enfants de la famille s'est avéré un facteur protecteur contre la mortinatalité, la mortalité néonatale précoce et la mortalité infantile pour les modèles concernant l'ensemble de la progéniture (garçons et filles considérés ensemble). Dans les modèles féminins, cette protection agit contre la mortalité néonatale précoce seulement tandis que du côté des garçons, son effet se manifesterait contre la mortinatalité. Rappelons que le nombre d'enfants par famille varie entre 1 et 19 et, qu'en moyenne, les familles de l'étude se composent d'environ une dizaine d'enfants. La protection entraînée par une famille nombreuse a, comme il a déjà été mentionné dans la section d'analyse des résultats, une connotation rétrospective car, à moins d'être le cadet de la famille, l'enfant qui naît n'est pas entouré de l'ensemble de tous ses frères et soeurs. On peut donc penser que les rapports de cotes protecteurs obtenus reflètent la santé et les ressources de la famille (les deux parents doivent demeurer en vie pour permettre la naissance de tous ces enfants) , et surtout la santé de la mère puisque, plus celle-ci est en forme, plus elle sera en mesure de s'occuper d'un grand nombre d'enfants. Il est important de remarquer que la variable nombre d'enfants de la famille influe sur le sort de la progéniture seulement pour la première année de vie. C'est pendant cette année que l'enfant requiert le plus de soins maternels.

Si on revient aux données de l'occupation des hopitaux dans la région de Charlevoix ouest (Rochette, Drouin et Tremblay, 1986), on se rappellera que les filles montraient une surhospitalisation pour des maladies périnatales (0-27 jours). Si l'augmentation du nombre d'enfants au sein d'une famille est un indicateur d'une meilleure santé de la mère, on peut aussi supposer que ses enfants sont en meilleure santé et que l'hospitalisation et la mortalité périnatale des filles est moins prononcée. Pour les garçons, l'augmentation de la taille de la famille préviendrait la mortinatalité. Ces deux phénomènes pris ensemble introduisent une nouvelle question : le nombre plus élevé d'enfants au sein de la famille serait-il dû à moins de pertes de foetus? Il existe une faible corrélation négative (-0,1189) entre l'intervalle génésique et le nombre d'enfants de la famille. On peut penser que plus cet intervalle génésique augmente, moins la taille de la famille est grande, donc on peut soupçonner une perte d'enfants in utero. Cependant, cette corrélation est très faible. Peut-être est-ce parce que l'intervalle génésique est exprimé en jours, ce qui amoindrit l'impact car les variations sont moins marquées?...

On remarque aussi que la variable « nombre d'enfants de la famille » se démarque plus avec la progéniture totale que lorsque l'on prend en considération les garçons seuls ou les filles seules. En effet, la protection contre la mortalité infantile ne s'est manifestée que lorsqu'on s'attardait à la progéniture totale. Dans les modèles axés sur la mortalité infantile séparés selon le sexe, la variable « taille de la famille » n'a pas atteint le seuil de signification.

#### 6.1.2 L'intervalle génésique

Si l'on s'attarde maintenant à la variable « intervalle génésique », on peut constater que l'augmentation de cet intervalle a un petit impact positif sur les chances de survie de la progéniture totale pendant la période où elle est exposée à la mortalité infantile (0 à 1 an). Cet impact est très mineur, presque neutre (r.c.=0,999). Il y a lieu de croire que cet effet protecteur pourrait être lié au poids du

bébé à la naissance. Plus il se sera écoulé de temps entre la naissance d'un premier enfant et de son cadet, plus la mère aura récupéré. Dans cette optique, sa grossesse aura plus de chances de s'être bien déroulée et le bébé qui naît aura donc plus de chances d'afficher un poids normal, ce qui augmente ses chances de survie. Le résultat obtenu va toutefois à l'encontre de ce qui a été avancé par Brittain (1992). Rappelons que dans son étude sur la population de l'île Saint-Barthélémy, la naissance d'un deuxième enfant dans la même année augmentait les chances de survie du premier enfant. L'étude de Brittain porte sur une population antillaise tandis que les résultats par rapport au lien intervalle génésique-poids du bébé à la naissance proviennent de sources québécoises. Probablement que ce dernier ouvrage, par sa nature québécoise, reflète mieux ce qui se passe à l'Île-aux-Coudres (Lepage et Levasseur, 1989).

Encore une fois, la variable « intervalle génésique » aurait probablement plus d'impact si elle n'était pas exprimée en jours. Peut-être que des classes seraient plus appropriées. Néanmoins, l'intervalle moyen entre les naissances est de deux ans et une augmentation substantielle de cet intervalle pourrait refléter une perte in utero (due à un problème d'histocompatibilité?) qui espacerait les naissances ou encore un allaitement prolongé. De plus, le fait que la mère n'ait à s'occuper que d'un seul enfant en très bas âge favorise la survie de cet enfant puisqu'il est le seul à demander beaucoup d'attention.

### 6.1.3 L'âge de la mère à la naissance et le rang de naissance de l'enfant

Attardons-nous maintenant à l'influence de l'âge de la mère à la naissance et du rang de naissance de l'enfant. L'interprétation de ces deux variables est très complémentaire et c'est pour cette raison que nous les analyserons ensemble. Toutes deux ont dévoilé que leur augmentation avait un impact néfaste sur les chances de survie de la progéniture. Cependant, elles n'ont pas eu cet effet sur les

mêmes modèles de régression logistique présentés : l'âge de la mère agit sur la mortalité infantile et le rang de naissance influence la mortalité pré-reproductive.

Sur le plan démographique, plus la mère est âgée, plus l'enfant a un rang de naissance élevé et, par conséquent, plus la charge de la mère est grande. Il est donc normal que cet enfant reçoive moins d'attention puisque la mère doit s'occuper de plusieurs enfants.

De plus, biologiquement parlant, plus la mère vieillit, plus les risques de formation de gamètes anormaux augmentent, plus les grossesses sont à risque. On a donc plus de chances de voir des irrégularités chromosomiques et génétiques chez la progéniture quand la mère est vieillissante (Marieb, 1999). L'âge du père étant corrélé à celui de la mère, il en est de même pour les risques de malformation des gamètes paternels. Plus la mère (et le père) vieillit (vieillissent), plus les enfants sont susceptibles d'avoir une santé fragile et d'exiger davantage de soins. Selon le modèle, les plus défavorisés mourront dans la première année (mortalité infantile).

Si l'enfant souffre d'une maladie génétique qui le rend dépendant d'un surplus de soins, ses chances de survie sont compromises. Étrangement, le rang de naissance n'a pas manifesté d'impact significatif sur la mortalité infantile comme on pourrait s'y attendre. Dans la section d'analyses descriptives, on fait état de la forte corrélation qui existe entre l'âge de la mère à la naissance et le rang de naissance (0,8037). On a choisi, malgré cette corrélation, de garder les deux variables pour les analyses dans le but de voir si les informations qu'elles véhiculent pouvaient se démarquer l'une de l'autre. Lorsque nous examinons la mortalité infantile chez la progéniture totale, il semblerait que ce soit l'âge de la mère qui ait le plus d'influence. Il est probable que l'essentiel de l'information ayant pu entrer de cette manière dans le modèle, le rang de naissance ne pouvait rien apporter de plus et n'a pu être inclus. Toutefois, pour la mortalité pré-reproductive de la progéniture totale, c'est le rang de naissance qui devient très influent. Plus le rang de

naissance augmente, plus la progéniture est à risque pour ce type de mortalité. On observe également le même phénomène pour la progéniture masculine, et ce, de façon plus prononcée (r.c. de 1,079 contre r.c. de 1,008). Contrairement aux garçons, le rang de naissance chez les filles ne semble pas avoir une importance quant au risque de mortalité pré-reproductive. On pourrait interpréter ces résultats comme suit : la famille ne défavorise pas ses cadets, au contraire, elle favorise ses aînés. Cette affirmation s'applique en fait à la progéniture mâle. En ce qui concerne la progéniture totale, c'est sa composante masculine seulement qui fait ressortir le rang de naissance comme significatif (les modèles féminins ne l'incluent pas). À l'Île-aux-Coudres, c'est en général le fils aîné qui hérite de la terre familiale (Le Querrec, 1981). Un aîné en santé qui atteint l'âge de travailler aux champs et/ou à la pêche devient un atout économique pour la famille. En plus de contribuer fortement à cette économie familiale, il assure une relève à son père, une continuité des traditions familiales et du patronyme. C'est donc probablement en lui que la famille met le plus d'espoirs et d'énergie. Quand la survie des fils aînés est assurée, les fils subséquents ne sont pas nécessairement négligés mais les aînés, dès leur naissance et par la suite, représentent l'assurance de la descendance et d'une transmission familiale des acquis. Pour les filles, l'importance du rang de naissance ne s'applique pas. Aînées ou cadettes, elles quitteront la famille pour se marier; les aînées ne bénéficieront pas de ce surplus d'attention entre 1 et 14 ans. On peut donc qualifier de « culturel » l'impact entraîné par le rang de naissance des garçons.

#### 6.1.4 La différence d'âge entre les conjoints

La différence d'âge entre les conjoints demeure plus une variable à connotation économique. C'est en fait le seul indice économique dont nous disposons pour une population où la situation sociale et économique est par ailleurs assez homogène (Martin 1957 et Boisvert, 1992). On a remarqué que cette variable était protectrice contre la mortalité infantile de la progéniture totale ainsi que contre la

mortalité infantile des filles. Seulement, cette variable est retenue dans tous les modèles de la progéniture totale (i.e. avec les variables d'apparentement et de consanguinité totales, proches et éloignées) tandis que pour les filles, seul le modèle qui met en jeu les variables génétiques proches inclut la différence d'âge entre les parents. Rappelons toutefois que dans ce dernier modèle, les variables d'apparentement et de consanguinité proches n'ont pas atteint le seuil de signification. Comme pour le SES (score socio-économique) présenté par Shull et ses co-auteurs (1970a, 1970b et 1972), la différence d'âge entre les conjoints donne un bon indice de la situation socio-économique du couple. Plus cet écart est marqué, (habituellement parce que l'homme est plus vieux que sa femme) plus le père de famille a de chances d'être « établi » et capable de subvenir aux besoins de sa famille, indépendamment de la force du réseau social qui entoure cette même famille. Pour cette raison, les enfants ont de meilleures chances de survie durant leur première année de vie. On pourrait penser, par exemple, que dans le cas où une maladie génétique (ou tout autre problème) serait déclarée chez l'un des enfants, la famille la mieux établie aurait plus de chances de disposer des ressources pour s'assurer du traitement et de la survie de l'enfant. Ce phénomène s'observe indépendamment des facteurs génétiques, ce qui montre bien que la différence d'âge entre les conjoints implique principalement des facteurs économiques et démographiques, pas des facteurs biologiques.

Pour les filles, la variable « différence d'âge entre les conjoints » ne s'est manifestée que lorsqu'on étudiait le modèle incluant les valeurs génétiques proches : les consanguinités (paternelle et maternelle) proches ainsi que l'apparentement proche n'ont pas été considérées comme significatives mais leur présence plutôt que celle de leurs homologues totales ou éloignées semble entraîner une différence entre les modèles. Est-ce parce que les variables génétiques totales et/ou éloignées ont pu se démarquer dans leurs modèles respectifs, masquant ainsi l'impact de la différence d'âge entre les conjoints? C'est fort possible puisque les interactions épidémiologiques entre les variables n'ont pas

été vérifiées. Il est également possible que seuls l'apparement et la consanguinité proche soient néfastes pour les fillettes et donc que l'augmentation de la différence d'âge entre les conjoints vienne palier ce risque supplémentaire en protégeant les filles (dû à une situation sociale plus avantageuse). Du même coup, la variable différence d'âge entre les conjoints empêcherait l'entrée significative des variables génétiques proches.

#### 6.1.5 L'année de naissance de l'enfant

Finalement, l'année de naissance est la seule variable qui se soit démarquée dans tous les modèles significatifs. Sa présence dénote toujours une augmentation du risque de mortalité avec l'augmentation des années quoique les rapports de cote demeurent toujours très faiblement différents de zéro. Il est important de ne pas oublier que la variable « année de naissance » n'est pas organisée par classe dans les régressions. L'augmentation d'une unité de cette variable correspond donc à l'augmentation d'une année de calendrier. Si on avait regroupé les années de naissance par décennie, peut-être aurait-on vu un rapport de cote plus prononcé. Néanmoins, deux explications se présentent pour l'interprétation de l'impact de l'année de naissance sur le risque de mortalité de la progéniture.

Tout d'abord, on peut supposer que l'augmentation du risque de mortalité dans toutes les catégories observées est probablement attribuable à l'augmentation de la population de l'île au fil des ans jusqu'à atteindre un niveau de surpopulation qui pousse à l'émigration une partie de la population à cause du manque de ressources (Le Querrec, 1981).

De plus, il n'y avait aucun médecin dans l'île avant 1930. Notre étude s'étend jusqu'aux années 50 pour les natalités. Cependant, la majeure partie des naissances ont lieu avant 1930 et pour cette raison, l'augmentation du risque de mortalité peut être le reflet de la surpopulation. Dans un article publié en 1992,

Brittain a mis en évidence que l'arrivée d'un médecin à Saint-Barthélémy en 1934 a significativement fait baisser la mortalité infantile. Rappelons que Martin (1957) a lui aussi relié la baisse de mortalité infantile à Île-aux-Coudres à l'arrivée d'un médecin résident.

En deuxième lieu, cette augmentation de la population entraîne un obstacle supplémentaire qui se manifeste surtout pour les dernières années de l'étude : la possibilité d'émigration de certains individus ou même de familles complètes. Comme il a déjà été mentionné, la nature de l'étude a mené à l'exclusion de tous les individus sans date de décès. Après vérification des dates moyennes de mariage des parents des individus formant le groupe inclus dans l'étude, c'est-à-dire l'effectif, en comparaison avec deux autres groupes d'individus ayant été exclus (le groupe des « exclus » et celui des « couples »), nous avons pu constater que ces moyennes étaient différentes entre elles de façon très hautement significative (fig. J.15). De plus, l'observation de ces moyennes révélait que celles-ci s'espaçaient entre elles d'une vingtaine d'années chacune (dates moyennes de mariages des parents : « inclus »=1869,292 (fig. J.2 b), « exclus »=1889,963 (fig. J.1 b) et « couples »=1907,051 (fig. J.3 b). Le groupe des « inclus » soumis à nos analyses est donc le plus ancien par rapport à la date de mariage des parents des individus à l'étude. Vient ensuite celui du groupe des « exclus » et finalement celui des « couples » qui est en fait un sous-groupe particulier des « exclus » (voir section 4.2). Les individus dont les parents ont une date moyenne de mariage située en l'année 1889 peuvent avoir émigré mais aussi avoir été encore en vie lors du dernier recensement de 1967. À plus forte raison, les individus faisant partie du groupe « couple » constitue la classe à part où tous les enfants du couple sont exclus des analyses. La date moyenne de mariage des parents est si « récente » (1907), qu'on peut supposer que la famille au complet a émigré à cause de la surpopulation mais on peut aussi croire que tous ces gens étaient encore en vie (et pas si âgés que ça) au recensement de 1967. En d'autres mots, comme il faut absolument être « officiellement mort » pour être inclu dans l'échantillon, il est

possible que certains de ces individus, nés vers la fin de l'étude, puissent avoir survécu passé l'âge de 14 ans et avoir émigré par la suite. On perd alors leur trace et on n'a aucune date de décès. Ces individus auraient été des « 0 » (des survivants) pour les mortalités étudiées. D'autres individus n'ont peut-être pas quitté l'île mais étaient encore en vie au moment de la constitution du Registre. Eux non plus n'ont pas de date de décès et ces « survivants » ne sont donc pas considérés. La conséquence demeure toutefois la même : plus l'étude avance dans le temps, plus on risque d'avoir une sous-représentation de survivants et une sur-représentation des morts pour les mortalités étudiées car on élimine les vivants de l'étude. En soustrayant ces gens qui n'ont pas encore « daigné mourir » on crée une augmentation artificielle de la proportion de décès vers la fin de l'étude. Ce phénomène explique donc partiellement le fait que plus les années sont récentes, plus le risque de mortalité augmente pour tous les types de mortalités étudiées.

## **6.2 Les variables génétiques**

Pour l'ensemble des modèles présentés, seulement deux des trois grandes catégories de variables génétiques ont démontré avoir un impact sur la mortalité de la progéniture : l'apparentement entre les parents et la consanguinité de la mère. La consanguinité du père n'a démontré d'impact significatif dans aucun des modèles. Selon les études les plus récentes (Ober, 1995, 1997 et 1998), la consanguinité du père n'a d'impact que dans la mesure où il est apparenté à sa conjointe et que les foetus produits par une telle union ont plus de chances d'être compatibles avec la mère. Il est donc normal que la variable « consanguinité paternelle » ait un effet beaucoup moins marqué (non mesurable, peut donc être nul) que la variable « apparentement » du couple ou la variable « consanguinité de la mère ».

Rappelons que chaque variable génétique a été divisée en trois composantes : totale, proche et éloignée. Nous tiendrons donc compte de chacune de ces composantes (qui sont dans les faits des variables distinctes les unes des autres) dans nos interprétations.

### 6.2.1 L'apparentement des conjoints

Commençons par expliquer l'impact sur la progéniture du coefficient d'apparentement entre les parents. Tout d'abord, chez les filles seulement, on a vu que l'apparentement total et l'apparentement éloigné des conjoints augmentaient les chances de mortalité infantile. En effet, lorsque les conjoints sont apparentés, il y a plus de chances de retrouver un allèle délétère récessif à l'état homozygote et il y a par conséquent une plus grande probabilité de voir une maladie génétique se déclarer durant la première année d'existence de la fillette.

Pourquoi n'observe-t-on pas ce même impact délétère de l'apparentement proche sur la mortalité féminine? Le fait que les parents soient apparentés par des liens proches augmente encore une fois la possibilité que la fillette soit homozygote. Cependant, si on observe ce phénomène sous l'optique de la biologie évolutive, des liens éloignés donnent la chance à des allèles trop délétères d'avoir été purgés du pool génique (par la mort in utero des individus qui les auraient portés); des liens proches ne donnent pas cette possibilité. Il est donc possible que l'apparentement proche n'ait pas eu d'impact significatif sur les risques de mortalité infantile féminine tout simplement parce la sélection s'opère in utero. Les foetus féminins trop tarés seraient éliminés sans qu'on puisse mesurer ce type de sélection et donc l'impact de la variable.

Chez les garçons, on n'observe pas d'impact significatif de l'apparentement des parents (que ce soit l'apparentement total, proche ou éloigné) sur les risques de mortalité dans la première année (que ce soit la mortinatalité, la mortalité néonatale

précoce ou la mortalité infantile). Vu le rapport de masculinité inversé à l'Île-aux-Coudres, serait-ce possible que les garçons affectés ne viennent jamais au monde? Ceci expliquerait qu'il y naisse moins de garçons, ceux-ci étant évincés avant même que la grossesse soit détectable (Gill, 1997). Cette stratégie évolutive sert à économiser les énergies de la mère en prévision d'une grossesse viable. Rappelons que ce type de relation mère-foetus a déjà été observé chez les souris. Les responsables de l'étude ont vu que lorsqu'elles étaient accouplées à des mâles apparentés, les femelles homozygotes avaient tendance à rejeter leurs foetus de sexe masculin. Cette mortalité in utero a été attribuée à un gène lié au chromosome Y influençant la production de HLA. On a également soulevé la possibilité d'un gène lié au MHC chez les Hutterites (Ober et al, 1998). Se pourrait-il qu'on puisse observer un phénomène similaire à l'Île-aux-Coudres? Y a-t-il un lien entre ce rapport de masculinité inversé et les interactions du MHC avec le chromosome Y lors de la grossesse des femmes de l'Île-aux-Coudres? Il est certain que la consanguinité favorise l'homozygotie, donc les femmes de l'île (plus fortement homozygotes) présentent peut-être une proportion anormalement élevée de rejet des foetus mâles à cause de leur chromosome Y qui interagirait avec le MHC de façon délétère; et ce, de façon amplifiée avec un conjoint apparenté. Les résultats suggèrent ce type de relation mais seules des études prospectives plus poussées sur les relations mère-foetus à l'Île-aux-Coudres pourraient en dire plus long.

Chez les garçons, les variables d'apparentement des parents (total, proche et éloigné) a un effet protecteur contre la mortalité entre 1 et 4 ans. Comment une variable peut-elle avoir un impact délétère pour une partie de la progéniture (les filles) et un impact protecteur pour une autre partie de cette même progéniture (les garçons) à un autre moment d'exposition?

L'apparentement, au niveau génétique est associé à une plus grande possibilité de retrouver des homozygotes désavantagés. Toutefois, sur le plan social,

l'apparentement proche peut présenter de très nombreux avantages pour une petite population. Comme il a été mentionné précédemment, les garçons représentent la garantie de continuité de la tradition familiale. Si un fils est en santé après la période de mortalité infantile passée, la famille voit en lui toutes les ressources familiales renouvelées (surtout s'il est l'aîné, c.f. l'analyse de la variable rang de naissance) et fera tout ce qu'elle peut pour l'amener en santé jusqu'à l'âge adulte. Si le réseau familial est agrandi et solidifié par la présence de liens familiaux reconnus entre les parents (par exemple, cousin de degré inégal) le couple et l'enfant bénéficieraient en plus d'un réseau d'entraide familial très serré et uni pour favoriser la survie du fils. Socialement parlant, plus les liens dans une famille se recourent, plus tous les membres se sentent concernés. Comme on peut le voir en appendice (tableau F.4) l'apparentement total a le même effet sur la survie des garçons entre 1 et 14 ans. La tendance est donc similaire pour les jeunes garçons entre 1 et 14 ans (les explications sont donc également les mêmes) mais cet impact n'a pas atteint le seuil de signification dans les modèles FORWARD.

Quand on arrive à étudier uniquement la composante génétique proche de l'apparentement, on voit que son impact est le même que pour son homologue total. Les seuils de signification ont de plus été atteints pour la mortalité masculine entre 1 et 4 ans ainsi que pour la mortalité masculine pré-reproductive (entre 1 et 14 ans). En plus, l'apparentement proche a montré un impact protecteur pour l'ensemble de la progéniture entre 1 et 14 ans et 1 et 4 ans ( non-significatif entre 1 et 4 ans).

Au contraire, ce même apparentement considéré pour ses paramètres génétiques éloignés seulement n'a eu aucun impact (positif ou négatif) sur la mortalité après la première année chez les garçons, les filles ou l'ensemble de la progéniture.

Ces résultats confirment bien ce qui vient d'être expliqué. En effet, le réseau social apparenté favorise la survie de la progéniture totale. Il est certain que si les ressources sont disponibles, on donne plus de chances de survie aux filles comme aux garçons. Toutefois, l'impact d'un réseau social étendu en fonction de l'apparentement des parents ne se détecte que lorsque cet apparentement est proche. Si les liens d'apparentement entre les époux sont trop éloignés, les individus ne se reconnaissent plus comme appartenant à la même famille (et, par exemple, il n'y a aucune dispense religieuse à demander lors du mariage). Par conséquent, le surplus de soutien dû aux forts liens unissant les familles n'existe plus. Quand l'apparentement total a un impact, c'est sa composante proche qui se manifeste. Ceci porte à croire que l'impact de l'apparentement sur la survie de la progéniture (après la première année de vie) serait de nature socio-culturelle et non génétique.

#### 6.2.2 La consanguinité maternelle

Voyons maintenant le rôle de la consanguinité maternelle pour la survie de sa progéniture. La consanguinité maternelle totale a un impact très protecteur sur l'ensemble de sa progéniture pendant la première année. Cet impact se manifeste contre la mortalité masculine, la mortalité infantile féminine et la mortalité néonatale précoce totale.

Si une femme est consanguine, elle porte donc potentiellement à l'état homozygote des gènes issus d'un ancêtre commun à ses deux parents. Cette femme a cependant survécu jusqu'à l'âge de la reproduction; ceci porte à croire que son bagage génétique ne lui a pas nuí. Les gènes qu'elle transmettra à sa progéniture sont donc profitables. Avec seulement cette information, on ne peut pas spéculer sur des interactions au niveau du MHC puisque les enfants décédés sont mort-nés ou très tard dans le développement foetal. On ne peut pas non plus savoir s'il s'agit de l'impact de la consanguinité maternelle proche ou éloignée.

Cependant, nous disposons d'un autre indicateur : la consanguinité maternelle éloignée qui se manifeste comme très protectrice contre la mortalité féminine. Est-ce cette même composante de la consanguinité maternelle qui se manifeste pour protéger la progéniture masculine du même type de décès? On ne peut le dire avec certitude. Cependant, en ce qui concerne les filles, la présence de consanguinité maternelle éloignée semble être le facteur le plus protecteur contre la mortalité (son rapport de cote est de 0,315; ce qui signifie qu'une augmentation d'une unité divise par trois la probabilité de décès).

La consanguinité et l'apparentement sont deux manifestations différentes d'une même réalité : l'endogamie au sein d'une population. L'apparentement se rapporte au lien entre deux individus tandis que la consanguinité révèle si un individu a des parents qui sont apparentés par des ancêtres. Dans une population endogame, il existe une forte relation entre l'apparentement des conjoints et la consanguinité de l'un ou l'autre d'entre eux. De plus, une tradition de mariages entre apparentés contribue à augmenter, en partie, l'homogénéisation de la population au cours des générations. Il n'est donc pas surprenant qu'un individu consanguin soit marié à un conjoint qui lui est apparenté. Si tel est le cas, tous deux ont donc plus de chances de produire une progéniture homozygote. Toutefois, si la femme, elle-même consanguine, a survécu, elle porte des allèles non-délétères ou avantageux. Donc, associée à un conjoint apparenté, il y aura peut-être sous-réactivité immunitaire lors des grossesses (foetus compatibles), ce qui pourrait la pousser à rejeter un surplus de foetus masculins. Ceux-ci ont moins de chances d'être différents de la mère puisqu'ils n'ont qu'un chromosome X. De plus, cette sous-réactivité immunitaire pourrait également être amplifiée par l'association du chromosome Y avec le MHC comme il a été mentionné plus haut. Les foetus restants, majoritairement féminins, héritent alors d'un génome très épuré. C'est probablement cet avantage génétique qui protège les fillettes de la mortalité infantile en général ainsi que des atteintes périnatales auxquelles les filles sont très sensibles dans Charlevoix ouest.

Si, au contraire, cette femme, dont la consanguinité est éloignée, n'est pas mariée à un conjoint apparenté, elle transmettra alors à sa progéniture une part d'allèles favorables. Toutefois, au niveau immunitaire de la relation mère-foetus, il n'y aura probablement pas de sous-réaction puisque le foetus (masculin ou féminin) aura plus de chances de posséder un haplotype différent de celui de la mère pour le MHC. Il n'y aura donc pas de surplus de rejet masculin dans ce cas. Garçons comme filles se verront donc avantagés génétiquement. En présence d'un conjoint non apparenté, c'est donc probablement la composante éloignée de la consanguinité maternelle qui agit comme protectrice contre la probabilité de mortinatalité chez les garçons. Pour les filles, cette composante est toujours protectrice, en présence ou non d'un conjoint apparenté.

La consanguinité maternelle proche s'est avérée être délétère pour la progéniture totale et les garçons dans le cas de la mortalité entre 1 et 4 ans. Cette action délétère est toujours associée à la présence d'apparentement proche au conjoint. Puisqu'il s'agit de consanguinité proche, peut-être que la mère a un génome qui n'est que partiellement épuré de ses allèles délétères et que, associé à ces mêmes allèles délétères fournis par le conjoint, les enfants se retrouvent avec, à l'état homozygote, des allèles qui seront la cause d'une maladie génétique en bas âge (entre 1 et 4 ans).

### **6.3 Les inclus et les exclus**

Suite aux analyses de la section 4.2, nous avons pu constater que les individus composant les groupes « exclus » étaient issus de familles plus nombreuses que les individus à l'étude (groupe des « inclus »). En effet, le nombre moyen d'enfants par famille dans notre effectif est de 9,531 tandis qu'il est de 10,237 chez l'ensemble des exclus. On a également observé que chez les individus de l'effectif, la date moyenne de mariage des parents se situait en 1869 tandis que chez les

individus exclus de l'étude, la date moyenne de mariage des parents se situe vingt ans plus tard, en 1889. Ces différences pourraient révéler une réalité sociale énoncée par Le Querrec (1981). En effet, plus les familles sont nombreuses, plus les ressources familiales sont divisées. Les individus appartenant à des familles nombreuses sont plus enclins à quitter l'île et aller chercher des ressources ailleurs. De plus, comme Le Querrec le mentionnait, une certaine surpopulation s'est installée à l'Île-aux-Coudres. La date moyenne de mariage du groupe des exclus étant de deux décennies plus récente que celle des individus de l'effectif, il est possible que ces deux phénomènes aient joué de concert pour favoriser l'exil de ces gens issus de familles nombreuses pouvant difficilement accéder à d'autres terres en raison de la surpopulation qui crée une pénurie. Ces individus ont donc probablement quitté l'île et sont morts à l'extérieur sans qu'on puisse avoir accès à l'information. Vu la date moyenne de mariage assez récente des parents de ces individus, il est également possible que certains n'aient pas quitté l'île mais aient été encore en vie au recensement de 1967. Pour combiner les deux problèmes, il est également fort probable que certains des émigrés aient été encore en vie en 1967. Toutes ces possibilités arrivent à la même conséquence : nous sommes privés de l'information sur leur âge au décès et nous devons les exclure de l'effectif.

En ce qui concerne le groupe « couple », la première explication de l'exclusion de la totalité des enfants d'une famille est que cette famille (parents et enfants) ont émigré. Il est important de noter que la date moyenne de mariage des parents de ces familles est en 1907. Cette émigration pourrait donc être due encore une fois à la surpopulation. Les hommes à la tête de ces familles d'émigrants seraient donc potentiellement des « emplacitaires » (Le Querrec, 1981). Une seconde explication de cette totale exclusion s'offre aussi à nous : vu la date de mariage relativement récente des parents, il est possible que la famille soit demeurée à l'Île-aux-Coudres et qu'en 1967, aucun des enfants ne soit encore décédé.

Nous avons voulu vérifier si, parmi les individus rejetés de l'effectif, nous avons conservé les mêmes proportions d'hommes et de femmes que dans l'effectif utilisé pour les analyses. Non seulement avons-nous constaté que les trois groupes présentaient des proportions significativement différentes, mais ces observations nous ont permis de pousser plus loin l'investigation. Selon les apparences de l'analyse, aucun groupe n'est représentatif du rapport de masculinité inversé à l'Île-aux-Coudres. Le groupe où tous les enfants d'une même famille sont exclus, c'est à dire le groupe « couples » compte 0,52918 homme pour 0,47082 femme. Chez le groupe des exclus, la proportion d'hommes et de femmes est très équilibrée (0,49546 homme pour 0,50454 femme) et dans l'effectif ayant servi à l'analyse, on voit des proportions de 0,48365 homme pour 0,48563 femme et 0,03072 individu de sexe inconnu. Toutefois, lorsque le rapport de masculinité était fait en se fiant aux variables « nombre d'enfants masculins »/ « nombre d'enfants féminins », on obtenait un rapport de masculinité de 0,952 (voir la section 5.2.2), typique de l'Île-aux-Coudres (Boisvert, 1992). Ce rapport n'inclut pourtant pas les individus de sexe inconnu. Le destin de ces individus a été analysé à la figure C.7. On voit que les individus de sexe inconnu (X), sont en moyenne âgés de 0 jours à leur décès. Ce sont donc des cas de mortalité qui n'ont été compilés dans les analyses que lorsque l'on s'attardait à la progéniture totale (on n'en tenait pas compte dans les modèles considérant les sexes séparément). S'il s'avérait que ces cas de mortalité étaient principalement des garçons, on se retrouverait face à un rapport de masculinité plus près de ce qui est observé dans le reste de la province, c'est-à-dire, plus de naissances masculines mais aussi plus de morts masculines. Peut-être alors que la sélection in utero est moins sévère que l'on ne le croyait... Il est impossible de vérifier ces hypothèses étant donné la non-identification du sexe de ces enfants. Cependant, le cas du groupe « couples » où le nombre de garçons dépasse le nombre de filles tend à soutenir cette hypothèse.

En ce qui concerne les valeurs de l'apparement et des consanguinités de ces trois groupes, on voit qu'en général, les groupes « exclus » et « couples »

présentent des valeurs d'apparentement et de consanguinités significativement supérieures à celles du groupe des « inclus ». En effet, les valeurs d'apparentement total et de consanguinités totales obtenues par les deux groupes soustraits à l'effectif se rapprochent beaucoup plus des valeurs calculées par Philippe en 1969. Les seules analyses n'ayant révélé aucune différence entre les variables génétiques communes d'un groupe à l'autre sont celles concernant la comparaison des consanguinités maternelles totales, des consanguinités maternelles éloignées ainsi que des consanguinités paternelles éloignées. L'interprétation de ces résultats est difficile... Est-ce un effet du temps? Rappelons-nous que vingt années séparent en moyenne les trois groupes les uns des autres en ce qui concerne la date de mariage des parents de ces individus. Cette hypothèse est bien soutenue par le fait que les consanguinités maternelle et paternelle éloignées ne tendent pas à augmenter et donc sont comparables d'un groupe à l'autre (et d'une tranche de vingt ans à l'autre par la même occasion) tandis que les consanguinités maternelle et paternelle proches, elles, diffèrent significativement d'un groupe à l'autre (et augmentent dans le temps). Il ne faut pas oublier que consanguinité proche, au sens de notre étude, ne réfère pas seulement aux dispenses de mariages obtenus pour des liens très près entre les parents (comme des cousins germains) mais bien à des liens pouvant remonter jusqu'à 5-5. Le fait que la consanguinité maternelle totale ne change pas non plus pourrait également révéler l'apport très grand de l'endogamie accumulée depuis des générations à l'île sur le coefficient total de consanguinité. Un fossé de vingt ans représente aisément une génération qui contribue potentiellement à faire augmenter les liens génétiques proches. Les aînés de familles élargies ont pu bénéficier du réseau social fort et des terres familiales mais, au fil des ans, l'augmentation de la population dans l'île a forcé d'abord les cadets puis des familles complètes à émigrer.

## CONCLUSION

La régression logistique utilisée dans le cadre de la présente étude nous a permis d'estimer avec succès les liens entre plusieurs facteurs d'exposition ainsi que leur impact sur la population à l'étude. Afin d'être rigoureux, toutes les variables significatives de chacun des modèles présentés ont été analysées et leur impact sur tous les types de mortalités étudiés a été développé dans la discussion. Lors de la vérification de l'ajustement des modèles de régression logistique, nous avons constaté, entre autres grâce aux tables de classification, que la sensibilité des régressions logistiques effectuées était très faible. La présence de fortes corrélations entre plusieurs des variables sélectionnées pour l'étude et le fait que plusieurs d'entre elles auraient dû être organisées par classes sont partiellement responsables de ce manque de sensibilité des modèles. Nous pouvons donc dire que le fait d'avoir voulu conserver la dynamique naturelle des variables d'exposition sur la mortalité a l'avantage de nous donner un reflet de ce qui se passe vraiment dans cette population aussi bien du côté social que génétique; cependant, la régression logistique devient moins fiable malgré sa grande tolérance aux corrélations entre variables indépendantes.

Malgré ces problèmes d'ajustement, nous avons été en mesure de dégager l'effet de chacune des variables indépendantes et d'analyser les phénomènes naturels de mortalité impliquant à la fois des composantes socio-culturelles, démographiques et biologiques dans le contexte naturel des habitants de l'Île-aux-Coudres. Vu l'ajustement des modèles, les analyses qui ont été développées dans la discussion sont donc applicables seulement à l'Île-aux-Coudres. De plus, les restrictions faites sur la population afin d'obtenir un effectif dont on peut bien contrôler les variables peuvent également restreindre les analyses et les conclusions à ce même effectif qui constitue la portion la plus stable de la population. Ainsi, les individus sans date de décès ont été exclus des analyses. Cette soustraction a donné lieu à des interrogations quant aux différences entre les individus faisant partie de

l'effectif et ceux ayant été rejetés; des analyses ont donc été menées afin de bien distinguer ce qui caractérise chacun de ces groupes. Une fois cette étape passée et après avoir terminé les analyses sur l'effectif conservé, il est possible, en toute connaissance de cause, d'utiliser les résultats obtenus ainsi que leurs interprétations comme pistes indicatives à d'autres études portant sur des populations endogames.

Dans la population à l'étude, le nombre d'enfants dans une famille est un gage de « santé familiale » qui indique que les deux parents ont survécu pendant la période génésique et sont également capables de subvenir aux besoins d'un grand nombre d'enfants vivants. Cependant l'intervalle entre les naissances vient poser une certaine limite au nombre d'enfants puisque si celui-ci augmente (espaçant ainsi les naissances) les chances de survie de la progéniture en bas âge augmentent. On sait que les mères ont tendance à donner naissance à des bébés de petit poids lorsque leurs grossesses sont trop rapprochées (deux dans la même année). Lorsque les enfants sont trop petits à la naissance, leurs chances de survie sont compromises. Le nombre d'enfants de la famille et l'intervalle entre les naissances n'ont toutefois plus d'importance significative après la première année de vie. Elles reflètent donc directement la demande en énergie de chaque enfant et la compétition fraternelle entre deux enfants trop rapprochés. La différence d'âge entre les parents reflète les ressources dont dispose la famille pour subvenir aux besoins de ses enfants. Plus cette différence augmente, plus on peut supposer que le père a une bonne stabilité financière à la naissance de son premier enfant. C'est donc normal que cet impact se manifeste pour la même période, celle qui demande le plus aux parents.

L'âge de la mère à la naissance de l'enfant et le rang de naissance de l'enfant sont deux variables corrélées qui augmentent le facteur de risque de mortalité. Cependant, même si elles évoluent ensemble, les deux variables n'ont ni le même impact ni la même cause. La première est très influente en bas âge quand l'enfant

est le plus demandant et la cause de son effet délétère est d'ordre biologique : plus la mère vieillit, plus elle a de chances d'avoir une mauvaise formation de gamètes. Ses chances d'avoir un enfant malade sont donc augmentées en conséquence. De plus, la résistance et l'énergie de la mère diminuent à mesure qu'elle vieillit tandis que le nombre de ses enfants, lui, augmente. L'augmentation de l'âge de la mère indique la possibilité que le rang de naissance de l'enfant soit plus élevé donc que la charge de la mère soit plus grande et que l'enfant doive partager l'attention maternelle avec plusieurs frères et soeurs.

Finalement, l'effet délétère de l'augmentation de l'année de naissance s'explique de deux façons non-exclusives. Tout d'abord par l'augmentation de la population jusqu'à saturation des terres et des ressources. Rappelons que la fraction de la population qui émigre a été exclue de notre effectif lors de la sélection des couples par les critères d'émigration et d'intervalles intergénéraliques présentés précédemment. Cette exclusion a été faite dans le but d'avoir une information fiable qui permette de créer des variables valides mais elle entraîne toutefois un biais d'échantillonnage puisque notre effectif est finalement constitué de la fraction la plus stable de la population de l'île. Dans un deuxième temps, il est important de rappeler l'effet trompeur de l'exclusion des groupes « exclus » et « couples ». En raison de la grande différence qui existe entre la date moyenne de mariage des « inclus » de l'effectif en comparaison avec celles de ces deux groupes évincés de l'étude, une source d'erreur s'est introduite dans notre effectif en augmentant la proportion des individus dont le destin était de mourir. Pour cette raison, la modélisation fait une surestimation de la mortalité (pour tous les types de mortalité) pour les années les plus récentes de l'étude en excluant plusieurs individus dont on ne connaît pas la date de décès. Ces individus ont soit émigré, soit étaient encore vivants lors de la constitution du Registre et constituaient potentiellement en majorité des survivants pour les types de mortalité étudiés. Les exclure a permis de rester très fidèle aux critères de sélection et donc de garder le contrôle sur la partie stable de la population à étudier. Cependant, ce choix a probablement

amplifié l'impact délétère de l'augmentation de l'année de naissance en biaisant à la hausse la proportion des mortalités pour la fin de la période à l'étude. On constate en plus que la présence bénéfique d'un médecin depuis l'année 1930 ne peut se manifester de façon tangible dans nos résultats à cause de ce dernier biais.

Le Registre de la population de l'Île-aux-Coudres est un outil d'analyse grâce auquel il a été possible d'obtenir avec précision toutes les variables utilisées dans cette étude. La grande profondeur généalogique dont dispose le Registre de population de l'Île-aux-Coudres, nous fait bénéficier de coefficients d'apparentement et de consanguinité dont l'estimation se rapproche énormément de la réalité biologique vécue. De plus, l'innovation de séparer les coefficients obtenus en des composantes proches et éloignées pour ensuite les confronter aux autres variables d'exposition permet d'une part de dégager leurs implications au niveau social et de les interpréter en fonction des liens humains tissés dans une population où l'endogamie fait partie de la culture et contribue à l'économie de la population. Nous avons constaté que cette implication sociale des variables génétiques s'appliquait seulement pour les variables génétiques totales et plus spécifiquement, leur composante proche (5-5 et moins pour la consanguinité et 4-4 et moins pour l'apparentement). La raison sous-jacente de ce phénomène est simple : des liens trop éloignés et dilués au fil des générations sont oubliés ou ignorés par les familles qui ne se reconnaissent plus comme liées; le lien social est brisé. D'autre part, l'impact génétique et biologique des variables d'apparentement et de consanguinité a pu être mis en valeur grâce à cette étude. Les résultats obtenus et leurs interprétations sont cohérents avec les dernières découvertes faites en génétique et en immunologie.

Tout d'abord, l'analyse des variables génétiques a fourni des informations très intéressantes sur les probabilités de mortalité de la progéniture en fonction de l'apparentement des conjoints et de la consanguinité maternelle dépendant de la période d'exposition. En effet, pour les variables génétiques (ainsi que pour les

variables socio-démographiques dans une moindre mesure), la période 0-1 an et la période 1-14 ans sont très distinctement séparées par les facteurs qui influencent les risques de mortalité.

Dans un premier temps, l'apparentement total et l'apparentement éloigné des conjoints sont délétères pour la progéniture féminine puisqu'ils augmentent les risques de mortalité infantile. L'apparentement proche n'a cependant pas d'impact significatif sur les types de mortalités mesurables dans le cadre de cette étude. On peut supposer une perte in utero de garçons ce qui nous empêche de mesurer pour eux l'impact de ces variables pour ce type de mortalité.

Cependant, pour la progéniture âgée de plus d'un an, nous avons remarqué que l'effet des variables d'apparentement avait une connotation beaucoup plus sociale que génétique. L'apparentement proche entre les conjoints correspond en effet à un lien que les gens se reconnaissent. On a donc pu voir que l'apparentement proche et l'apparentement total entraînaient un surplus de protection pour la progéniture entre 1 et 4 ans ainsi qu'entre 1 et 14 ans. Cette protection serait attribuable au soutien du réseau social envers la famille pour la progéniture (surtout masculine) qui semble être viable (i.e. qui a survécu à sa première année d'existence). L'apparentement éloigné ne se manifeste pas comme significatif dans les modèles. On peut donc supposer que cet apparentement est trop éloigné (liens 4-4 ou plus) pour que les gens le considèrent lié à des obligations et des responsabilités sociales.

Au niveau de l'impact de la consanguinité maternelle totale, nous avons pu remarquer que son effet était significativement protecteur pour la progéniture masculine et féminine âgée de moins d'un an. Ces seuls résultats ne nous permettaient pas de nous avancer davantage dans nos conclusions. La présence de l'effet remarquablement protecteur (contre la mortalité féminine) de la part éloignée de la consanguinité maternelle nous a permis de faire un lien avec l'impact

de l'apparement des conjoints et d'avancer une explication plausible de la différence entre la mortalité masculine et féminine en fonction des variables d'exposition génétiques. La consanguinité maternelle proche a également eu un impact délétère sur la probabilité de mortalité entre 1 et 4 ans chez les garçons. Ce dernier résultat demeure difficile à interpréter. La consanguinité proche de la mère entraîne un risque supplémentaire d'engendrer une progéniture homozygote pour des gènes délétères pour les fils seulement. Le niveau d'endogamie est peut-être insuffisant et l'histoire de cette population trop courte pour avoir permis d'éliminer tous les gènes délétères du pool génique de la population. S'il s'agit d'allèles délétères et récessifs appartenant à un gène du chromosome X, on peut s'attendre à ce que leur élimination soit lente et qu'ils continuent d'affecter principalement la progéniture masculine. Pourtant, les résultats issus de la comparaison des consanguinités maternelles entre les groupes exclus de l'analyse et l'effectif inclus dénote que la consanguinité maternelle éloignée a assez de poids sur la consanguinité maternelle totale pour lui permettre de demeurer stable sur une période allant de 1896 à 1907. Ce poids de la consanguinité éloignée refléterait donc une longue histoire endogame.

À la lumière de ces informations, nous pouvons affirmer que l'hypothèse principale de cette étude est confirmée puisqu'il semble que la consanguinité éloignée possédée par certaines des mères constitue une protection pour la progéniture. Cette protection se manifeste principalement contre toute mortalité périnatale (mortalité néonatale précoce) ainsi que contre la mortalité infantile. De plus, on remarque ce phénomène de façon plus évidente chez les filles. Il est donc fort probable que cet avantage sélectif transmis de la mère à sa progéniture (surtout les filles) soit le résultat direct de l'épuration des gènes délétères au fil des générations. Ainsi, lorsque la mère est apparentée à son conjoint, la sélection qu'à subi son génome vient compenser le risque de retrouver chez l'enfant des gènes récessifs létaux à l'état homozygote. Le génome de ces enfants se retrouve lui aussi épuré, ceux-ci héritent donc aussi d'une plus grande proportion d'allèles

avantageux issus de leur mère et ont moins de risques d'être des homozygotes désavantagés.

La difficulté rencontrée avec la progéniture masculine était de taille puisqu'il semble que les variables sélectionnées pour les modèles étaient moins explicatives pour les garçons que pour les filles pendant la première année de vie. Après analyse des modèles, nous sommes arrivés à la conclusion que malgré l'effet protecteur de la consanguinité éloignée, les garçons font face à un problème supplémentaire associé à l'histocompatibilité entre la mère et son foetus (interaction probable du MHC avec le chromosome Y). La grande sélection à laquelle les enfants mâles font face se passe donc in utero; ce qui explique que les modèles post-nataux soient moins parlants que ceux des filles.

L'apparement des parents rend les foetus mâles plus fragiles parce que trop compatibles avec la mère. Ce surplus de compatibilité empêche une reconnaissance réciproque de la mère et de son foetus; l'avortement spontané se produirait avant même que la grossesse soit conscientisée. Ce phénomène pourrait expliquer le surplus de naissances féminines à l'Île-aux-Coudres. Les garçons issus de parents apparentés et donc eux-mêmes compatibles ne viendraient jamais à terme et les filles issues de ce même couple pourraient voir le jour. Le problème de la compatibilité entre les antigènes de la mère et ceux de son foetus appuie l'hypothèse que les gènes du MHC ont potentiellement un lien avec le chromosome Y comme le suggèrent certaines des études déjà mentionnées. La deuxième hypothèse de la présente étude est donc également confirmée par l'analyse des résultats de modélisation mais, malheureusement, on ne peut investiguer plus loin dans cet aspect de l'étude, faute de données sur le MHC à l'Île-aux-Coudres.

Finalement, il est impossible de vérifier ce qui s'est passé in utero puisqu'aucune analyse d'ordre immunologique ou génétique n'a été effectuée dans le cadre de la

présente étude. Les résultats et les conclusions obtenus sont donc des indicateurs qui pourront orienter les prochaines recherches sur le sujet. Il a été mis en lumière qu'en plus des facteurs démographiques, les facteurs génétiques issus de l'histoire démographique et culturelle d'une population jouent un grand rôle dans la destinée de la progéniture.

De plus, il a également été démontré que, dépendamment de son origine, la consanguinité n'est pas toujours une menace pour la santé de la progéniture. Les mariages entre apparentés seraient parfois un gage de santé à la fois économique et biologique pour plusieurs sociétés qui ont entretenu ces pratiques pendant des dizaines de générations...

La modélisation par régression logistique a été un outil très efficace pour donner une représentation la plus réelle possible d'une situation à la fois sociale et biologique. Cependant, le nombre élevé de régressions effectuées ainsi que la nature stricte des critères appliqués lors de la sélection de l'effectif et le contrôle des variables nous a forcé à faire des choix qui ont entraîné de nombreuses limites d'interprétations. En effet, chaque type de mortalité étudié présente ses mécanismes propres et le type de recherche que nous avons effectué ne met pas à notre disposition tous les éléments nécessaires à l'approfondissement de ces mécanismes. Les résultats obtenus et leurs analyses constituent donc autant d'indices pour investiguer plus loin la compréhension des mécanismes complexes auxquels sont exposés les populations humaines tant sur le plan biologique que démographique.

Par conséquent, nous sommes très satisfaits de ces résultats que nous considérons assez novateurs dans ce type de recherche. En effet, ce qui a été observé est à la fois en accord avec les grandes notions démographiques et les connaissances immuno-génétiques que nous avons sur les populations. En se basant sur des coefficients d'apparentement et de consanguinité, nous croyons

avoir réussi à mettre en valeur une partie de l'impact réel de l'accumulation de liens entre apparentés dans une population.

Il s'est avéré fort intéressant de vérifier le destin des individus exclus de l'étude par l'absence de date de décès. Ces individus dont on perd le trace ont des caractéristiques différentes de ceux qui ont été conservés dans l'étude. De plus, 41 familles sélectionnées initialement, par l'entremise de la sélection des 308 couples, ont été exclues complètement de l'étude. Afin de pouvoir bien critiquer les résultats obtenus avec notre effectif, ces familles (le groupe « couples ») ainsi que l'ensemble des individus exclus (le groupe « exclus) ont été l'objet d'analyses comparatives avec les individus de l'effectif. Ces résultats sommaires ont donc sérieusement aidé à soutenir plusieurs interprétations concernant l'impact de certaines variables sur la mortalité de la progéniture (principalement pour l'explication de l'impact légèrement délétère de l'année de naissance). Il serait toutefois très enrichissant de caractériser de façon encore plus précise ces groupes qui échappent aux analyses puisqu'ils n'ont pas de date de décès enregistrée. L'idée de leur fixer des âges et des dates de décès fictifs et pondérés à partie des proportions obtenues dans l'effectif pour tous les types de mortalité étudiés semble intéressante pour tenter de voir la vraie nature de l'impact de la variable « année de naissance » mais de nombreux biais devront être contrôlés afin d'en tirer des conclusions claires. De plus, cette procédure n'est utile que dans le cadre de l'étude du même effectif que dans le présent travail. En effet, des dates fictives faussent les données des exclus et d'autres méthodes devront être mises au point pour les étudier sans modifier leur destin artificiellement.

Enfin, pour pousser plus loin l'investigation, des études démographiques prospectives incluant un volet d'analyses génétiques et immunologiques sur des populations endogames pourraient éclairer encore plus les règles complexes qui régissent la mortalité dans une population.

## **APPENDICE A**

**Liste des noms complets des variables associés à leur acronyme**

Tableau A.1: Liste des noms complets des variables utilisées dans les modèles de régression logistiques associés à leur acronyme.

Nom complet de la variable	Acronyme
Nombre d'enfants (taille de la famille)	nbenf
Différence d'âge entre les conjoints (en années)	difconj
Rang de naissance de l'enfant	rangnai
Année de naissance de l'enfant (année de calendrier)	year
Âge de la mère à la naissance de l'enfant	agemm
Intervalle génésique entre l'enfant et son aîné immédiat (en mois)	interidi
Coefficient de parenté (apparentement) total des parents (forme binaire)	xcoefp
Coefficient de consanguinité total du père (forme binaire)	xcngh
Coefficient de consanguinité total de la mère (forme binaire)	xcngf
Coefficient de parenté (apparentement) proche des parents (forme binaire)	xapp44
Coefficient de consanguinité proche du père (forme binaire)	xcngh55
Coefficient de consanguinité total de la mère (forme binaire)	xcngf55
Coefficient de parenté (apparentement) éloigné des parents (forme binaire)	xcél
Coefficient de consanguinité éloigné du père (forme binaire)	xchél
Coefficient de consanguinité total de la mère (forme binaire)	xcfél

## **APPENDICE B**

### **Corrélations entre les variables utilisées**

Correlations					
Variable	nbenf	datem	rangnai	year	agemn
nbenf	1.0000	0.1654	0.4079	0.2155	0.0192
datem	0.1654	1.0000	0.0624	0.9827	-0.0401
rangnai	0.4079	0.0624	1.0000	0.2368	0.8037
year	0.2155	0.9827	0.2368	1.0000	0.1182
agemn	0.0192	-0.0401	0.8037	0.1182	1.0000

Figure B.1 : Table de corrélation entre certaines variables socio-démographiques : le nombre d'enfants du couple (nbenf), la date de mariage des conjoints (datem), le rang de naissance de chaque enfant (rangnai), l'âge de la mère à la naissance de l'enfant (agemn) et l'année de naissance de l'enfant (year).

Correlations					
Variable	coefp	entrop	varian	cngf	cngh
coefp	1.0000	0.3018	-0.1225	0.0769	0.0808
entrop	0.3018	1.0000	0.1046	0.2669	0.1413
varian	-0.1225	0.1046	1.0000	0.0186	-0.0200
cngf	0.0769	0.2669	0.0186	1.0000	-0.1000
cngh	0.0808	0.1413	-0.0200	-0.1000	1.0000

Figure B.2 : Table de corrélation entre certaines variables généalogiques : le coefficient de parenté (coefp), le coefficient de consanguinité du père (cngf), le coefficient de consanguinité de la mère (cngh), l'entropie (entrop) et la variance (varian).

Correlations		
Variable	nbenf	interidi
nbenf	1.0000	-0.1189
interidi	-0.1189	1.0000

Figure B.3 : Table de corrélation entre le nombre d'enfants du couple et l'intervalle génésique.

## **APPENDICE C**

### **Distribution des variables utilisées**

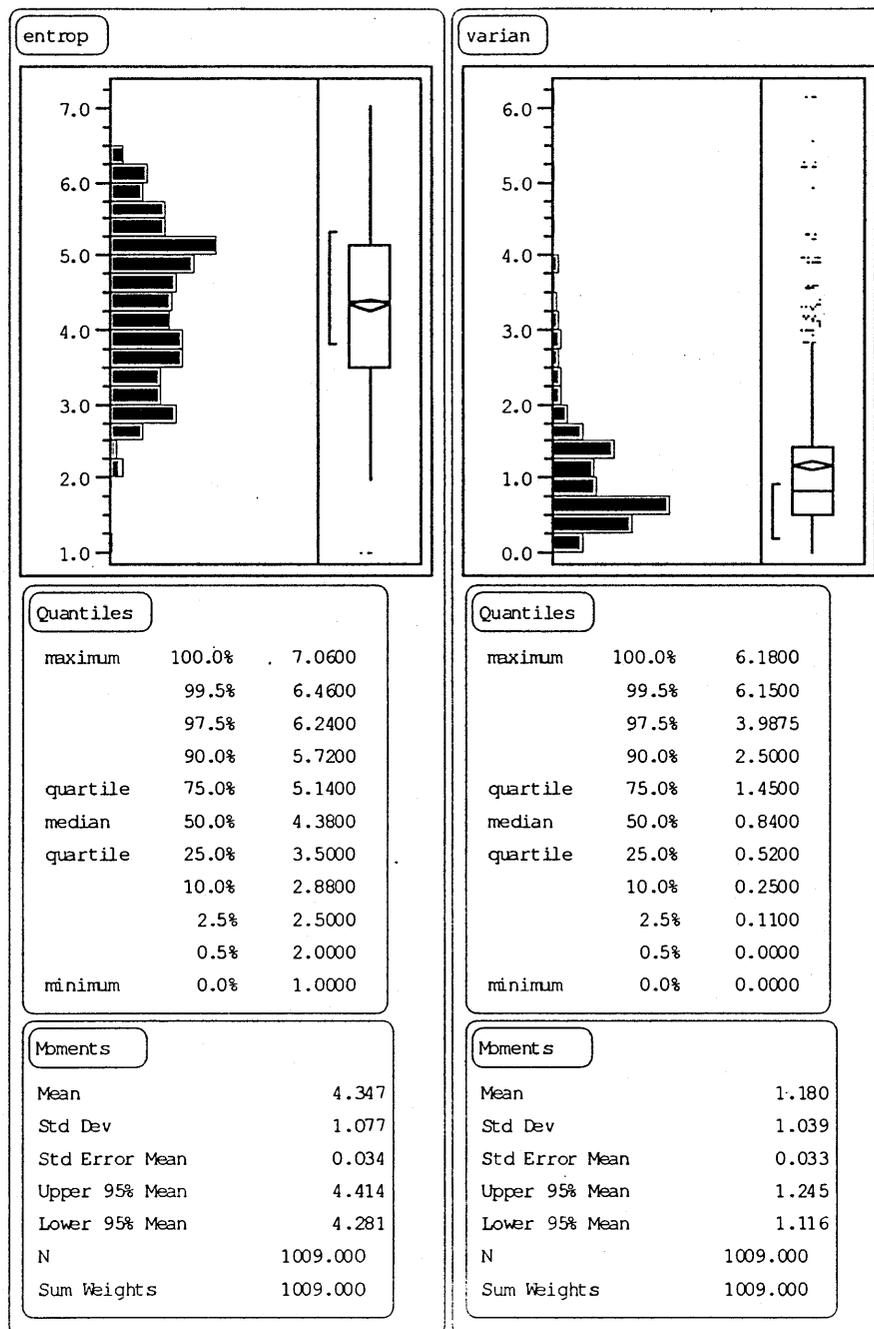
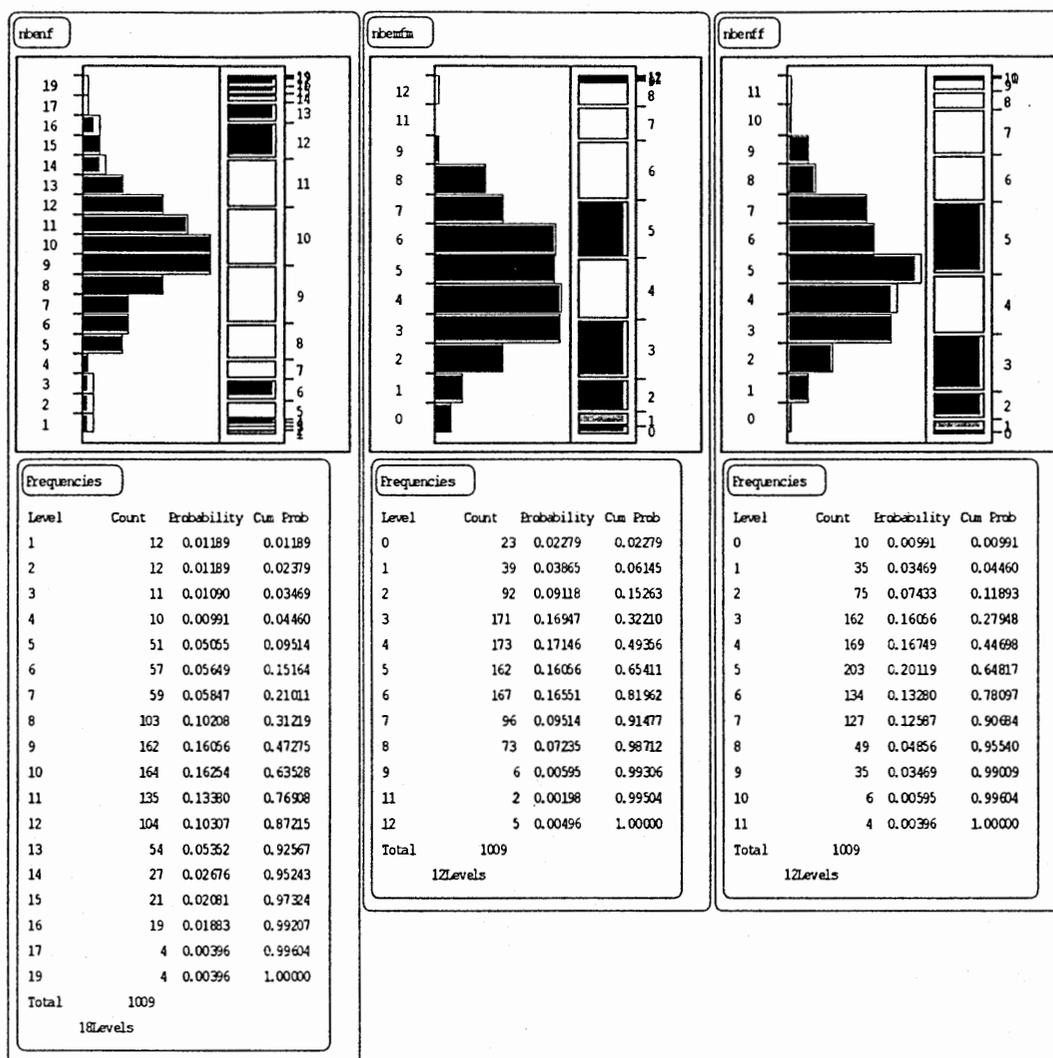
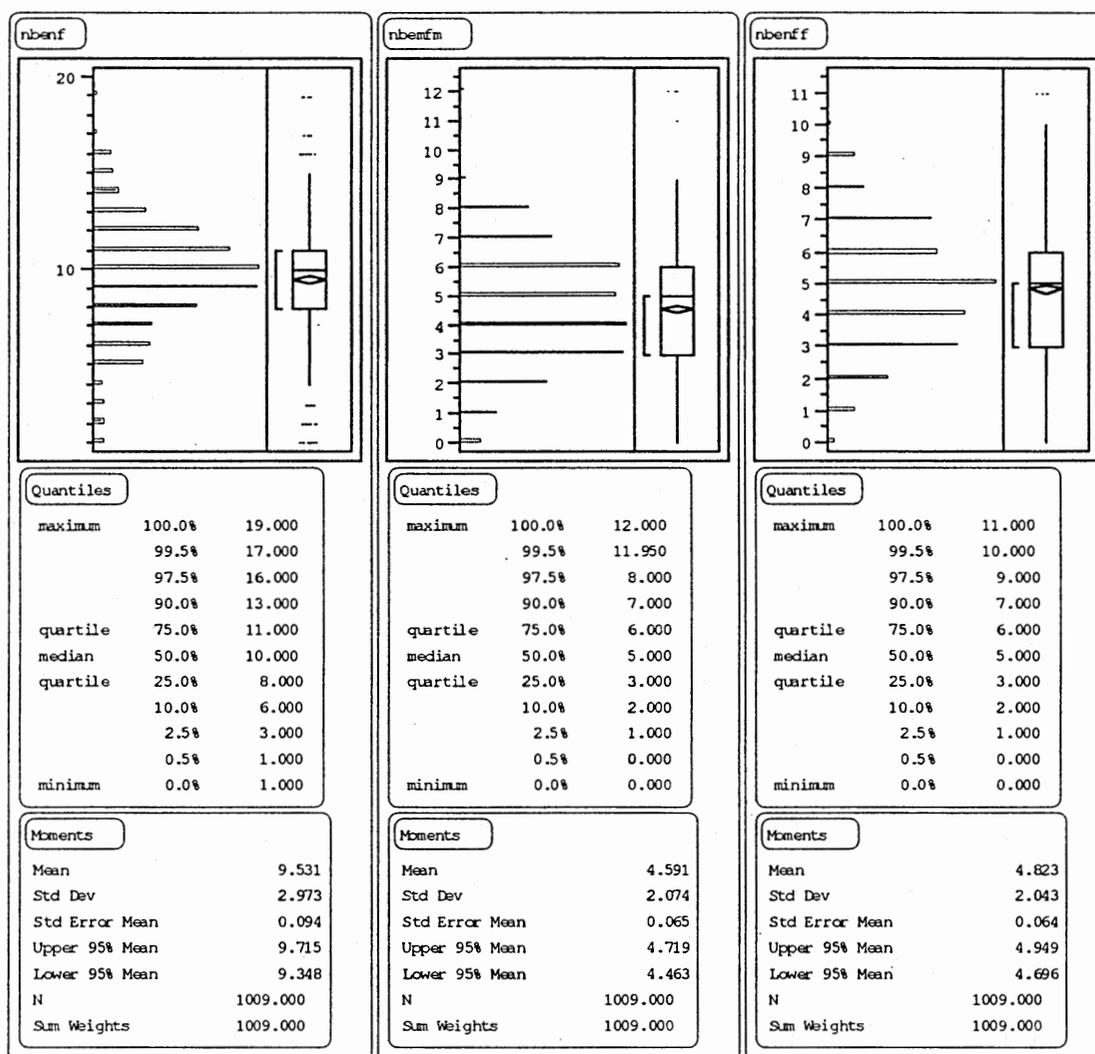


Figure C.1 : Moyennes et écart-types des variables entropie (à gauche) et variance (à droite).



a)

Figure C.2: Distribution des fréquences des variables nombre d'enfants total (à gauche), nombre d'enfants masculins (au centre) et nombre d'enfants féminins (à droite); a) présentation détaillée de chaque catégorie, b) présentation des moyennes et écart-types.



b)

Figure C.2: Distribution des fréquences des variables nombre d'enfants total (à gauche), nombre d'enfants masculins (au centre) et nombre d'enfants féminins (à droite); a) présentation détaillée de chaque catégorie, b) présentation des moyennes et écart-types.

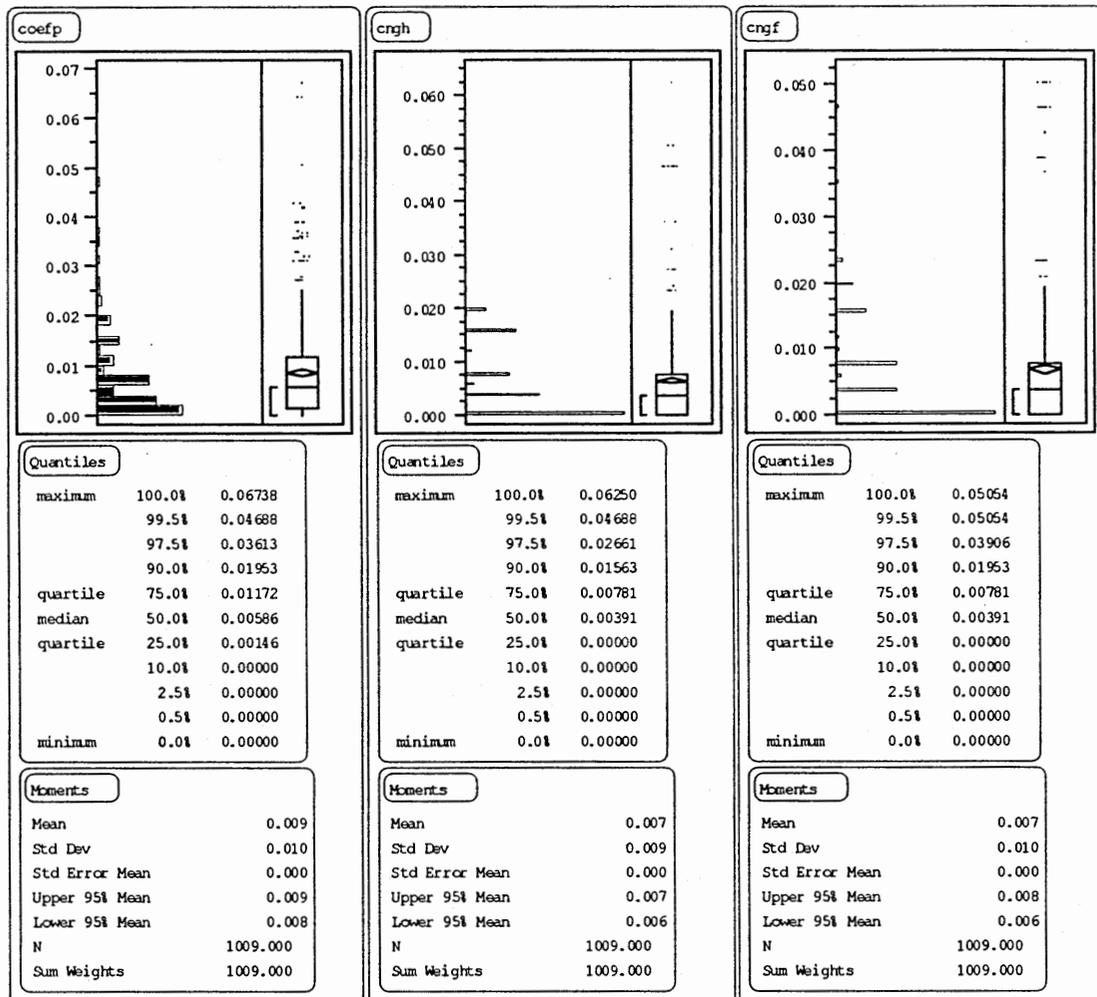


Figure C.3 : Distribution, moyenne et écart-type des variables génétiques totales : le coefficient de parenté, le coefficient de consanguinité du père et le coefficient de consanguinité de la mère.

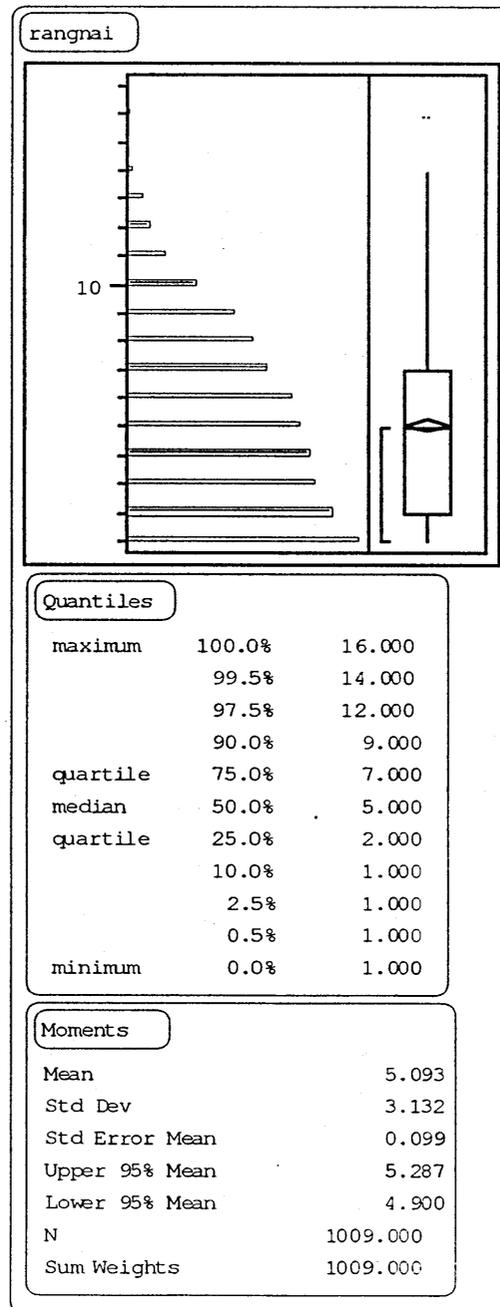
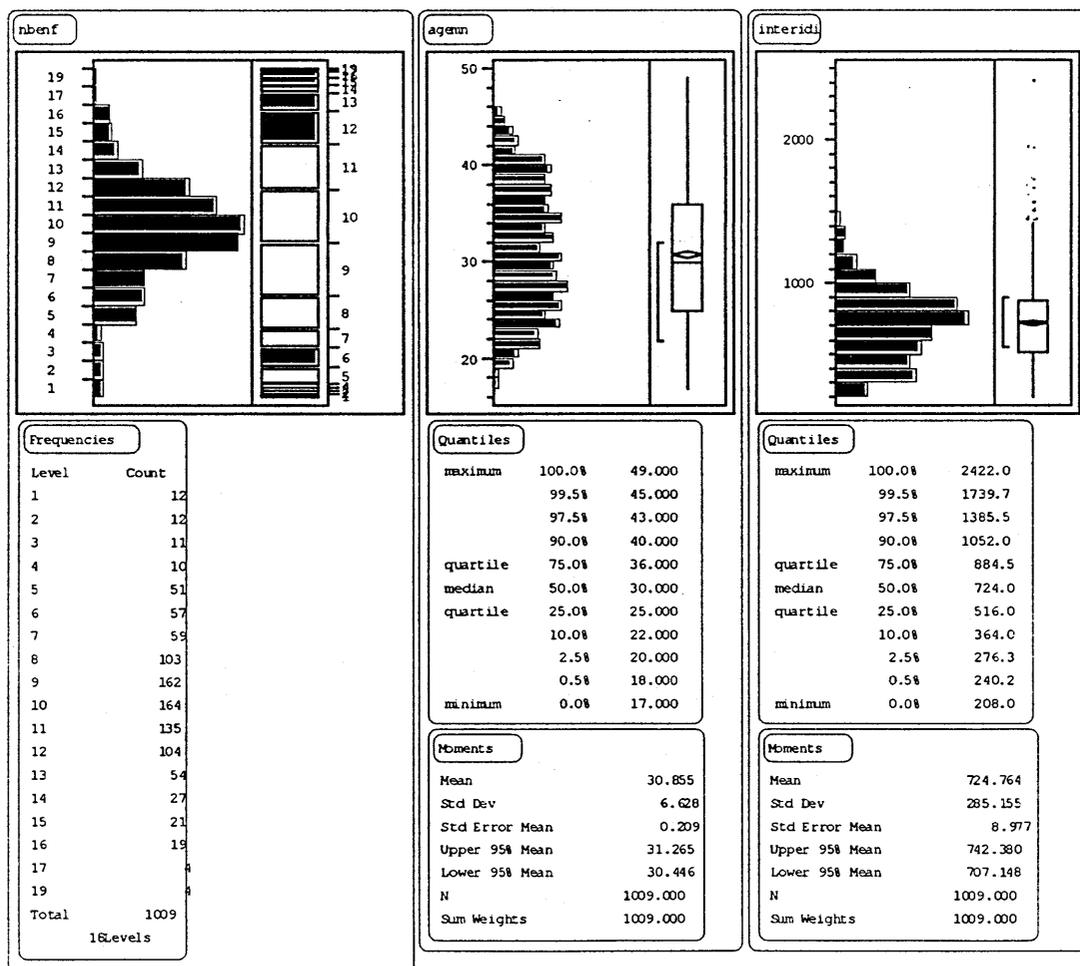
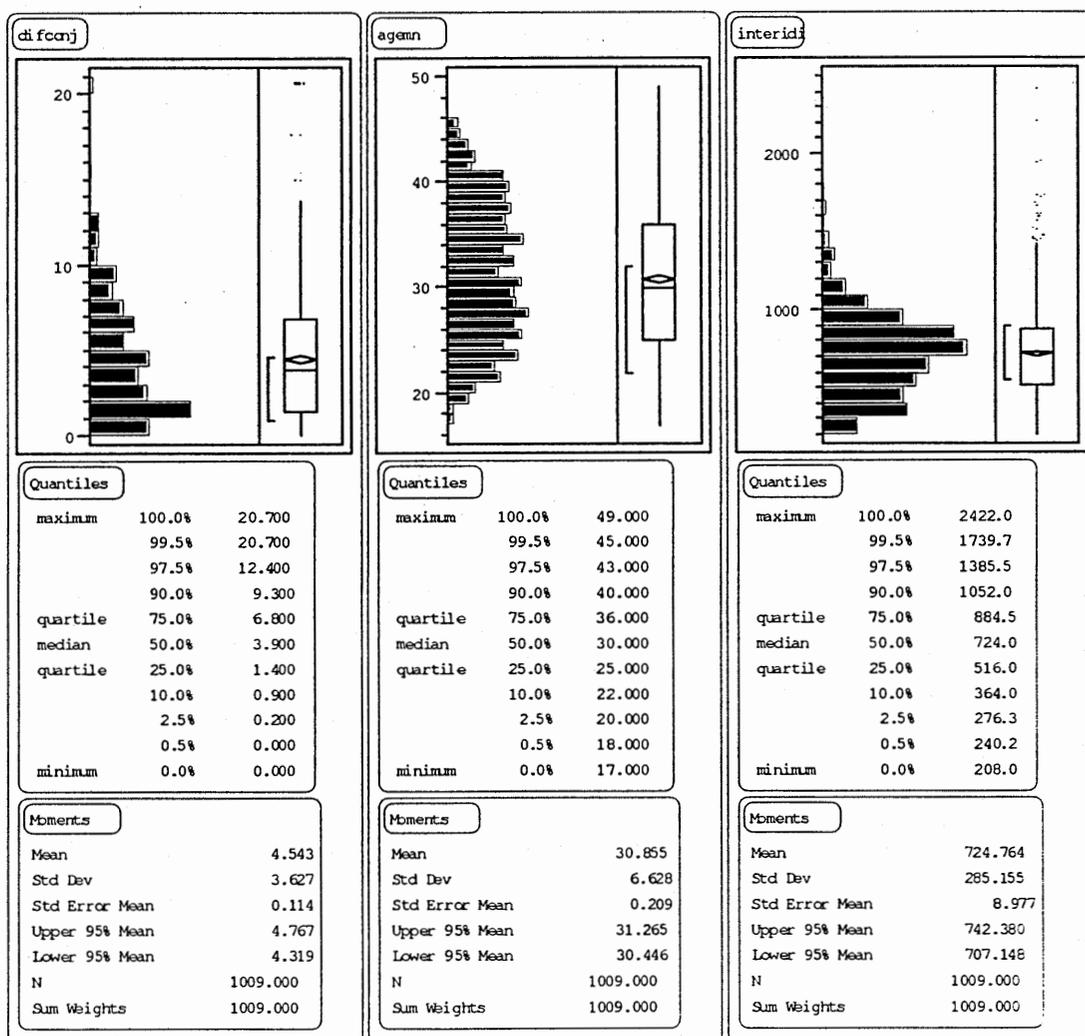


figure C.4 : Moyenne et écart-type de la variable rang de naissance.



a)

Figure C.5 : Distribution des variables différence d'âge entre les conjoints (à gauche), âge de la mère à la naissance de l'enfant (au centre) et intervalle génésique (à droite); a) présentation détaillée de chaque catégorie et b) présentation des moyennes et des écart-types.



b)

Figure C.5 : Distribution des variables différence d'âge entre les conjoints (à gauche), âge de la mère à la naissance de l'enfant (au centre) et intervalle génésique (à droite); a) présentation détaillée de chaque catégorie et b) présentation des moyennes et des écart-types.

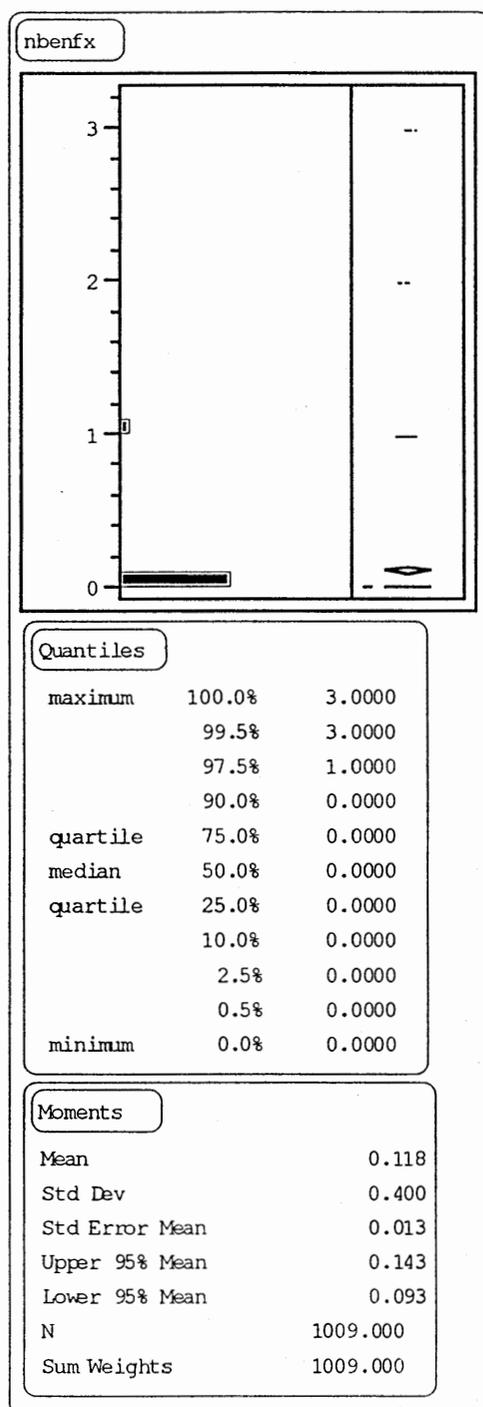


Figure C.6: Distribution des fréquences de la variable nombre d'enfants de sexe inconnu présentation de la moyenne et de l'écart-type.

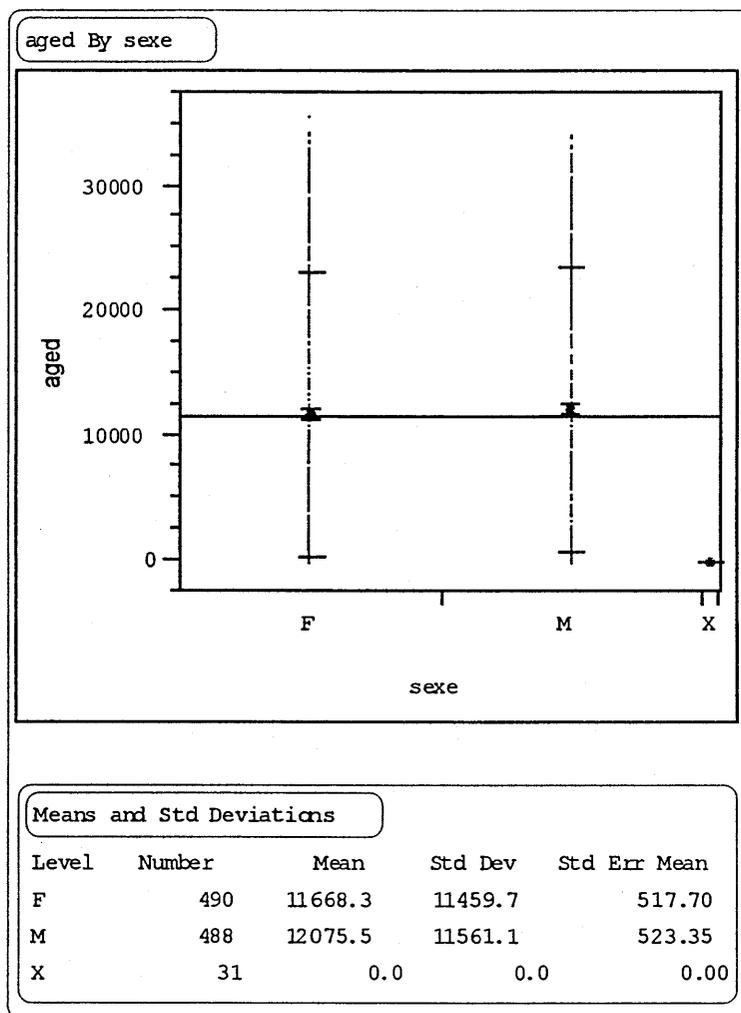


fig C.7. Représentation des moyennes et des écart-types de l'âge moyen au décès (en jours) chez les filles, les garçons et les enfants de sexe inconnu de l'effectif.

## **APPENDICE D**

**Analyse de la mortalité par rapport à l'année de naissance**

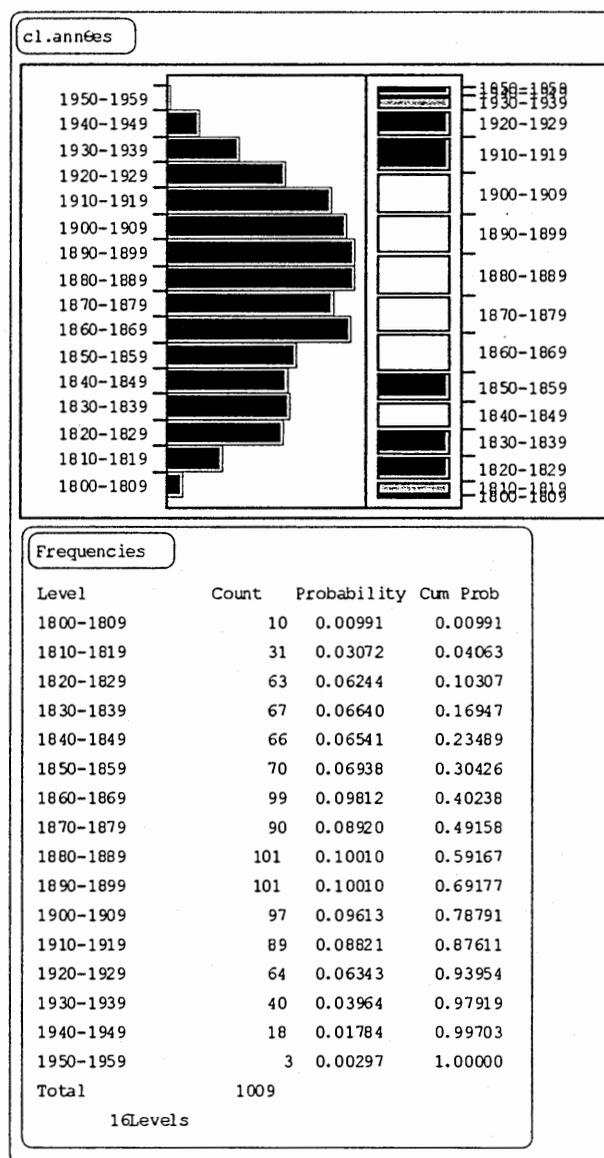


Figure D.1 : Distribution de la variable année de naissance par classes de 10 ans.

mort01 By cl.années		cl.années																	
Crosstabs		1800-1809	1810-1819	1820-1829	1830-1839	1840-1849	1850-1859	1860-1869	1870-1879	1880-1889	1890-1899	1900-1909	1910-1919	1920-1929	1930-1939	1940-1949	1950-1959		
Count		10	31	57	62	62	67	98	89	98	95	89	79	61	32	12	2	944	
Expected		9.3558	29.003	58.9415	62.6838	61.7483	65.4906	92.6224	84.2022	94.4936	94.4936	90.7512	83.2666	59.8771	37.4232	16.8404	2.80674		
Cell Chi^2		0.0444	0.1375	0.0640	0.0075	0.0010	0.0348	0.3122	0.2734	0.1301	0.0027	0.0338	0.2186	0.0211	0.7859	1.3913	0.2319		
		0	0	6	5	4	3	1	1	3	6	8	10	3	8	6	1	65	
1		0.6442	1.99703	4.05847	4.31615	4.25173	4.50942	6.3776	5.79782	6.50644	6.50644	6.24876	5.7334	4.12289	2.57681	1.15956	0.19226		
		0.6442	1.9970	0.9288	0.1083	0.0149	0.5052	4.5344	3.9703	1.8897	0.0394	0.4908	3.1751	0.3068	11.4137	20.2057	3.3676	1009	
		10	31	63	67	66	70	99	90	101	101	97	89	64	40	18	3		

Tests	
Source	DF -LogLikelihood RSquare (R)
Model	15 24.30057 0.1008
Error	993 216.91061
C Total	1008 241.11118
Total Count	1009
Test	ChiSquare ProbChiSq
Likelihood Ratio	48.601 <.0001
Pearson	57.281 <.0001

Warning: 20% of cells have expected count less than 5, Chi-squares suspect

Figure D.2 : Table de contingence de la mortalité de la progéniture totale par tranches de 10 ans.

mott1-6j By cl.années		cl.années															
Crosstabs		1800-1809	1810-1819	1820-1829	1830-1839	1840-1849	1850-1859	1860-1869	1870-1879	1880-1889	1890-1899	1900-1909	1910-1919	1920-1929	1930-1939	1940-1949	1950-1959
Count		10	31	61	66	66	68	98	90	100	100	95	87	63	34	18	3
Expected		9.81169	30.4163	61.8137	65.7384	64.7572	68.6619	97.1358	88.3053	99.0981	99.0981	95.1734	87.3241	62.7948	39.2468	17.6511	2.94351
Cell Chi-2		0.0036	0.0112	0.0107	0.0010	0.0239	0.0068	0.0077	0.0225	0.0082	0.0082	0.0003	0.0002	0.0007	0.7014	0.0065	0.0001
0		0	0	0	0	0	2	1	0	1	1	2	2	1	2	0	0
1		0.18831	0.58375	1.18632	1.26165	1.24281	1.31814	1.86422	1.69475	1.90188	1.90188	1.82656	1.67592	1.20515	0.75322	0.33895	0.05649
2		0.1883	0.5837	0.5581	0.0543	1.2428	0.3527	0.4006	1.6947	0.4277	0.4277	0.0165	0.0627	0.0349	36.5480	0.3389	0.0565
Total		10	31	63	67	66	70	99	90	101	101	97	89	64	40	18	3

Tests	
Source	DF -LogLikelihood RSquare (0)
Model	15 12.965838 0.1375
Error	993 81.327383
C Total	1008 94.293221
Total Count	1009
Test	ChiSquare ProbChiSq
Likelihood Ratio	25.932 0.0387
Pearson	43.813 0.0001

Warning: 20% of cells have expected count less than 5. Chi-squares suspect

Figure D.3 : Table de contingence de la mortalité néonatale précoce de la progéniture totale par tranches de 10 ans.

mortinf by cl.années		cl.années															
Count	1800-1809	1810-1819	1820-1829	1830-1839	1840-1849	1850-1859	1860-1869	1870-1879	1880-1889	1890-1899	1900-1909	1910-1919	1920-1929	1930-1939	1940-1949	1950-1959	
Expected	8	25	49	52	48	60	90	78	88	83	64	57	43	12	4		
Cell Chi-2	7.55203	23.4113	47.5778	50.5986	49.8434	52.8642	74.7651	67.9683	76.2755	76.2755	73.2547	67.2131	48.333	30.2001	13.5937	2.26561	
	0.0266	0.1078	0.0425	0.0388	0.0682	0.9632	3.1044	1.4806	1.8022	0.5228	1.1692	1.5519	0.5984	10.9751	6.7707	0.7070	
3	2.44797	7.5887	15.4222	16.4014	16.1566	17.1358	24.2349	22.0317	24.7245	24.7245	23.7453	21.7869	15.667	9.79187	4.40634	0.73439	
	0.0820	0.3326	0.1312	0.1197	0.2103	2.9715	9.5772	4.5677	5.5598	1.8289	3.6070	4.7876	1.8153	33.8583	20.8877	2.1811	
	10	31	63	67	66	70	99	90	101	101	97	89	64	40	18	3	
Total	109	108	109	108	108	109	109	108	109	109	108	108	109	108	109	109	

Tests	
Source	DF -LogLikelihood RSquare (U)
Model	15 56.61939 0.1008
Error	993 504.93397
C. Total	1008 561.55526
Total Count	1009
Test	ChiSquare ProbChiSq
likelihood Ratio	113.239 <.0001
Pearson	122.507 <.0001

Figure D.4 : Table de contingence de la mortalité infantile de la progéniture totale par tranches de 10 ans.

mort-4 By ci. années		ci. années															
Crosstabs		1800-1809	1810-1819	1820-1829	1830-1839	1840-1849	1850-1859	1860-1869	1870-1879	1880-1889	1890-1899	1900-1909	1910-1919	1920-1929	1930-1939	1940-1949	1950-1959
Count		10	31	55	63	64	61	82	79	84	94	89	71	54	35	16	891
Expected		8.87953	27.3746	55.6323	59.1645	58.2815	61.8137	87.4222	79.4747	89.1883	89.1883	85.6561	78.5917	56.5154	35.3221	15.8949	2.64916
Cell Chi^2		0.1519	0.4801	0.0072	0.2486	0.5611	0.0107	0.3383	0.0028	0.3018	0.2596	0.1395	0.7333	0.1120	0.0029	0.0007	0.0465
0																	
4		1.16947	3.62537	7.36769	7.83548	7.71853	8.18632	11.5778	10.5253	11.8117	11.8117	11.3439	10.4083	7.48464	4.6779	2.10505	0.35084
1		1.1695	3.6254	0.0543	1.8775	4.2368	0.0809	2.5394	0.0214	2.2790	1.9001	0.9857	5.5273	0.8403	0.0222	0.0052	0.35084
Total		10	31	63	67	66	70	99	90	101	101	97	89	64	40	18	1009

Tests			
Source	DF	-LogLikelihood	Resquare (U)
Model	15	17.54396	0.0482
Error	993	346.50182	
C Total	1008	364.04578	
Total Count	1009		
Test	ChiSquare	Prob>ChiSq	
likelihood Ratio	35.068	0.0024	
Pearson	28.90	0.0162	

Figure D.5 : Table de contingence de la mortalité entre 1 et 4 ans de la progéniture totale par tranches de 10 ans.

mort1-14 By cl.années		cl. années															
Crosstabs		1800-1809	1810-1819	1820-1829	1830-1839	1840-1849	1850-1859	1860-1869	1870-1879	1880-1889	1890-1899	1900-1909	1910-1919	1920-1929	1930-1939	1940-1949	1950-1959
Count		10	30	51	59	61	60	71	70	80	83	82	61	41	32	15	
Expected		8.00793	24.8246	50.45	53.6531	52.8523	56.0555	79.2785	72.0714	80.8801	80.8801	77.6769	71.2706	51.2507	32.0317	14.4143	2.40238
Cell Chi^2		0.4956	1.0790	0.0060	0.5329	1.2560	0.2776	0.8645	0.0595	0.0096	0.0556	0.2406	1.4801	2.0503	0.0000	0.0238	0.0674
0																	
5		0	1	12	8	5	10	28	20	21	18	15	28	23	8	3	1
10		1.99207	6.17542	12.55	13.3469	13.1477	13.9445	19.7215	17.9286	20.1199	20.1199	19.3231	17.7294	12.7493	7.96829	3.58573	0.59762
14		1.9921	4.3374	0.0241	2.1420	5.0491	1.1158	3.4751	0.2393	0.0285	0.2234	0.9672	5.9497	8.2419	0.0001	0.0957	0.2709
1009		10	31	63	67	66	70	99	90	101	101	97	89	64	40	18	3

Tests			
Source	DF	-LogLikelihood	RSquare (U)
Model	15	23.14734	0.0469
Error	993	480.64769	
C Total	1008	503.79503	
Total Count	1009		
Test	ChiSquare	ProbChiSq	
Likelihood Ratio	46.295	<.0001	
F Pearson	42.660	0.0002	

Figure D.6 : Table de contingence de la mortalité pré-reproductive de la progéniture totale par tranches de 10 ans.

mort0m By cl.années		cl.années															
Crosstabs		1800-1809	1810-1819	1820-1829	1830-1839	1840-1849	1850-1859	1860-1869	1870-1879	1880-1889	1890-1899	1900-1909	1910-1919	1920-1929	1930-1939	1940-1949	1950-1959
Count		4	11	27	33	27	30	52	48	51	45	41	40	26	13	5	1
Expected		3.72131	10.2336	26.9795	31.6311	26.9795	30.7008	48.3717	44.6557	48.377	44.6557	40.9344	42.7951	26.0492	17.6762	9.30328	0.93033
Cell Chi^2		0.0209	0.0574	0.0200	0.0592	0.0200	0.0160	0.2713	0.2505	0.1422	0.027	0.0001	0.1826	0.0001	1.2271	1.9205	0.0062
0																	
1		0	0	2	1	2	3	0	0	1	3	3	6	2	6	5	0
Expected		0.27869	0.7664	2.02049	2.36885	2.02049	2.2918	3.62295	3.34426	3.62295	3.34426	3.06557	3.20492	1.95082	1.32377	0.69672	0.06967
Cell Chi^2		0.2787	0.7664	0.0202	0.7910	0.0202	0.2136	3.6230	3.3443	1.8990	0.0264	0.004	2.4377	0.0012	16.5188	26.5791	0.0697
0																	
1		4	11	29	34	29	33	52	48	52	48	44	46	28	19	10	1
Total		4	11	29	34	29	33	52	48	52	48	44	46	28	19	10	1

Tests	
Source	DF -Loglikelihood RSquare (U)
Model	15 23.32843 0.1891
Error	472 100.03310
C Total	487 123.36153
Total Count	488
Test	ChiSquare ProbChiSq
Likelihood Ratio	46.657 <.0001
Pearson	60.795 <.0001

Warning: 20% of cells have expected count less than 5, Chi-squares suspect

Figure D.7: Table de contingence de la mortalité masculine par tranches de 10 ans.

portof by cl.anses

Crosstabs

Count	cl.anses																
	1800-1809	1810-1819	1820-1829	1830-1839	1840-1849	1850-1859	1860-1869	1870-1879	1880-1889	1890-1899	1900-1909	1910-1919	1920-1929	1930-1939	1940-1949	1950-1959	
Expected	6	20	28	26	31	36	46	39	47	47	46	34	29	17	7	1	460
Cell Chi^2	5.6385	18.7755	30.0408	28.1633	30.9796	33.7959	44.1224	37.551	46.9388	47.8776	47.8776	34.7347	28.1633	17.8367	7.5102	1.87755	
	0.0240	0.0799	0.1386	0.1662	0.0000	0.1637	0.0799	0.0559	0.0217	0.0001	0.0736	0.0155	0.0249	0.0093	0.0047	0.4102	
1	0	0	4	4	2	0	1	1	2	3	5	3	1	2	1	1	30
	0.36735	1.22449	1.95918	1.83673	2.02041	2.20408	2.87755	2.44898	3.06122	3.12245	3.12245	2.26531	1.83673	1.16327	0.4898	0.12245	
	0.3673	1.2245	2.1259	2.5478	0.0002	2.2041	1.2251	0.8373	0.3333	0.0012	1.1290	0.2393	0.3812	0.6039	0.5313	6.2891	
Total Count	6	20	32	30	33	36	47	40	49	50	51	37	30	19	8	2	490

Tests

Source	DF	-LogLikelihood	RSquare (U)
Model	15	10.30826	0.0913
Error	474	102.55028	
C Total	489	112.85854	
Total Count	490		
Test	chi-square	ProbChiSq	
likelihood ratio	20.617	0.1495	
Wald	21.366	0.1255	

Warning: 20% of cells have expected count less than 5, Chi-squares suspect

Figure D.8 : Table de contingence de la mortalité féminine par tranches de 10 ans.

mort0-6jm By ci.années		ci.années															
CrossTabs		1800-1809	1810-1819	1820-1829	1830-1839	1840-1849	1850-1859	1860-1869	1870-1879	1880-1889	1890-1899	1900-1909	1910-1919	1920-1929	1930-1939	1940-1949	1950-1959
Count	Expected	4	11	28	34	29	33	52	48	51	48	43	45	28	16	10	
Cell Chi-2		3.94262	10.8422	28.984	30.5123	28.984	30.5266	51.2541	47.3115	51.2541	47.3115	43.3689	45.3402	27.5984	18.7275	9.8566	
		0.008	0.023	0.0119	0.0071	0.0061	0.0059	0.0109	0.0100	0.0013	0.0100	0.0031	0.0056	0.0068	0.3972	0.0021	0.0002
2		0	0	1	0	0	0	0	0	1	0	1	1	0	3	0	0
		0.05738	0.15779	0.41598	0.4877	0.41598	0.47336	0.7459	0.68852	0.7459	0.68852	0.63115	0.65984	0.40164	0.27254	0.14344	0.01434
		0.0574	0.1578	0.8199	0.4877	0.4160	0.4734	0.7459	0.6885	0.0866	0.6885	0.2156	0.1754	0.4016	27.2951	0.1434	0.0143
		4	11	29	34	29	33	52	48	52	48	44	46	28	19	10	1
																	488

Test	
Source	DF -LogLikelihood RSquare (U)
Model	15 9.491456 0.2589
Error	472 27.168394
C Total	487 36.660390
Total Count	488
Test	ChiSquare ProbChiSq
LiKelihood Ratio	18.993 0.2145
Barnson	33.345 0.0002

Warning: 20% of cells have expected count less than 5, Chi-squares suspect

Figure D.9 : Table de contingence de la mortalité néonatale précoce masculine par tranches de 10 ans.

mort0-6f By cl.années		cl.années																
Count	Expected	1800-1809	1810-1819	1820-1829	1830-1839	1840-1849	1850-1859	1860-1869	1870-1879	1880-1889	1890-1899	1900-1909	1910-1919	1920-1929	1930-1939	1940-1949	1950-1959	
0	5.85306	6	20	31	29	33	34	46	40	49	49	50	36	29	16	8	2	478
	0.0077	0.0213	0.0223	0.0015	0.0024	0.0203	0.0356	0.0005	0.0216	0.0301	0.0010	0.0012	0.0002	0.0024	0.3466	0.0049	0.0012	
1	0.14634	0	0	1	1	0	2	1	1	0	1	1	1	1	3	0	0	12
	0.1469	0.4898	0.4898	0.78367	0.73469	0.80816	0.88163	1.15102	0.97959	1.2	1.22449	1.24898	0.90612	0.73469	0.46531	0.19592	0.04898	
				0.0597	0.0968	0.8092	1.4187	0.0198	0.9796	1.2000	0.0412	0.0496	0.0097	0.0268	13.8074	0.1959	0.0490	
		6	20	32	30	33	36	47	40	49	50	51	37	30	19	8	2	490

Tests	
Source	DF -LogLikelihood RSquare (R)
Model	15 7.875361 0.1397
Error	474 48.490471
C Total	489 56.365832
Total Count	490
Test	ChiSquare Prob>ChiSq
Likelihood Ratio	15.751 0.3988
Pearson	19.956 0.1736
Warning: 20% of cells have expected count less than 5, Chi-squares suspect	

Figure D.10 : Table de contingence de la mortalité néonatale précoce féminine par tranches de 10 ans.

Mortality by cl.enfres		cl.enfres															
Crosstabs		1800-1809	1810-1819	1820-1829	1830-1839	1840-1849	1850-1859	1860-1869	1870-1879	1880-1889	1890-1899	1900-1909	1910-1919	1920-1929	1930-1939	1940-1949	1950-1959
Count		3	8	21	31	21	28	47	43	48	38	33	28	16	6	2	
Expected		3.05738	8.40779	22.166	25.9877	22.166	25.2234	39.7459	36.6885	39.7459	36.6885	33.631	35.1598	21.4016	14.5225	7.64344	0.76404
Cell Chi^2		0.0011	0.0198	0.0613	0.9667	0.0613	0.3057	1.3240	1.0858	1.7141	0.0469	0.0118	1.4580	1.3633	5.0014	4.1668	0.7643
Count		1	3	8	3	8	5	5	5	4	10	11	18	12	13	8	1
Expected		0.94262	2.59221	6.83402	8.0123	6.83402	7.77664	12.2541	11.3115	12.2541	11.3115	10.389	10.8402	6.59836	4.47746	2.35666	0.23366
Cell Chi^2		0.0035	0.0641	0.1989	3.1356	0.1989	0.9914	4.2942	3.5216	5.5598	0.1321	0.0384	4.7290	4.4220	16.2221	13.5148	2.4791
Total		4	11	23	34	23	33	52	48	52	48	44	46	28	19	10	1

Tests	
Source	DF -LogLikelihood RSquare (U)
Model	15 36.74685 0.1379
Error	472 229.71112
C Total	487 266.45797
Total Count	488
Test	ChiSquare ProbChiSq
Likelihood Ratio	73.494 <.0001
NPearson	77.878 <.0001

Warning: 20% of cells have expected count less than 5, Chi-squares suspect

Figure D.11 : Table de contingence de la mortalité infantile masculine par tranches de 10 ans.

mortInf By ci.annes		ci.annes															
Cronologia		1800-1809	1810-1819	1820-1829	1830-1839	1840-1849	1850-1859	1860-1869	1870-1879	1880-1889	1890-1899	1900-1909	1910-1919	1920-1929	1930-1939	1940-1949	1950-1959
Count	Expected	5	17	26	20	25	31	43	33	40	42	30	26	22	5	2	
Cell Chi2		4.50612	15.0204	24.0227	22.5306	24.7837	27.0357	35.298	30.0408	36.8	37.551	36.32	27.7878	22.5306	14.2694	6.00816	1.50204
0		0.0541	0.2609	0.1610	0.2842	0.0019	0.3820	1.6806	0.2913	0.2783	0.5271	1.7993	0.1110	0.0123	6.0214	2.6739	0.1678
3		1.49288	4.97959	7.96735	7.46939	8.21633	8.96327	11.702	9.95918	12.2	12.449	12.698	9.21224	7.46939	4.73061	1.99184	0.49796
MortInf		0.1633	0.7870	0.4858	0.8574	0.0057	1.7324	5.0693	0.8793	0.8393	1.5900	5.4279	0.3469	0.0377	18.1629	8.0656	0.5062
		6	20	32	30	33	36	47	40	49	50	51	37	30	19	8	2
		368															
		122															
		490															

**Tests**

Source DF -LogLikelihood RSquare (U)

Model 15 27.63159 0.1005

Error 474 247.36197

C Total 489 274.99355

Total Count 490

Fret ChiSquare ProbChiSq

Likelihood Ratio 55.263 <.0001

Bonferroni 59.887 <.0001

Warning: 20% of cells have expected count less than 5, Chi-square suspect

Figure D.12 : Table de contingence de la mortalité infantile féminine par tranches de 10 ans.

Mortale by cl.ans		cl.ans																
Crosstabs		1800-1809	1810-1819	1820-1829	1830-1839	1840-1849	1850-1859	1860-1869	1870-1879	1880-1889	1890-1899	1900-1909	1910-1919	1920-1929	1930-1939	1940-1949	1950-1959	
Count		4	11	26	32	23	27	43	42	42	44	38	39	25	17	10	1	480
Expected		3.55459	9.6242	25.5533	29.979	25.5533	29.079	45.8197	42.2961	45.8197	42.2961	36.7705	40.538	24.6721	16.7418	8.81148	0.88115	
Cell Chi^2		0.0641	0.1763	0.0078	0.1390	0.4643	0.1485	0.1735	0.0021	0.3184	0.0687	0.0153	0.0580	0.0044	0.0040	0.1503	0.0160	
		0	0	3	2	0	6	9	6	10	4	6	7	3	2	0	0	58
Total		4	11	29	34	29	33	52	48	52	48	44	46	28	19	10	1	488
Test		0.47541	1.3078	3.44672	4.0408	3.44672	3.92213	6.18033	5.70492	6.18033	5.70492	5.22961	5.46721	3.32787	2.2982	1.18852	0.11885	
ChiSquare		0.4754	1.3074	0.0579	1.0308	3.4467	1.1008	1.2864	0.0153	2.3607	0.5095	0.1135	0.4237	0.0323	0.0295	1.1885	0.1189	
ProbChiSq																		
Likelihood Ratio		21.409																
Rearson		15.325																

Warning: 20% of cells have expected count less than 5, Chi-squares suspect

DF -LogLikelihood RSquare (U)  
 Model 15 10.70434 0.0602  
 Error 472 167.23624  
 C Total 487 177.94058  
 Total Count 488

Figure D.13 : Table de contingence de la mortalité masculine entre 1 et 4 ans par tranches de 10 ans.

max1-f By cl.années		cl.années															
Crosstabs		1800-1809	1810-1819	1820-1829	1830-1839	1840-1849	1850-1859	1860-1869	1870-1879	1880-1889	1890-1899	1900-1909	1910-1919	1920-1929	1930-1939	1940-1949	1950-1959
Count	Expected	6	20	28	28	32	33	39	36	42	47	49	28	25	16	6	
	Cell Chi*2	5.35102	17.8367	28.5388	26.7551	29.4306	32.1061	41.9163	35.6735	43.7	44.5918	45.4837	32.998	26.7551	16.9449	7.13469	1.78367
		0.0787	0.2624	0.0102	0.0579	0.2243	0.0249	0.2029	0.0020	0.0661	0.1301	0.2718	0.7570	0.1151	0.0527	0.1803	0.0262
4		0	0	4	2	1	3	8	4	7	3	2	9	5	3	2	0
5		0.64898	2.16327	3.46122	3.2449	3.56939	3.89388	5.09367	4.32653	5.3	5.40816	5.51633	4.00204	3.2449	2.0551	0.86531	0.21633
6		0.6490	2.1633	0.0839	0.4716	1.8495	0.2052	1.6730	0.0246	0.5453	1.0723	2.2414	6.2417	0.9493	0.4344	1.4879	0.2163
	Total	6	20	32	30	33	36	47	40	49	50	51	37	30	19	8	2

Tests	
Source	DF -LogLikelihood RSquare (U)
Model	15 12.53219 0.0746
Error	474 155.36917
C Total	489 167.90236
Total Count	490
Test	ChiSquare Prob-CHISq
Likelihood Ratio	25.066 0.0491
Pearson	22.779 0.0890

Warning: 20% of cells have expected count less than 5, Chi-squares suspect

Figure D.14 : Table de contingence de la mortalité féminine entre 1 et 4 ans par tranches de 10 ans.

mortality by ci. années																
ci. années																
Count	1800-1809	1810-1819	1820-1829	1830-1839	1840-1849	1850-1859	1860-1869	1870-1879	1880-1889	1890-1899	1900-1909	1910-1919	1920-1929	1930-1939	1940-1949	1950-1959
Expected																
Cell Chi-2																
0	4	11	25	30	29	26	38	36	40	40	33	34	13	14	9	389
	3.18852	8.76844	23.1166	27.1025	23.1168	26.3053	41.4508	38.2823	41.4508	38.2823	35.0738	36.668	22.3197	15.1455	7.97131	0.79713
	0.205	0.5679	0.1534	0.3098	1.4773	0.0035	0.2872	0.1398	0.0508	0.0789	0.1226	0.1941	0.4977	0.0866	0.1338	0.0516
5	0	0	4	4	0	7	14	12	12	8	11	12	9	5	1	0
	0.81148	2.23156	5.8832	6.89754	5.8832	6.69467	10.5492	9.7377	10.5492	9.7377	8.95623	9.33197	5.68033	3.85451	2.02869	0.20287
	0.8115	2.2316	0.6028	1.2172	5.8832	0.0139	1.1288	0.5256	0.1995	0.3101	0.4818	0.7628	1.9401	0.3904	0.5216	0.2029
	4	11	29	34	29	33	52	48	52	48	44	46	28	19	10	488

Test #

Source: DF LogLikelihood Ratio = (0)

Model: 15 15.19226 0.0017

Error: 472 230.93043

C Total: 487 246.12469

Total Count: 488

Test: ChiSquare ProbChiSq

LogLikelihood Ratio: 30.389 0.0106

Inversion: 21.544 0.1203

Warning: 20% of cells have expected count less than 5, Chi-squares suspect

Figure D.15 : Table de contingence de la mortalité pré-reproductive masculine par tranches de 10 ans.

mort-14f By cl.ansées		cl.ansées															
		1800-1809	1810-1819	1820-1829	1830-1839	1840-1849	1850-1859	1860-1869	1870-1879	1880-1889	1890-1899	1900-1909	1910-1919	1920-1929	1930-1939	1940-1949	1950-1959
Count	Expected	6	19	25	26	29	33	33	33	40	40	47	23	19	16	6	396
	Cell Chi^2	4.84899	16.1633	25.8612	24.2449	26.6694	29.0529	37.9837	32.3265	39.6	40.4082	41.2163	29.902	24.2449	15.3551	6.46531	1.61633
		0.2732	0.4979	0.0287	0.1271	0.2037	0.3244	0.6539	0.0140	0.0040	0.0041	0.8116	1.5931	1.1346	0.0271	0.0335	0.2350
5		0	1	7	4	4	3	14	7	9	10	4	14	11	3	2	1
		1.15102	3.83673	6.13878	5.7551	6.33061	6.90612	9.01633	7.67347	9.4	9.59184	9.78367	7.09796	5.7551	3.6449	1.53469	0.38367
		1.1510	2.0974	0.1208	0.5352	0.8980	2.2093	2.7547	0.0591	0.0170	0.0174	3.4191	6.7115	4.7799	0.1141	0.1411	0.9901
		6	20	32	30	33	36	47	40	49	50	51	37	30	19	8	2

Tests	
Source	DF -LogLikelihood Rsquare (0)
Model	15 16.46186 0.0687
Error	474 223.08705
C Total	489 239.54890
Total Count	490
Test	ChiSquare ProbChiSq
Likelihood Ratio	32.984 0.008
F Pearson	32.142 0.0062

Warning: 20% of cells have expected count less than 5, Chi-squares suspect

Figure D.16 : Table de contingence de la mortalité pré-reproductive féminine par tranches de 10 ans.

## **APPENDICE E**

### **Tests de symétrie des variables génétiques**

Tableau E.1 : Test de symétrie entre le coefficient de parenté total et le coefficient de parenté proche.

		Value	Asymp. Std. Error	Approx. T	Approx. Sig.
Ordinal	b Kendall's	.607	.018	23.258	.000
Ordinal	tau-b				
	Kendall's	.437	.019	23.258	.000
	tau-c				
	Gamma	.895	.015	23.258	.000
N of Valid Cases		1009			

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

Tableau E.2 : Test de symétrie entre le coefficient de parenté total et le coefficient de parenté éloigné.

		Value	Asymp. Std. Error	Approx. T	Approx. Sig.
Ordinal	b Kendall's	.549	.028	18.668	.000
Ordinal	tau-b				
	Kendall's	.426	.023	18.668	.000
	tau-c				
	Gamma	.690	.031	18.668	.000
N of Valid Cases		1009			

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

Tableau E.3 : Test de symétrie entre le coefficient de parenté éloigné et le coefficient de parenté proche.

		Value	Asymp. Std. Error	Approx. T	Approx. Sig.
Ordinal	b Kendall's	-.035	.028	-1.250	.211
Ordinal	tau-b				
	Kendall's	-.024	.019	-1.250	.211
	tau-c				
	Gamma	-.063	.050	-1.250	.211
N of Valid Cases		1009			

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

Tableau E.4 : Test de symétrie entre le coefficient de consanguinité maternel total et le coefficient de consanguinité maternel proche.

		Value	Asymp. Std. Error	Approx. T	Approx. Sig.
Ordinal	b Kendall's	.452	.022	14.534	.000
Ordinal	tau-b				
	Kendall's	.265	.018	14.534	.000
	tau-c				
	Gamma	.812	.022	14.534	.000
N of Valid Cases		1009			

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

Tableau E.5 : Test de symétrie entre le coefficient de consanguinité maternel total et le coefficient de consanguinité maternel éloigné.

		Value	Asymp. Std. Error	Approx. T	Approx. Sig.
Ordinal	b Kendall's	.765	.021	34.456	.000
Ordinal	tau-b				
	Kendall's	.580	.017	34.456	.000
	tau-c				
	Gamma	.865	.017	34.456	.000
N of Valid Cases		1009			

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

Tableau E.6 : Test de symétrie entre le coefficient de consanguinité maternel éloigné et le coefficient de consanguinité maternel proche.

		Value	Asymp. Std. Error	Approx. T	Approx. Sig.
Ordinal	b Kendall's	-.009	.028	-.321	.748
Ordinal	tau-b				
	Kendall's	-.005	.016	-.321	.748
	tau-c				
	Gamma	-.020	.063	-.321	.748
N of Valid Cases		1009			

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

Tableau E.7 : Test de symétrie entre le coefficient de consanguinité paternel total et le coefficient de consanguinité paternel proche.

		Value	Asymp. Std. Error	Approx. T	Approx. Sig.
Ordinal	b Kendall's	.459	.021	14.850	.000
Ordinal	tau-b				
	Kendall's	.255	.017	14.850	.000
	tau-c				
	Gamma	.837	.020	14.850	.000
N of Valid Cases		1009			

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

Tableau E.8 : Test de symétrie entre le coefficient de consanguinité paternel total et le coefficient de consanguinité paternel éloigné.

		Value	Asymp. Std. Error	Approx. T	Approx. Sig.
Ordinal	b Kendall's	.750	.023	32.171	.000
Ordinal	tau-b				
	Kendall's	.574	.018	32.171	.000
	tau-c				
	Gamma	.836	.020	32.171	.000
N of Valid Cases		1009			

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

Tableau E.9 : Test de symétrie entre le coefficient de consanguinité paternel éloigné et le coefficient de consanguinité paternel proche.

		Value	Asymp. Std. Error	Approx. T	Approx. Sig.
Ordinal	b Kendall's	-.016	.027	-.602	.547
Ordinal	tau-b				
	Kendall's	-.009	.014	-.602	.547
	tau-c				
	Gamma	-.036	.060	-.602	.547
N of Valid Cases		1009			

a Not assuming the null hypothesis.

b Using the asymptotic standard error assuming the null hypothesis.

## **APPENDICE F**

**Modèles de régressions logistiques effectués avec la sélection ENTER**

## Modèles de régression logistique faits avec la sélection

### ENTER

#### Classes de mortalités touchant la progéniture totale

Tableau F.1: Modèles de régressions logistiques effectués par sélection ENTER touchant la progéniture totale et soumettant les variables génétiques totales.

variables dans le modèle	mortalité (seuil de signification des variables) (rapports de cotes des variables)				
	infantile	morti-natalité	néonatale précoce	1-4 ans	pré-reproductive
nbenf	0,1061	<b>0,0003</b>	<b>0,0210</b>	0,1873	0,6216
	0,9431	<b>0,8174</b>	<b>0,8238</b>	1,0619	1,0185
difconj	<b>0,0414</b>	0,5307	0,9443	0,6060	0,6576
	<b>0,9482</b>	0,9744	1,0054	0,9851	0,9897
rangnai	0,2997	0,1804	0,8259	0,6456	0,3494
	1,0613	1,1255	0,9658	0,9667	1,0565
year	<b>0,0000</b>	<b>0,0005</b>	<b>0,0063</b>	0,0504	<b>0,0055</b>
	<b>1,0183</b>	<b>1,0160</b>	<b>1,0235</b>	1,0064	<b>1,0073</b>
agemn	0,7751	0,8384	0,9778	0,9128	0,9304
	1,0075	1,0077	1,0017	0,9963	0,9976
interidi	<b>0,0086</b>	0,0660	0,5524	0,3897	0,6784
	<b>0,9991</b>	0,9990	0,9994	1,0003	0,9999
xcoefp	0,1527	0,6747	0,1640	0,1837	0,2351
	1,1048	1,0475	1,3192	0,8927	0,9213
xcngh	0,1597	0,6004	0,2505	0,9235	0,9012
	0,9115	0,9459	0,8002	0,9923	0,9920
xcngf	<b>0,0130</b>	0,3965	<b>0,0244</b>	0,0620	0,0582
	<b>0,8525</b>	0,9159	<b>0,6198</b>	1,1526	1,1245
constante	<b>0,0017</b>	<b>0,0004</b>	<b>0,0060</b>	<b>0,0167</b>	<b>0,0017</b>

Tableau F.2: Modèles de régressions logistiques effectués par sélection ENTER touchant la progéniture totale et soumettant les variables génétiques proches.

variables dans le modèle	mortalité (seuil de signification des variables) (rapports de cotes des variables)				
	infantile	morti-natalité	néonatale précoce	1-4 ans	pré-reproductive
nbenf	<b>0,0070</b> <b>0,9123</b>	<b>0,0002</b> <b>0,8136</b>	<b>0,0315</b> <b>0,8318</b>	0,1653 1,0651	0,6354 1,0177
difconj	<b>0,0339</b> <b>0,9503</b>	0,5161 0,9736	0,9790 0,9980	0,5824 0,9837	0,7373 0,9920
rangnai	0,0993 1,0931	0,1873 1,1234	0,6986 0,9397	0,6221 0,9644	0,3652 1,0545
year	<b>0,0000</b> <b>1,0154</b>	<b>0,0033</b> <b>1,0149</b>	<b>0,0120</b> <b>1,0246</b>	<b>0,0387</b> <b>1,0074</b>	<b>0,0054</b> <b>1,0081</b>
agemn	0,9706 1,0009	0,7534 1,0118	0,7878 1,0172	0,9253 0,9968	0,9408 0,9980
interidi	<b>0,0005</b> <b>0,9989</b>	0,0543 0,9990	0,4874 0,9993	0,3596 1,0004	0,7303 0,9999
xapp44	0,7400 0,9731	0,6753 0,9580	0,2831 1,1897	<b>0,0284</b> <b>0,8262</b>	<b>0,0103</b> <b>0,8372</b>
xcngh55	0,6176 0,9577	0,3351 1,1169	0,6647 0,9094	0,4114 0,9181	0,7795 1,0224
xcngf55	0,9839 1,0018	0,8615 0,9795	0,1565 0,6788	<b>0,0111</b> <b>1,2371</b>	<b>0,0487</b> <b>1,1522</b>
constante	<b>0,0000</b>	<b>0,0023</b>	<b>0,0095</b>	<b>0,0138</b>	<b>0,0018</b>

Tableau F.3: Modèles de régressions logistiques effectués par sélection ENTER touchant la progéniture totale et soumettant les variables génétiques éloignées.

variables dans le modèle	mortalité (seuil de signification des variables) (rapports de cotes des variables)				
	infantile	morti-natalité	néonatale précoce	1-4 ans	pré-reproductive
nbenf	<b>0,0078</b> <b>0,9143</b>	<b>0,0003</b> <b>0,8174</b>	<b>0,0248</b> <b>0,8284</b>	0,2357 1,0547	0,7357 1,0125
difconj	<b>0,0363</b> <b>0,9510</b>	0,5674 0,9763	0,7436 0,9741	0,7833 0,9919	0,9265 0,9978
rangnai	0,1166 1,0880	0,1925 1,1213	0,6929 0,9401	0,7396 0,9759	0,2787 1,0654
year	<b>0,0000</b> <b>1,0137</b>	<b>0,0004</b> <b>1,0157</b>	<b>0,0086</b> <b>1,0230</b>	<b>0,0497</b> <b>1,0062</b>	<b>0,0036</b> <b>1,0073</b>
agemn	0,8619 1,0042	0,8024 1,0093	0,7805 1,0171	0,7846 0,9908	0,7815 0,9926
interidi	<b>0,0006</b> <b>0,9989</b>	0,0767 0,9990	0,4518 0,9993	0,3634 1,0004	0,7522 0,9999
xcél1	0,1795 1,1104	0,4039 1,1219	0,8804 0,9636	0,6350 0,9543	0,8420 1,0162
xché1	0,7444 1,0219	0,6038 0,9413	0,7050 0,9226	0,4460 1,0666	0,8574 1,0125
xcfél1	0,2327 0,9246	0,2385 0,8713	0,0737 0,6569	0,7699 0,9754	0,9829 1,0015
constante	<b>0,0000</b>	<b>0,0004</b>	<b>0,0084</b>	<b>0,0189</b>	<b>0,0013</b>

Classes de mortalités touchant la progéniture masculine

Tableau F.4: Modèles de régressions logistiques effectués par sélection ENTER touchant la progéniture masculine et soumettant les variables génétiques totales.

variables dans le modèle	mortalité (seuil de signification des variables) (rapports de cotes des variables)				
	infantile	morti- natalité	néonatale précoce	1-4 ans	pré- reproductive
nbenf	0,1782 0,9285	0,0668 0,8126	0,0855 0,7277	0,0877 1,1434	0,3313 1,0632
difconj	0,1247 0,9469	0,3104 0,9084	0,3274 0,8377	0,8598 0,9924	0,5416 1,0202
rangnai	0,3831 1,0745	0,5253 0,8962	0,5280 1,2055	0,2200 0,8707	0,9646 1,0040
year	<b>0,0000</b> <b>1,0164</b>	<b>0,0000</b> <b>1,0834</b>	0,5673 1,0077	0,1210 1,0077	0,0697 1,0072
agemn	0,9111 0,9958	0,2978 1,0868	0,5715 0,9372	0,2963 1,0567	0,2979 1,0443
interidi	0,1345 0,9993	0,9381 0,9999	0,6734 1,0006	0,7220 1,0002	0,2112 0,9994
xcoefp	0,7996 0,9760	0,1277 1,3783	0,9099 0,9593	<b>0,0445</b> <b>0,7800</b>	0,3228 0,9061
xcngh	0,3162 0,9101	0,1260 0,7400	0,4555 0,7468	0,8776 1,0183	0,7100 1,0359
xcngf	0,1046 0,8618	<b>0,0214</b> <b>0,6147</b>	0,0824 0,3912	0,3436 1,1165	0,3994 1,0823
constante	<b>0,0000</b>	<b>0,0000</b>	0,6290	<b>0,0419</b>	<b>0,0240</b>

Tableau F.5: Modèles de régressions logistiques effectués par sélection ENTER touchant la progéniture masculine et soumettant les variables génétiques proches.

variables dans le modèle	mortalité (seuil de signification des variables) (rapports de cotes des variables)				
	infantile	morti-natalité	néonatale précoce	1-4 ans	pré-reproductive
nbenf	0,2454 0,9378	0,0815 0,8228	0,0847 0,7140	0,0549 1,1690	0,2857 1,0703
difconj	0,0773 0,9397	0,3562 0,9190	0,2266 0,8161	0,7620 0,9863	0,5536 1,0200
rangnai	0,3763 1,0762	0,5397 0,9026	0,4518 1,2583	0,1564 0,8510	0,8933 0,9880
year	<b>0,0002</b> <b>1,0160</b>	<b>0,0000</b> <b>1,0881</b>	0,6955 1,0057	<b>0,0410</b> <b>1,0110</b>	<b>0,0142</b> <b>1,0106</b>
agemn	0,9417 0,9972	0,3046 1,0812	0,5642 0,9349	0,2469 1,0641	0,2520 1,0488
interidi	0,1193 0,9993	0,9588 1,0000	0,7028 1,0006	0,6151 1,0003	0,2620 0,9995
xapp44	0,9566 0,9952	0,2187 1,2345	0,7257 0,8795	<b>0,0005</b> <b>0,5914</b>	<b>0,0086</b> <b>0,7616</b>
xcngh55	0,3841 0,9111	0,4016 0,8576	0,8685 0,9218	0,4834 0,8982	0,5277 0,9280
xcngf55	0,6340 0,9499	0,0563 0,6299	0,8403 0,0253	<b>0,0144</b> <b>1,3724</b>	0,1982 1,1533
constante	<b>0,0001</b>	<b>0,0000</b>	0,8410	<b>0,0114</b>	<b>0,0044</b>

Tableau F.6: Modèles de régressions logistiques effectués par sélection ENTER touchant la progéniture masculine et soumettant les variables génétiques éloignées.

variables dans le modèle	mortalité (seuil de signification des variables) (rapports de cotes des variables)				
	infantile	morti- natalité	néonatale précoce	1-4 ans	pré- reproductive
nbenf	0,2305 0,9358	0,0873 0,8227	0,0731 0,7138	0,0821 1,1475	0,3028 1,0677
difconj	0,1420 0,9487	0,5159 0,9408	0,2383 0,7972	0,7673 1,0128	0,2821 1,0366
rangnai	0,3362 1,0828	0,4550 0,8823	0,4796 1,2389	0,2747 0,8835	0,8751 1,0143
year	<b>0,0002</b> <b>1,0140</b>	<b>0,0000</b> <b>1,0788</b>	0,8913 1,0018	0,1474 1,0069	0,0692 1,0069
agemn	0,9037 0,9954	0,2538 1,0940	0,5784 0,9377	0,3883 1,0467	0,3818 1,0371
interidi	0,1536 0,9994	0,9619 1,0000	0,6724 1,0006	0,5893 1,0003	0,2952 0,9995
xcél1	0,7108 1,0436	0,1990 1,5316	0,5124 0,7633	0,5414 1,0951	0,0906 1,2272
xché1	0,5782 1,0560	0,2644 0,7704	0,9440 0,9732	0,3148 1,1328	0,2014 1,1367
xcfél1	0,3349 0,9072	0,0873 0,6655	0,1549 0,4663	0,9849 0,9976	0,7100 1,0393
constante	<b>0,0002</b>	<b>0,0000</b>	0,9491	<b>0,0446</b>	<b>0,0185</b>

Classes de mortalités touchant la progéniture féminine

Tableau F.7: Modèles de régressions logistiques effectués par sélection ENTER touchant la progéniture féminine et soumettant les variables génétiques totales.

variables dans le modèle	mortalité (seuil de signification des variables) (rapports de cotes des variables)				
	infantile	morti-natalité	néonatale précoce	1-4 ans	pré-reproductive
nbenf	0,3219 0,9525	0,3494 0,8846	0,1554 0,8753	0,8288 1,0123	0,8546 0,9913
difconj	0,1908 0,9504	0,6457 1,0532	0,2485 1,1074	0,6034 0,9793	0,2040 0,9572
rangnai	0,5296 1,0525	0,9242 0,9783	0,4231 0,8504	0,6926 1,0393	0,2500 1,0959
year	<b>0,0000</b> <b>1,0204</b>	<b>0,0224</b> <b>1,0301</b>	<b>0,0043</b> <b>1,0354</b>	0,3254 1,0044	0,0623 1,0068
agemn	0,5630 1,0215	0,8154 1,0227	0,4505 1,0600	0,3855 0,9617	0,3432 0,9657
interidi	<b>0,0253</b> <b>0,9988</b>	0,4081 1,0008	0,3099 0,9985	0,4391 1,0004	0,5782 1,0003
xcoefp	<b>0,0270</b> <b>1,2578</b>	0,6193 1,1486	0,0874 1,5120	0,9875 1,0019	0,3936 0,9200
xcngh	0,3123 0,9070	0,7631 1,0835	0,4049 0,8265	0,7808 0,9695	0,7368 0,9696
xcngf	<b>0,0484</b> <b>0,8357</b>	0,0785 0,5855	0,1587 0,7037	0,0969 1,1862	0,0522 1,1803
constante	<b>0,0000</b>	<b>0,0000</b>	<b>0,0033</b>	0,2361	<b>0,0480</b>

Tableau F.8: Modèles de régressions logistiques effectués par sélection ENTER touchant la progéniture féminine et soumettant les variables génétiques proches.

variables dans le modèle	mortalité (seuil de signification des variables) (rapports de cotes des variables)				
	infantile	morti-natalité	néonatale précoce	1-4 ans	pré-reproductive
nbenf	0,3984 0,9599	0,3969 0,8922	0,1459 0,8733	0,8700 1,0092	0,7825 0,9871
difconj	0,0997 0,9383	0,8726 1,0175	0,2815 1,0939	0,6643 0,9825	0,2630 0,9617
rangnai	0,6329 1,0394	0,8157 0,9474	0,3889 0,8423	0,6791 1,0412	0,2641 1,0929
year	<b>0,0000</b> <b>1,0216</b>	<b>0,0287</b> <b>1,0301</b>	<b>0,0060</b> <b>1,0379</b>	0,5262 1,0031	0,2281 1,0048
agemn	0,5180 1,0243	0,6797 1,0418	0,4817 1,0569	0,3932 0,9623	0,3694 0,9676
interidi	<b>0,0195</b> <b>0,9987</b>	0,4462 1,0008	0,3319 0,9986	0,4286 1,0005	0,5643 1,0003
xapp44	0,3710 1,0841	0,2234 0,6961	0,1039 1,3712	0,7124 1,0414	0,2511 0,8959
xcngh55	0,4251 0,9142	0,2632 1,3283	0,5500 0,8515	0,6711 0,9408	0,2482 1,1346
xcngf55	0,1913 0,8710	0,9204 1,0292	0,2792 0,7382	0,2135 1,1529	0,0757 1,1894
constante	<b>0,0000</b>	<b>0,0188</b>	<b>0,0046</b>	0,4197	0,1797

Tableau F.9: Modèles de régressions logistiques effectués par sélection ENTER touchant la progéniture féminine et soumettant les variables génétiques éloignées.

variables dans le modèle	mortalité (seuil de signification des variables) (rapports de cotes des variables)				
	infantile	morti-natalité	néonatale précoce	1-4 ans	pré-reproductive
nbenf	0,2935 0,9504	0,2410 0,8538	0,1487 0,8744	0,8482 1,0108	0,8111 0,9887
difconj	0,1339 0,9425	0,5887 1,0700	0,4913 1,0656	0,5629 0,9763	0,2692 0,9626
rangnai	0,6157 1,0412	0,8837 1,0356	0,3684 0,8391	0,6949 1,0390	0,2320 1,0997
year	<b>0,0000</b> <b>1,0203</b>	<b>0,0151</b> <b>1,0305</b>	<b>0,0038</b> <b>1,0366</b>	0,2515 1,0049	<b>0,0370</b> <b>1,0072</b>
agemn	0,4857 1,0257	0,8130 1,0228	0,3728 1,0690	0,3668 0,9602	0,3253 0,9646
interidi	<b>0,0300</b> <b>0,9988</b>	0,3995 1,0009	0,2345 0,9983	0,5565 1,0003	0,7149 1,0002
xcél1	<b>0,0416</b> <b>1,2919</b>	0,1699 1,7817	0,9290 1,0295	0,2621 0,8599	0,2152 0,8708
xché1	0,6679 0,9559	0,4522 0,7835	0,8625 0,9555	0,7975 1,0307	0,4896 0,9342
xcfél1	0,0991 0,8449	<b>0,0224</b> <b>0,2923</b>	0,2416 0,7187	0,7205 0,9596	0,7853 0,9747
constante	<b>0,0000</b>	<b>0,0123</b>	<b>0,0033</b>	0,2205	<b>0,0392</b>

## **APPENDICE G**

**Évaluation de l'ajustement et des pourcentages explicatifs  
des modèles de régressions logistiques**

Tableau G.1 : Évaluation de l'ajustement et des pourcentages explicatifs des modèles de régressions logistiques soumettant les variables génétiques totales.

types de consanguinité et de parenté	type de mortalité	genre des individus visés (et numéro du modèle)	-2Log likelihood	Chi carré (Model) et signification du modèle	indice de Cox et S. / indice de Nagelkerke
<b>consanguinité et parenté totales</b>	infantile	totale (1)	1053,592	69,519 0,000	0,067 0,099
		masculin (5)	504,156	21,597 0,000	0,043 0,066
		féminin (7)	459,562	47,016 0,000	0,091 0,142
	mortalité (0 jour)	totale (10)	456,972	25,250 0,000	0,025 0,065
		masculin (13)	130,726	60,663 0,000	0,117 0,360
		féminin (16)	91,499	6,132 0,013	0,012 0,069
	néo-natale précoce (1-6 jours)	totale (19)	196,949	18,637 0,000	0,018 0,107
		masculin (22)	modèle NS	---	---
		féminin (25)	105,757	14,261 0,001	0,029 0,132

(suite)

types de consanguinité et de parenté	types de mortalité	genre des individus visés (et numéro du modèle)	-2Log likelihood	Chi carré (Model) et signification du modèle	indice de Cox et S. / indice de Nagelkerke
<b>consanguinité et parenté totales</b>	1-4 ans	totale (28)	722,692	5,399 0,020	0,005 0,010
		masculin (31)	335,844	7,838 0,020	0,005 0,010
		féminin (34)	modèle NS	—	—
	pré-reproductive	totale (37)	988,352	19,238 0,000	0,019 0,030
		masculin (40)	470,583	13,341 0,001	0,027 0,043
		féminin (43)	503,583	5,605 0,018	0,011 0,018

Tableau G.2 : Évaluation de l'ajustement et des pourcentages explicatifs des modèles de régressions logistiques soumettant les variables génétiques proches.

types de consanguinité et de parenté	types de mortalité	genre des individus visés (et numéro du modèle)	-2Log likelihood	Chi carré (Model) et signification du modèle	indice de Cox et S. / indice de Nagelkerke	
<b>consanguinité et parenté proches</b>	infantile	totale (2)	identique au (1)	---	---	
		masculin (6)	identique au (5)			
		féminin (8)	463,842	42,736 0,000	0,084 0,130	
	mortalité (0 jour)	totale (11)	identique au (10)			
		masculin (14)	135,503	55,886 0,000	0,108 0,334	
		féminin (17)	identique au (16)			
	néo-natale précoce (1-6 jours)	totale (20)	174,376	14,211 0,001	0,014 0,082	
		masculin (23)	modèle NS	---	---	
		féminin (26)	identique au (25)			

(suite)

types de consanguinité et de parenté	types de mortalité	genre des individus visés (et numéro du modèle)	-2Log likelihood	Chi carré (Model) et signification du modèle	indice de Cox et S. / indice de Nagelkerke
<b>consanguinité et parenté proches</b>	1-4 ans	totale (29)	720,050	8,042 0,005	0,008 0,015
		masculin (32)	320,559	23,123 0,000	0,046 0,092
		féminin (35)	modèle NS	---	---
	pré-reproductive	totale (38)	986,467	21,123 0,000	0,021 0,033
		masculin (41)	466,015	17,909 0,000	0,036 0,057
		féminin (44)	identique au (43)	---	---

Tableau G.3 : Évaluation de l'ajustement et des pourcentages explicatifs des modèles de régressions logistiques soumettant les variables génétiques éloignées.

types de consanguinité et de parenté	types de mortalité	genre des individus visés (et numéro du modèle)	-2Log likelihood	Chi carré (Model) et signification du modèle	indice de Cox et S. / indice de Nagelkerke
<b>consanguinité et parenté éloignées</b>	infantile	totale (3)	identique au (1)	---	---
		masculin (4)	identique au (5)	---	---
		féminin (9)	469,246	42,806 0,000	0,084 0,130
	mortalité (0 jour)	totale (12)	identique au (10)	---	---
		masculin (15)	identique au (14)	---	---
		féminin (18)	82,641	14,990 0,001	0,030 0,167
	néo-natale précoce (1-6 jours)	totale (21)	identique au (20)	---	---
		masculin (24)	modèle NS	---	---
		féminin (27)	identique au (25)	---	---

(suite)

types de consanguinité et de parenté	types de mortalité	genre des individus visés (et numéro du modèle)	-2Log likelihood	Chi carré (modèle) et signification du modèle	indice de Cox et S. / indice de Nagelkerke
<b>consanguinité et parenté éloignées</b>	1-4 ans	totale (30)	identique au (28)	---	---
		masculin (33)	339,744	3,939 0,047	0,008 0,016
		féminin (36)	modèle NS	---	---
	pré-reproductive	totale (39)	identique au (37)	---	---
		masculin (42)	identique au (40)	---	---
		féminin (45)	identique au (43)	---	---

## **APPENDICE H**

**Évaluation de la calibration des modèles de régressions logistiques**

Tableau H.1 : Évaluation de la calibration des modèles de régressions logistiques soumettant les variables génétiques totales (test de Hosmer et Lemeshow).

type de consanguinité et de parenté	type de mortalité	genre des individus visés (et numéro du modèle)	Valeur des tests de Hosmer et Lemeshow (calibration) et signification du test (seuil de 0,05)
<b>consanguinité et parenté totales</b>	infantile	totale (1)	23,521 0,003
		masculin (5)	50,639 0,000
		féminin (7)	9,086 0,335 NS
	mortalité (0 jour)	totale (10)	11,664 0,167 NS
		masculin (13)	4,763 0,783 NS
		féminin (16)	15,162 0,056 NS

(suite)

type de consanguinité et de parenté	type de mortalité	genre des individus visés (et numéro du modèle)	Valeur des tests de Hosmer et Lemeshow (calibration) et signification du test (seuil de 0,05)	
<b>consanguinité et parenté</b>	néonatale précoce (1-6 jours)	totale (19)	11,024 0,200 NS	
		masculin (22) modèle NS	---	
		féminin (25)	7,584 0,485 NS	
	<b>totales</b>	1-4 ans	totale (28)	11,056 0,199 NS
			masculin (31)	7,912 0,442 NS
			féminin (34) modèle NS	---

(suite)

type de consanguinité et de parenté	type de mortalité	genre des individus visés (et numéro du modèle)	Valeur des tests de Hosmer et Lemeshow (calibration) et signification du test (seuil de 0,05)
<b>consanguinité et parenté</b>	pré-reproductive	totale (37)	8,227 0,412 NS
		masculin (40)	9,916 0,217 NS
		féminin (43)	10,298 0,245 NS
<b>totales</b>			

Tableau H.2 : Évaluation de la calibration des modèles de régressions logistiques soumettant les variables génétiques proches (test de Hosmer et Lemeshow).

type de consanguinité et de parenté	type de mortalité	genre des individus visés (et numéro du modèle)	Valeur des tests de Hosmer et Lemeshow (calibration) et signification du test (seuil de 0,05)
<b>consanguinité et parenté proches</b>	infantile	totale (2) identique au (1)	
		masculin (6) identique au (5)	
		féminin (8)	15,523 0,05
	mortalité (0 jour)	totale (11) identique au (10)	
		masculin (14)	10,259 0,247
		féminin (17) identique au (16)	

(suite)

type de consanguinité et de parenté	type de mortalité	genre des individus visés (et numéro du modèle)	Valeur des tests de Hosmer et Lemeshow (calibration) et signification du test (seuil de 0,05)
<b>consanguinité et parenté proches</b>	néonatale précoce (1-6 jours)	totale (20)	8,506 0,386 NS
		masculin (23) modèle NS	---
		féminin (26) identique au (25)	
	1-4 ans	totale (29)	8,168 0,004
		masculin (32)	11,366 0,182 NS
		féminin (35) modèle NS	---

(suite)

type de consanguinité et de parenté	type de mortalité	genre des individus visés (et numéro du modèle)	Valeur des tests de Hosmer et Lemeshow (calibration) et signification du test (seuil de 0,05)
<b>consanguinité et parenté proches</b>	pré-reproductive	totale (38)	8,593 0,378 NS
		masculin (41)	7,102 0,526 NS
		féminin (44) identique au (43)	

Tableau H.3 : Évaluation de la calibration des modèles de régressions logistiques soumettant les variables génétiques éloignées (test de Hosmer et Lemeshow).

type de consanguinité et de parenté	type de mortalité	genre des individus visés (et numéro du modèle)	Valeur des tests de Hosmer et Lemeshow (calibration) et signification du test (seuil de 0,05)
<b>consanguinité et parenté éloignées</b>	infantile	totale (3) identique au (1)	
		masculin (4) identique au (5)	
		féminin (9)	8,645 0,373 NS
	mortalité (0 jour)	totale (12) identique au (10)	
		masculin (15) identique au (14)	
		féminin (18)	10,895 0,208 NS

(suite)

type de consanguinité et de parenté	type de mortalité	genre des individus visés (et numéro du modèle)	Valeur des tests de Hosmer et Lemeshow (calibration) et signification du test (seuil de 0,05).
<b>consanguinité et parenté éloignées</b>	néonatale précoce (1-6 jours)	totale (21) identique au (20)	
		masculin (24) modèle NS	---
		féminin (27) identique au (25)	
	1-4 ans	totale (30) identique au (28)	
		masculin (33)	13,013 0,111 NS
		féminin (36) modèle NS	---

(suite)

type de consanguinité et de parenté	type de mortalité	genre des individus visés (et numéro du modèle)	Valeur des tests de Hosmer et Lemeshow (calibration) et signification du test (seuil de 0,05)
<b>consanguinité et parenté éloignées</b>	pré-reproductive	totale (39) identique au (37)	
		masculin (42) identique au (40)	
		féminin (45) identique au (43)	

## **APPENDICE I**

### **Évaluation de la capacité de discrimination des modèles de régressions logistiques**

L'utilisation de la courbe ROC mesure en fait ce que l'on appelle la « statistique C ». Celle-ci représente la proportion de paires de cas ayant des destinées différentes (l'un meurt, l'autre pas) pour lesquelles le modèle a attribué un risque plus élevé à l'individu qui meurt (« 1 ») qu'à celui qui ne meurt pas (« 0 ») pour la période donnée. Cette affirmation n'exige toutefois pas que la personne « 1 » soit prédite comme morte mais seulement que son facteur de risque soit plus élevé que celui de l'autre membre de la paire. La valeur totale obtenue correspond à l'aire sous la courbe. Si l'aire sous la courbe pour un modèle donné est égal à 0,5, la prédiction ne vaudrait pas plus que de lancer une pièce de monnaie pour un pile ou face.

Les courbes ROC obtenues sont donc elles fort intéressantes. Malgré le problème de sensibilité dénoté par les tables de classification, on voit que la grande majorité des modèles présentent des aires sous la courbe supérieure à 0,6 avec un intervalle de confiance n'incluant pas les valeurs de 0,5 ou inférieures. Ce résultat indique donc que les modèles, même s'ils sont dans l'incapacité de discriminer correctement les individus qui meurent de ceux qui survivent pour une catégorie de mortalité donnée, peuvent donner une bonne indication de la tendance de l'exposition puisque la probabilité que cette prédiction soit vraie est plus grande que si la prédiction était faite au hasard.

Tableau I.1: Tables de classification des modèles de régression logistique utilisant la progéniture totale.

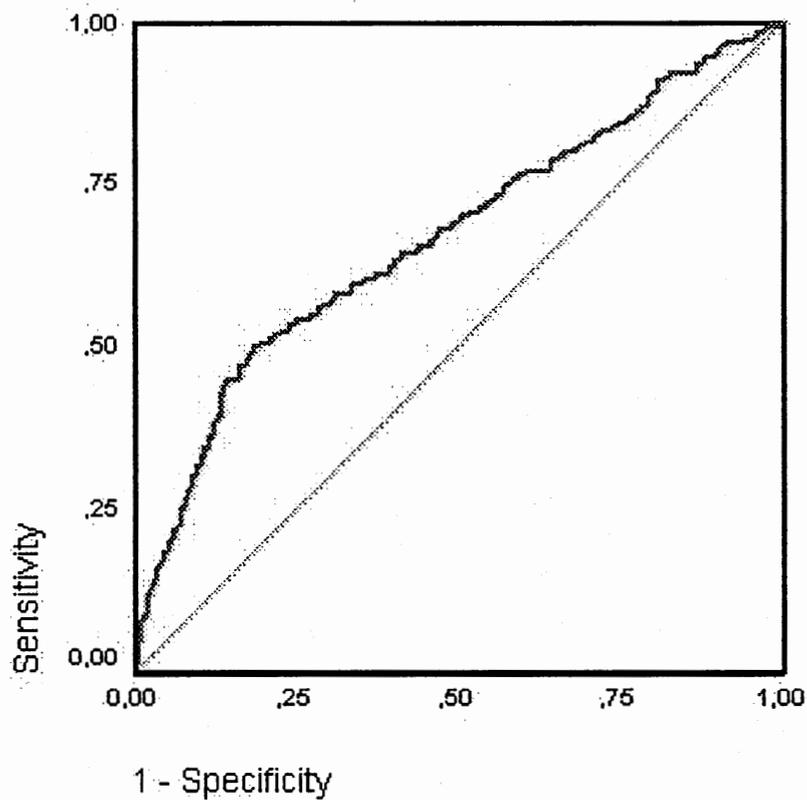
			Predit		
			MORTINFB		pourcentage correctement classé
Observé			,00	1,00	
	modèles 1, 2 et 3	MORTINF	,00	756	6
B		1,00	229	18	7,3
Overall Percentage				76,7	
a The cut value is ,500					
modèles 10 et 12	MORT0JB	,00	944	0	100,0
		1,00	65	0	,0
	Overall Percentage				93,6
a The cut value is ,500					
modèle 11	MORT0JB	,00	944	0	100,0
		1,00	65	0	,0
	Overall Percentage				93,6
a The cut value is ,500					
modèle 19	MORT06B	,00	990	0	100,0
		1,00	19	0	,0
	Overall Percentage				98,1
a The cut value is ,500					
modèles 20 et 21	MORT06B	,00	990	0	100,0
		1,00	19	0	,0
	Overall Percentage				98,1
a The cut value is ,500					
modèle 28 et 30	MORT4B	,00	891	0	100,0
		1,00	118	0	,0
	Overall Percentage				88,3
a The cut value is ,500					

(suite)

modèle 29	MORT4B	,00	891	0	100,0
		1,00	118	0	,0
	Overall Percentage				88,3
a The cut value is ,500					
modèles 37 et 39	MORT14B	,00	808	0	100,0
		1,00	201	0	,0
	Overall Percentage				80,1
a The cut value is ,500					
modèle 38	MORT14 B	,00	808	0	100,0
		1,00	201	0	,0
	Overall Percentage				80,1
a The cut value is ,500					

variable: MORTINFB	nombre d'individus
(1) nb d'individus mort pendant la période	247
(0) nb d'individus non-mort pendant la période	762
cas sans date de mort	1542

### ROC Curve



#### Area Under the Curve

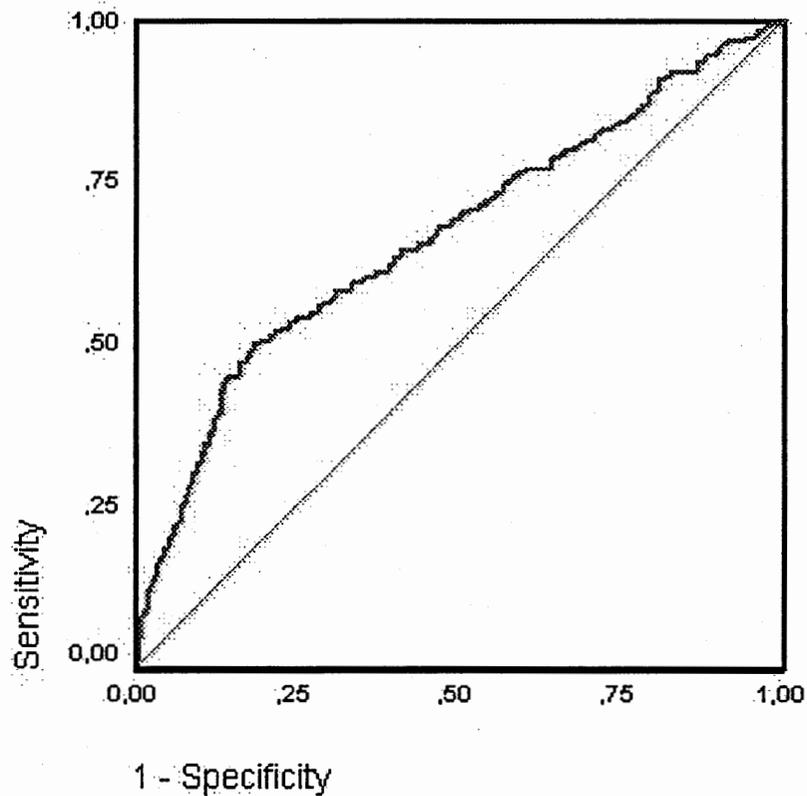
Test Result Variable(s): Predicted probability

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,669	,021	,000	,627	,710
a Under the nonparametric assumption				
b Null hypothesis: true area = 0.5				

Figure I.1 : « Roc Curve » pour les modèles 1,2 et 3.

MORTINFB	Valid N (listwise)
(1) nb d'individus mort pendant la période	247
(0) nb d'individus non-mort pendant la période	762
cas sans date de mort	1542

### ROC Curve



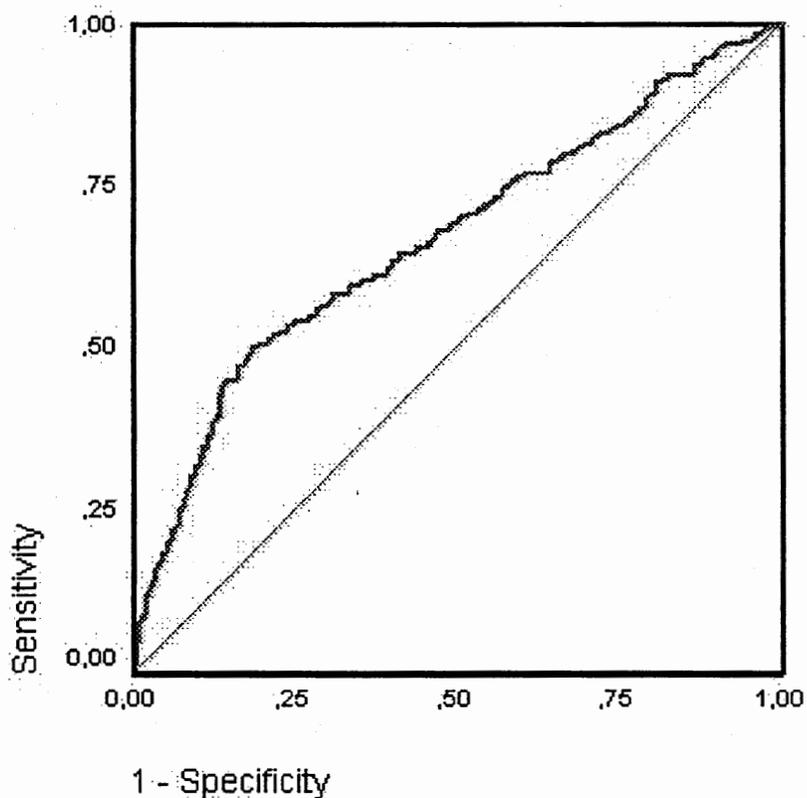
Area Under the Curve  
Test Result Variable(s): Predicted probability

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.632	.022	.000	.589	.676
The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.				
a Under the nonparametric assumption				
b Null hypothesis: true area = 0.5				

Figure I.2 : « Roc Curve » pour les modèles 10 et 12.

MORTINFB	Valid N (listwise)
(1) nb d'individus mort pendant la période	247
(0) nb d'individus non-mort pendant la période	762
cas sans date de mort	1542

### ROC Curve



#### Area Under the Curve

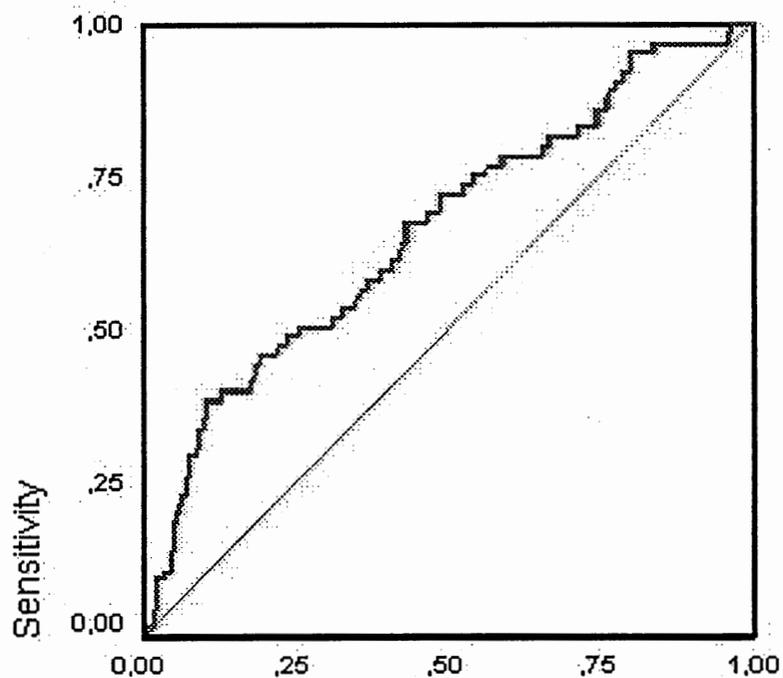
Test Result Variable(s): Predicted probability

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.632	.022	.000	.589	.676
The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.				
a Under the nonparametric assumption				
b Null hypothesis: true area = 0.5				

Figure I.3 : « Roc Curve » pour le modèle 11.

MORT0JB	Valid N (listwise)
(1) nb d'individus mort pendant la période	65
(0) nb d'individus non-mort pendant la période	944
cas sans date de mort	1542

### ROC Curve



1 - Specificity

Diagonal segments are produced by ties.

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,663	,037	,000	,592	,735

The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

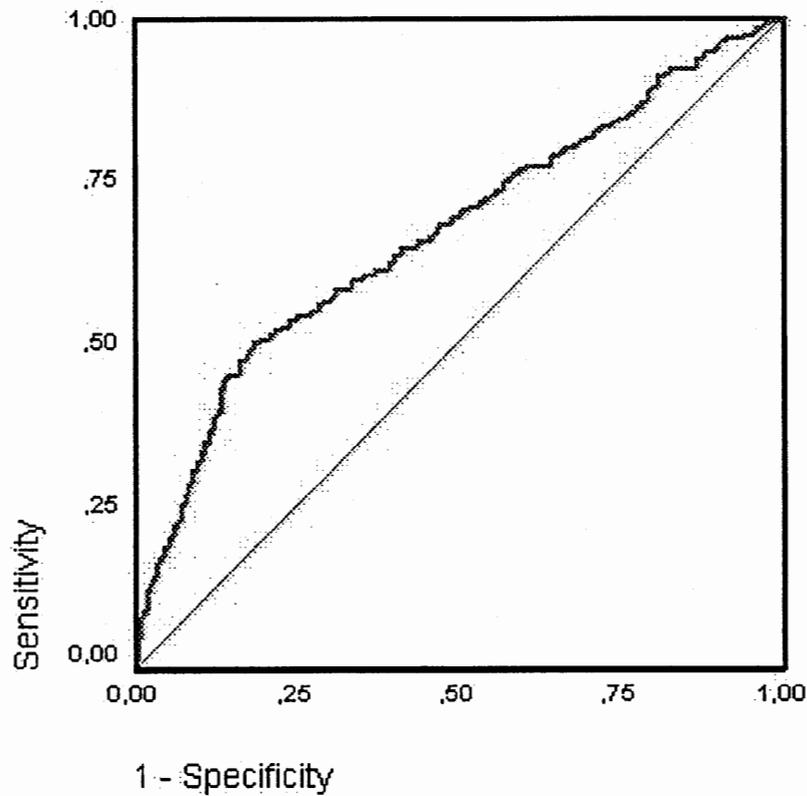
a Under the nonparametric assumption

b Null hypothesis: true area = 0.5

Figure I.4 : « Roc Curve » pour le modèle 19.

MORT0JB	Valid N (listwise)
(1) nb d'individus mort pendant la période	65
(0) nb d'individus non-mort pendant la période	944
cas sans date de mort	1542

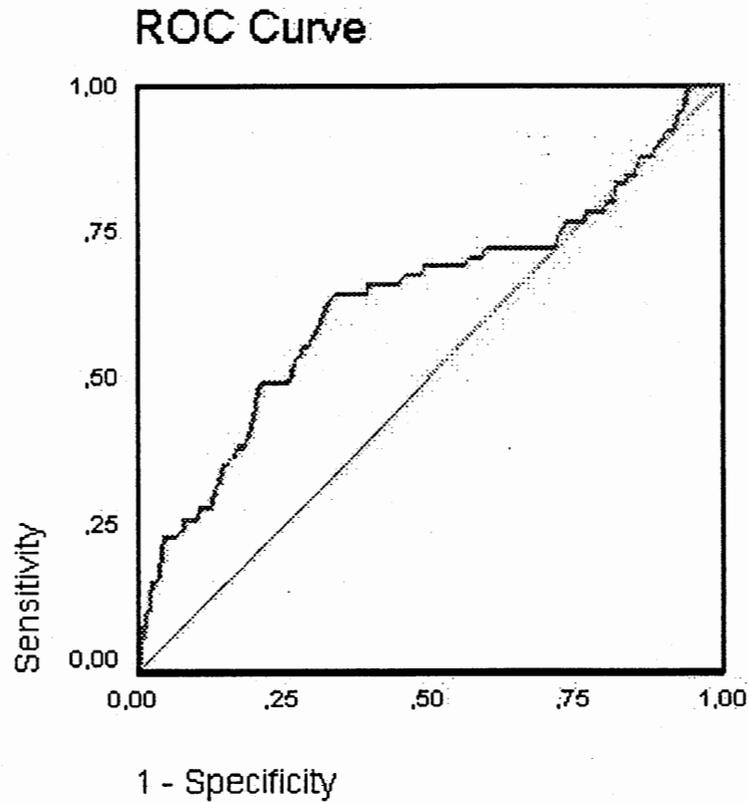
### ROC Curve



Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,682	,035	,000	,615	,750
The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.				
a Under the nonparametric assumption				
b Null hypothesis: true area = 0.5				

Figure I.5 : « Roc Curve » pour les modèles 20 et 21.

MORT0JB	Valid N (listwise)
(1) nb d'individus mort pendant la période	65
(0) nb d'individus non-mort pendant la période	944
cas sans date de mort	1542

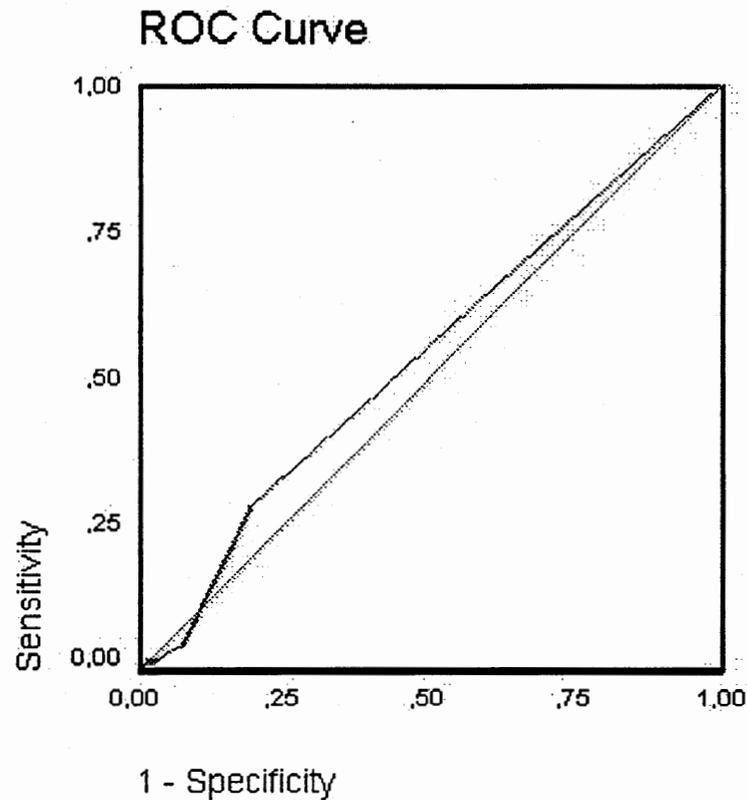


Diagonal segments are produced by ties.

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,635	,042	,000	,554	,717
The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.				
a Under the nonparametric assumption				
b Null hypothesis: true area = 0.5				

Figure I.6 : « Roc Curve » pour les modèles 28 et 30.

MORT0JB	Valid N (listwise)
(1) nb d'individus mort pendant la période	65
(0) nb d'individus non-mort pendant la période	944
cas sans date de mort	1542



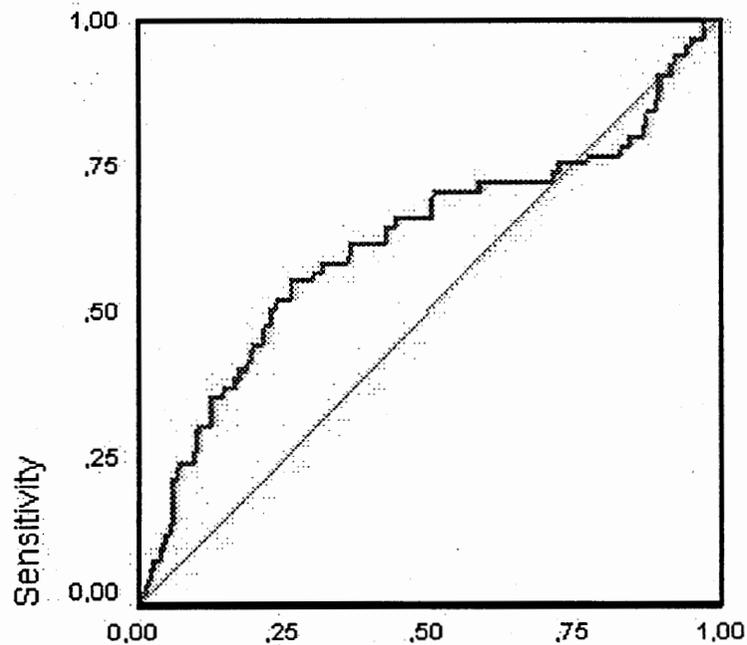
Diagonal segments are produced by ties.

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,538	,038	,303	,465	,612
The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.				
a Under the nonparametric assumption				
b Null hypothesis: true area = 0.5				

Figure I.7 : « Roc Curve » pour le modèle 29.

MORT0JB	Valid N (listwise)
(1) nb d'individus mort pendant la période	65
(0) nb d'individus non-mort pendant la période	944
cas sans date de mort	1542

### ROC Curve



1 - Specificity

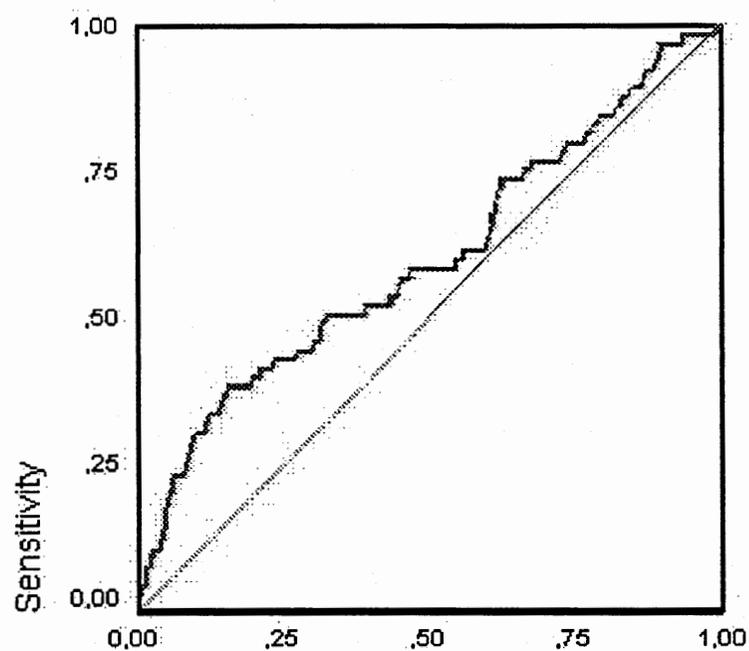
Diagonal segments are produced by ties.

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.621	.042	.001	.538	.704
The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.				
a Under the nonparametric assumption				
b Null hypothesis: true area = 0.5				

Figure I.8 : « Roc Curve » pour les modèles 37 et 39.

MORT0JB	Valid N (listwise)
(1) nb d'individus mort pendant la période	65
(0) nb d'individus non-mort pendant la période	944
cas sans date de mort	1542

### ROC Curve



1 - Specificity

Diagonal segments are produced by ties.

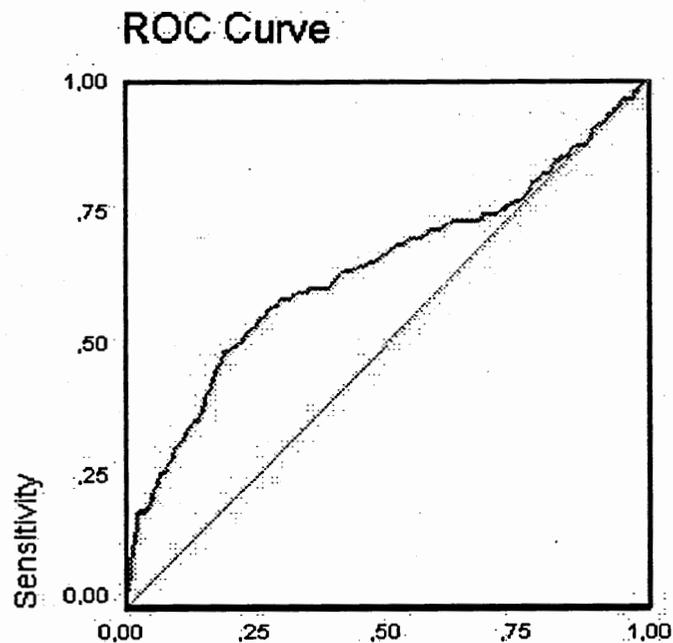
Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,601	,040	,006	,522	,680
The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.				
a Under the nonparametric assumption				
b Null hypothesis: true area = 0.5				

Figure I.9 : « Roc Curve » pour le modèle 38.

Tableau I.2: Tables de classification des modèles de régression logistique utilisant la progéniture masculine.

		Predicted						
		Selected Cases(a)			Unselected Cases(b,c)			
		MORTINFB		Percentage Correct	MORTINFB		Percentage Correct	
Observed	,00	1,00	,00		1,00			
modèles 4, 5 et 6	MORTINFB	,00	376	0	100,0	386	0	100,0
		1,00	112	0	,0	135	0	,0
	Overall Percentage				77,0			74,1
d The cut value is ,500								
modèle 13	MORT0JB	,00	462	2	99,6	473	7	98,5
		1,00	21	3	12,5	40	1	2,4
	Overall Percentage				95,3			91,0
d The cut value is ,500								
modèles 14 et 15	MORT0JB	,00	463	1	99,8	474	6	98,8
		1,00	22	2	8,3	41	0	,0
	Overall Percentage				95,3			91,0
d The cut value is ,500								
modèle 31	MORT4B	,00	433	0	100,0	458	0	100,0
		1,00	55	0	,0	63	0	,0
	Overall Percentage				88,7			87,9
d The cut value is ,500								
modèle 32	MORT4B	,00	433	0	100,0	456	2	99,6
		1,00	55	0	,0	63	0	,0
	Overall Percentage				88,7			87,5
d The cut value is ,500								
modèles 40 et 42	MORT14B	,00	392	0	100,0	416	0	100,0
		1,00	96	0	,0	105	0	,0
	Overall Percentage				80,3			79,8
d The cut value is ,500								
modèle 41	MORT14B	,00	392	0	100,0	416	0	100,0
		1,00	96	0	,0	105	0	,0
	Overall Percentage				80,3			79,8
d The cut value is ,500								

MORTINFB	Valid N (listwise)
(1) nb d'individus mort pendant la période	247
(0) nb d'individus non-mort pendant la période	762
cas sans date de mort	1542
Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.	
a The positive actual state is 1,00.	



1 - Specificity

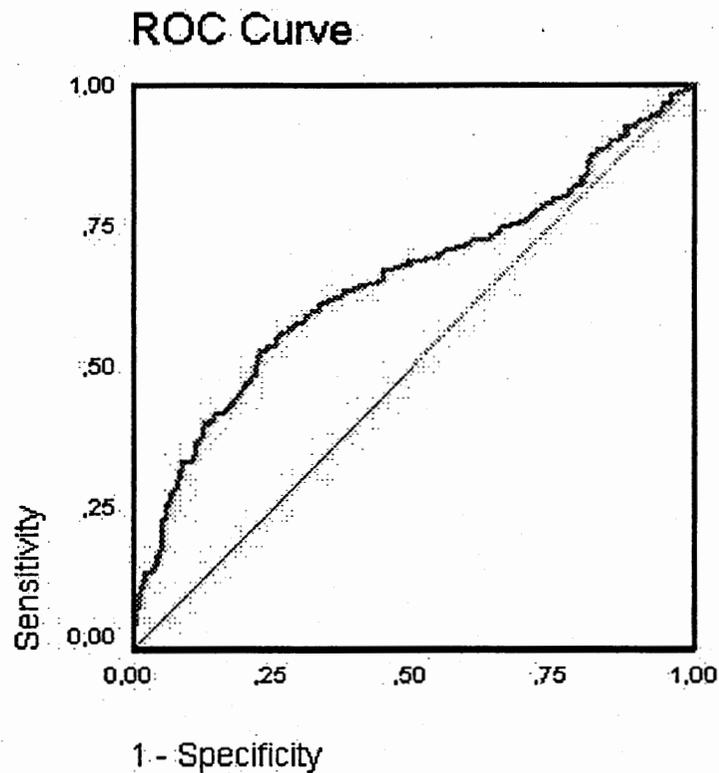
Diagonal segments are produced by ties.

**Area Under the Curve**  
Test Result Variable(s): Predicted probability

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,641	,023	,000	,597	,686
The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.				
a Under the nonparametric assumption				
b Null hypothesis: true area = 0.5				

Figure I.10 : « Roc Curve » pour les modèles 4, 5 et 6.

MORTINFB	Valid N (listwise)
(1) nb d'individus mort pendant la période	247
(0) nb d'individus non-mort pendant la période	762
cas sans date de mort	1542
Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.	
a The positive actual state is 1,00.	



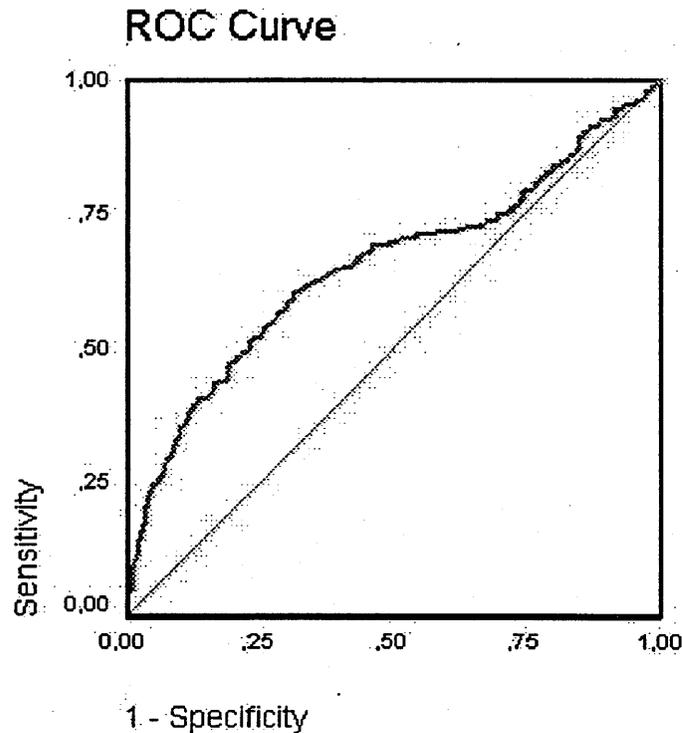
Diagonal segments are produced by ties.

**Area Under the Curve**  
Test Result Variable(s): Predicted probability

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,652	,022	,000	,609	,696
The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.				
a Under the nonparametric assumption				
b Null hypothesis: true area = 0.5				

Figure I.11: « Roc Curve » pour le modèle 13

MORTINFB	Valid N (listwise)
(1) nb d'individus mort pendant la période	247
(0) nb d'individus non-mort pendant la période	762
cas sans date de mort	1542
Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.	
a The positive actual state is 1,00.	



Diagonal segments are produced by ties.

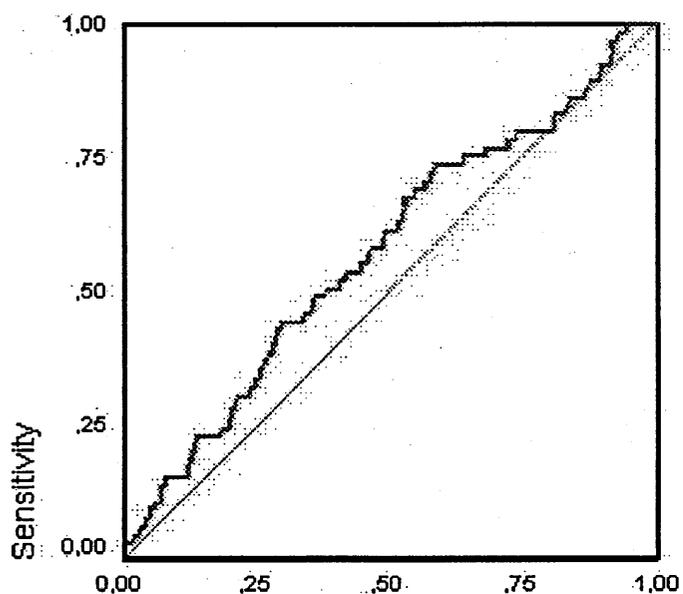
**Area Under the Curve**  
Test Result Variable(s): Predicted probability

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,652	,022	,000	,609	,696
The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.				
a Under the nonparametric assumption				
b Null hypothesis: true area = 0.5				

Figure I.12 : « Roc Curve » pour les modèles 14 et 15

MORT0JB	Valid N (listwise)
(1) nb d'individus mort pendant la période	65
(0) nb d'individus non-mort pendant la période	944
cas sans date de mort	1542
Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.	
a The positive actual state is 1,00.	

### ROC Curve



1 - Specificity

Diagonal segments are produced by ties.

#### Area Under the Curve

Test Result Variable(s): Predicted probability

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,574	,037	,046	,500	,647

The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

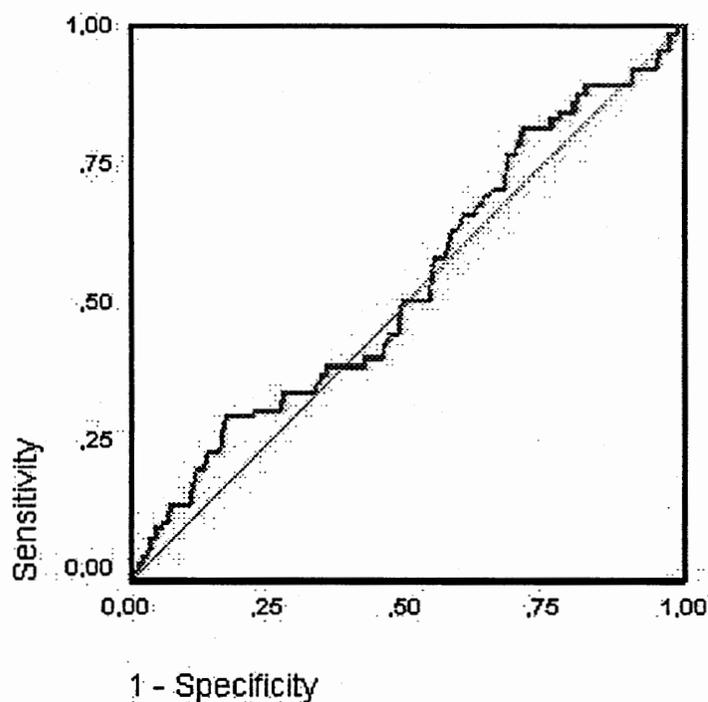
a Under the nonparametric assumption

b Null hypothesis: true area = 0.5

Figure I.13 : « Roc Curve » pour le modèle 31.

MORT0JB	Valid N (listwise)
(1) nb d'individus mort pendant la période	65
(0) nb d'individus non-mort pendant la période	944
cas sans date de mort	1542
Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.	
a The positive actual state is 1,00.	

### ROC Curve



Diagonal segments are produced by ties.

#### Area Under the Curve

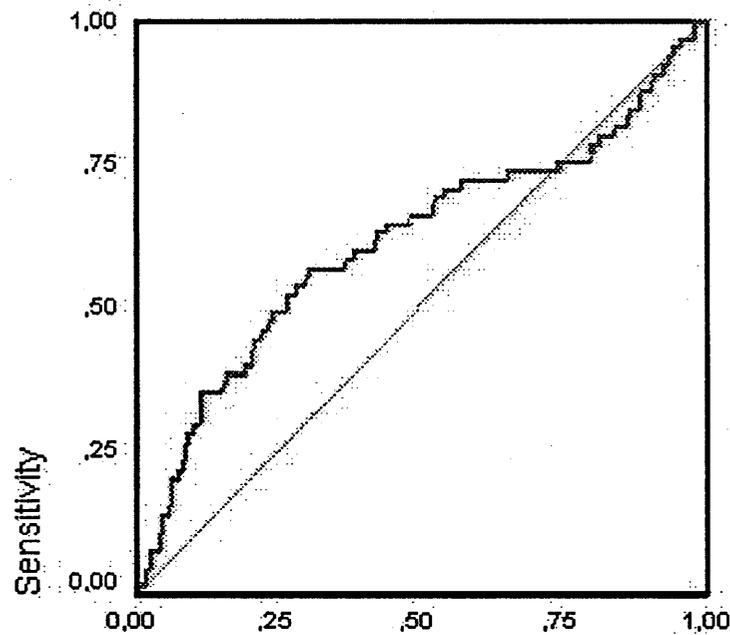
Test Result Variable(s): Predicted probability

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,534	,038	,360	,460	,608
The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.				
a Under the nonparametric assumption				
b Null hypothesis: true area = 0.5				

Figure I.14 : « Roc Curve » pour le modèle 32.

MORT0JB	Valid N (listwise)
(1) nb d'individus mort pendant la période	65
(0) nb d'individus non-mort pendant la période	944
cas sans date de mort	1542
Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.	
a The positive actual state is 1,00.	

### ROC Curve



1 - Specificity

Diagonal segments are produced by ties.

#### Area Under the Curve

Test Result Variable(s): Predicted probability

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.613	.042	.002	.531	.696

The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

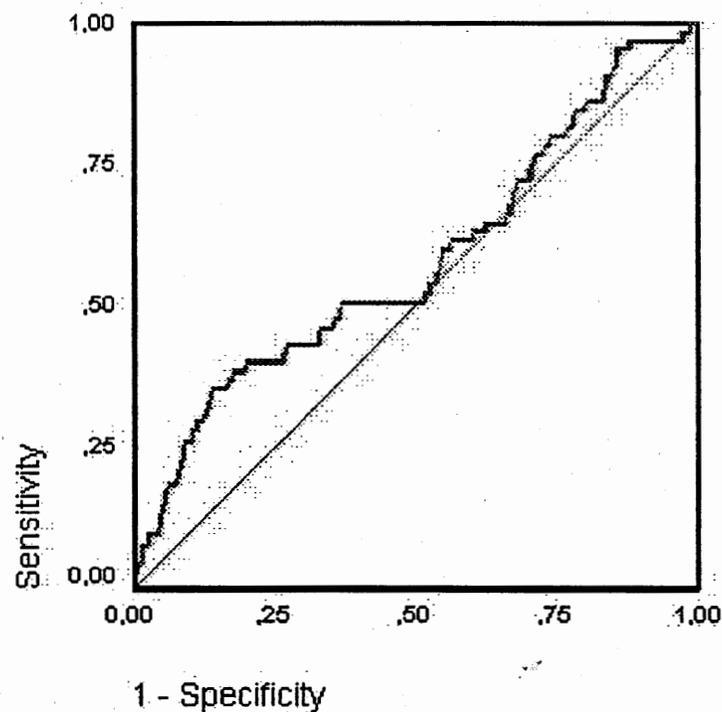
a Under the nonparametric assumption

b Null hypothesis: true area = 0.5

Figure I.15 : « Roc Curve » pour les modèles 40 et 42.

MORT0JB	Valid N (listwise)
(1) nb d'individus mort pendant la période	65
(0) nb d'individus non-mort pendant la période	944
cas sans date de mort	1542
Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.	
a The positive actual state is 1,00.	

### ROC Curve



Diagonal segments are produced by ties.

#### Area Under the Curve

Test Result Variable(s): Predicted probability

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,579	,040	,032	,500	,658
The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.				
a Under the nonparametric assumption				
b Null hypothesis: true area = 0.5				

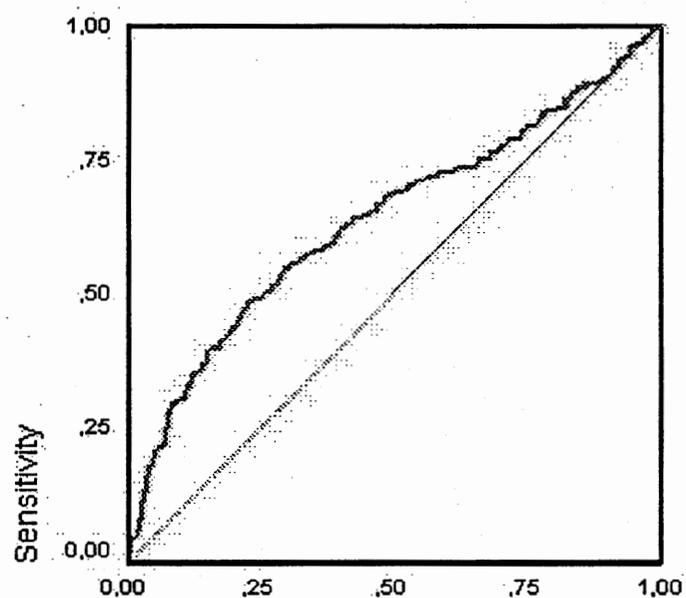
Figure I.16 : « Roc Curve » pour le modèle 41.

Tableau I.3: Tables de classification des modèles de régression logistique utilisant la progéniture féminine.

		Predicted						
		Selected Cases(a)			Unselected Cases(b,c)			
		MORTINFB		Percentage Correct	MORTINFB		Percentage Correct	
Observed	,00	1,00	,00		1,00			
modèle 7	MORTINFB	,00	379	7	98,2	369	7	98,1
		1,00	99	5	4,8	133	10	7,0
	Overall Percentage				78,4			73,0
d The cut value is ,500								
modèle 8	MORTINFB	,00	385	1	99,7	374	2	99,5
		1,00	101	3	2,9	137	6	4,2
	Overall Percentage				79,2			73,2
d The cut value is ,500								
modèle 9	MORTINFB	,00	384	2	99,5	373	3	99,2
		1,00	99	5	4,8	136	7	4,9
	Overall Percentage				79,4			73,2
d The cut value is ,500								
modèles 16 et 17	MORT0JB	,00	480	0	100,0	464	0	100,0
		1,00	10	0	,0	55	0	,0
	Overall Percentage				98,0			89,4
d The cut value is ,500								
modèle 18	MORT0JB	,00	480	0	100,0	464	0	100,0
		1,00	10	0	,0	55	0	,0
	Overall Percentage				98,0			89,4
d The cut value is ,500								
modèle 25, 26 et 27	MORT06B	,00	477	0	100,0	513	0	100,0
		1,00	13	0	,0	6	0	,0
	Overall Percentage				97,3			98,8
d The cut value is ,500								
modèles 43, 44 et 45	MORT14B	,00	385	0	100,0	423	0	100,0
		1,00	105	0	,0	96	0	,0
	Overall Percentage				78,6			81,5
d The cut value is ,500								

MORTINFB	Valid N (listwise)
(1) nb d'individus mort pendant la période	247
(0) nb d'individus non-mort pendant la période	762
cas sans date de mort	1542
Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.	
a The positive actual state is 1,00.	

### ROC Curve



1 - Specificity

Diagonal segments are produced by ties.

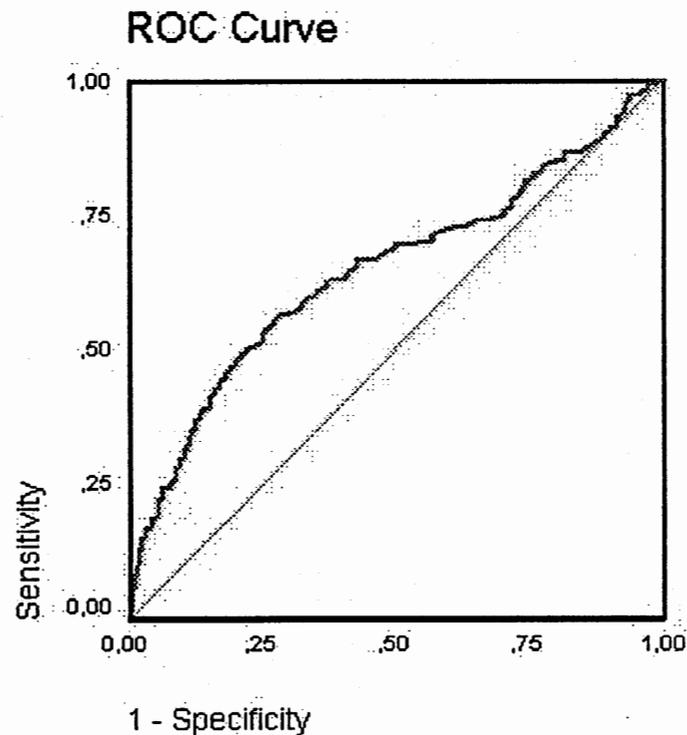
#### Area Under the Curve

Test Result Variable(s): Predicted probability

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,639	,022	,000	,596	,682
The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.				
a Under the nonparametric assumption				
b Null hypothesis: true area = 0.5				

Figure I.17 : « Roc Curve » pour le modèle 7.

MORTINFB	Valid N (listwise)
(1) nb d'individus mort pendant la période	247
(0) nb d'individus non-mort pendant la période	762
cas sans date de mort	1542
Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.	
a The positive actual state is 1,00.	



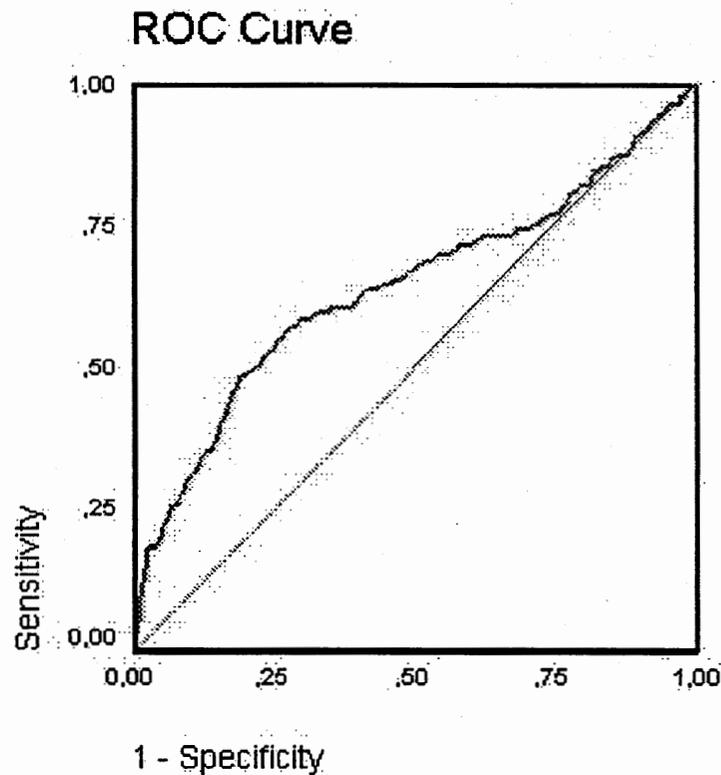
Diagonal segments are produced by ties.

**Area Under the Curve**  
Test Result Variable(s): Predicted probability

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,648	,022	,000	,605	,692
The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.				
a Under the nonparametric assumption				
b Null hypothesis: true area = 0.5				

Figure I.18 : « Roc Curve » pour le modèle 8.

MORTINFB	Valid N (listwise)
(1) nb d'individus mort pendant la période	247
(0) nb d'individus non-mort pendant la période	762
cas sans date de mort	1542
Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.	
a The positive actual state is 1,00.	



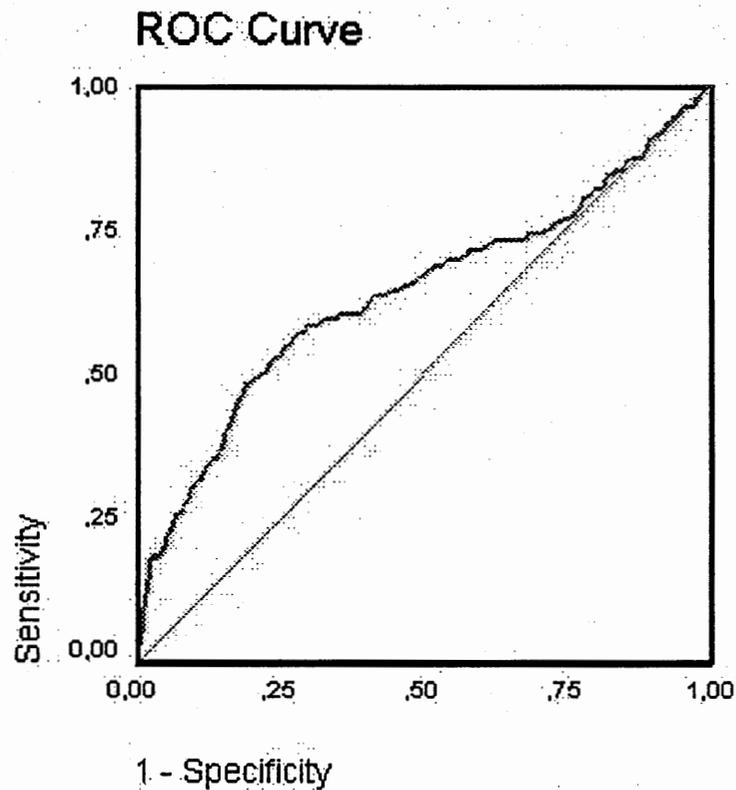
Diagonal segments are produced by ties.

**Area Under the Curve**  
Test Result Variable(s): Predicted probability

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,642	,022	,000	,598	,686
The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.				
a Under the nonparametric assumption				
b Null hypothesis: true area = 0.5				

Figure I.19 : « Roc Curve » pour le modèle 9.

MORTINFB	Valid N (listwise)
(1) nb d'individus mort pendant la période	247
(0) nb d'individus non-mort pendant la période	762
cas sans date de mort	1542
Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.	
a The positive actual state is 1,00.	



Diagonal segments are produced by ties.

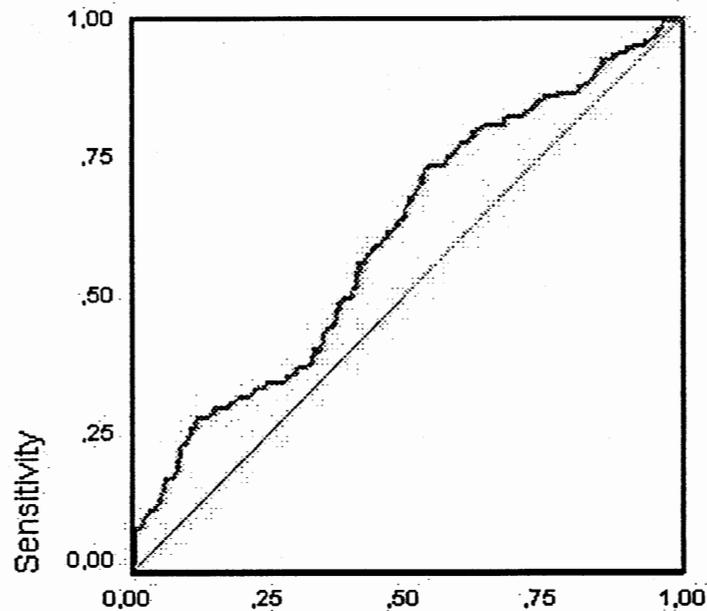
**Area Under the Curve**  
Test Result Variable(s): Predicted probability

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,641	,023	,000	,597	,686
The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.				
a Under the nonparametric assumption				
b Null hypothesis: true area = 0.5				

Figure I.20 : « Roc Curve » pour les modèles 16 et 17.

MORTINFB	Valid N (listwise)
(1) nb d'individus mort pendant la période	247
(0) nb d'individus non-mort pendant la période	762
cas sans date de mort	1542
Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.	
a The positive actual state is 1,00.	

### ROC Curve



1 - Specificity

Diagonal segments are produced by ties.

#### Area Under the Curve

Test Result Variable(s): Predicted probability

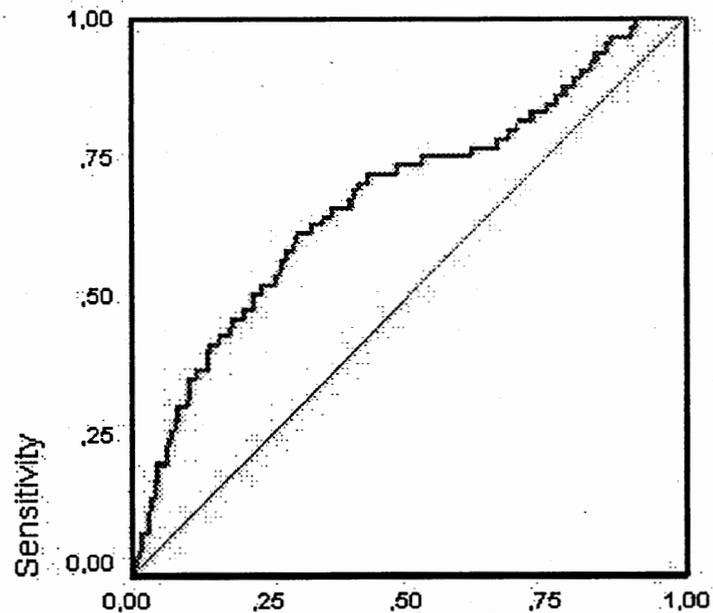
Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
.603	.021	.000	.562	.643
The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.				
a Under the nonparametric assumption				
b Null hypothesis: true area = 0.5				

Figure I.21 : « Roc Curve » pour le modèle 18.

MORT0JB	Valid N (listwise)
(1) nb d'individus mort pendant la période	65
(0) nb d'individus non-mort pendant la période	944
cas sans date de mort	1542
Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.	
a The positive actual state is 1,00.	

Area Under the Curve

### ROC Curve



1 - Specificity

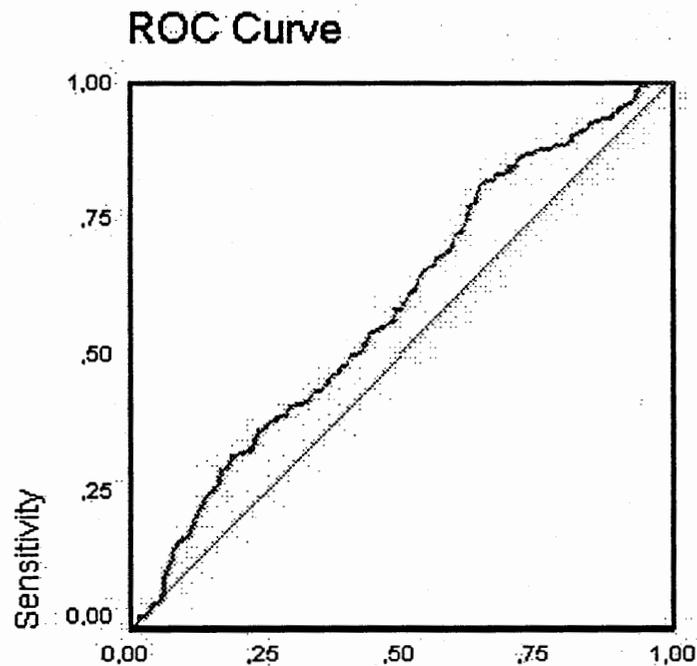
Diagonal segments are produced by ties.

Test Result Variable(s): Predicted probability

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,678	,037	,000	,606	,751
The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.				
a Under the nonparametric assumption				
b Null hypothesis: true area = 0.5				

Figure I.22 : « Roc Curve » pour les modèles 25, 26 et 27.

MORT14B	Valid N (listwise)
(1) nb d'individus mort pendant la période	201
(0) nb d'individus non-mort pendant la période	808
cas sans date de mort	1542
Larger values of the test result variable(s) indicate stronger evidence for a positive actual state.	
a The positive actual state is 1,00.	



1 - Specificity

Diagonal segments are produced by ties.

**Area Under the Curve**

Test Result Variable(s): Predicted probability

Area	Std. Error(a)	Asymptotic Sig.(b)	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
,587	,022	,000	,544	,630

The test result variable(s): Predicted probability has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

a Under the nonparametric assumption

b Null hypothesis: true area = 0.5

Figure I.23 : « Roc Curve » pour les modèles 43, 44 et 45.

## **APPENDICE J**

### **Analyse des groupes exclus**

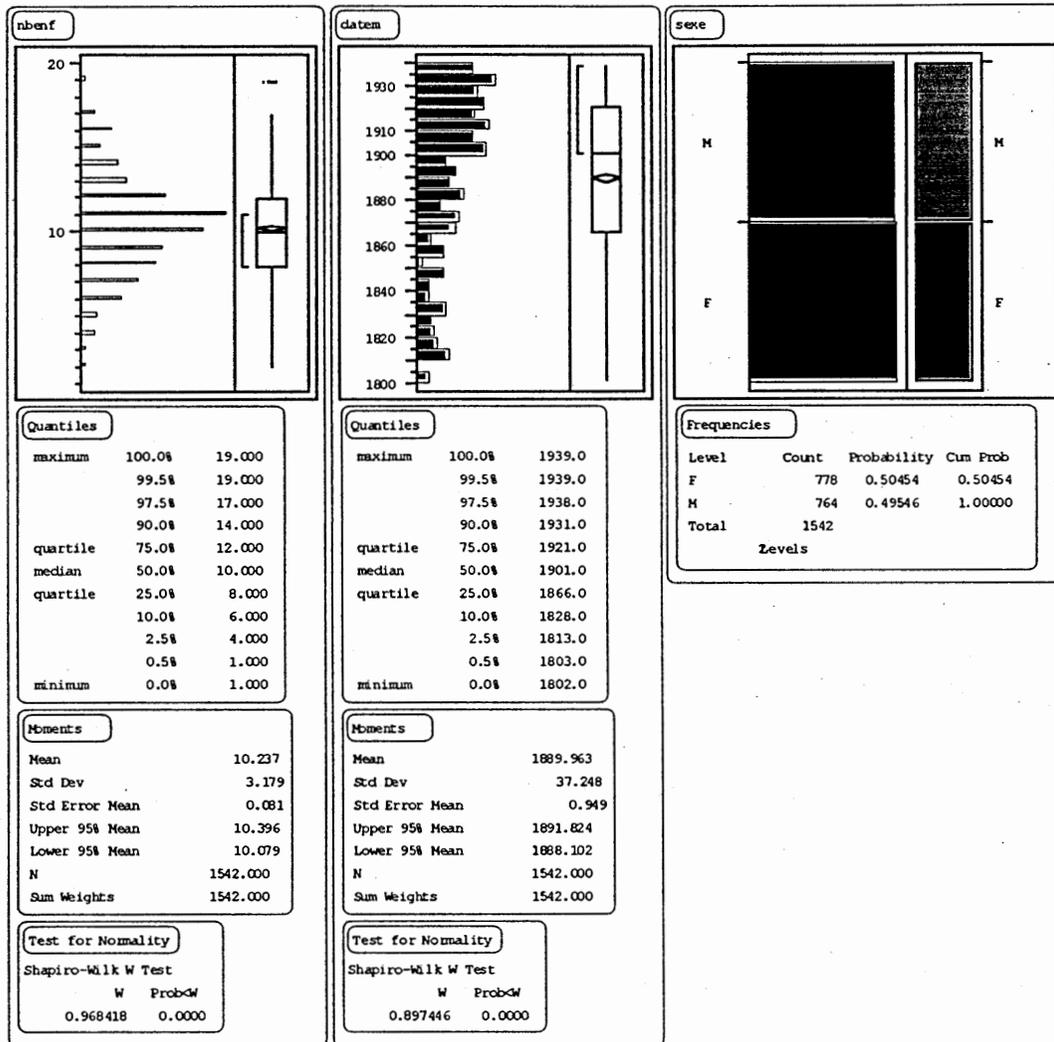


Figure J.1. Distribution, moyenne et écart-type des variables nombre d'enfants du couple (à gauche), date de mariage des parents (au centre) et sexe de l'enfant (à droite) pour le groupe des « exclus ».

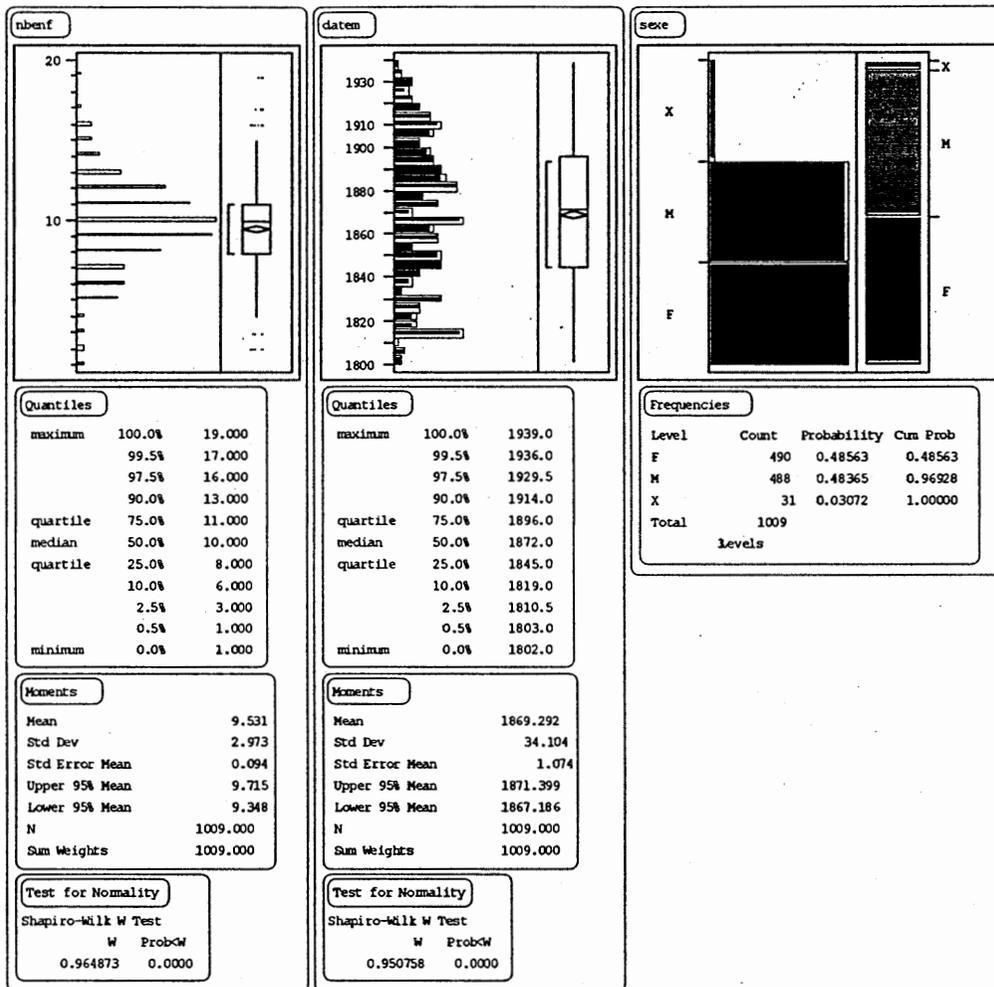


Figure J.2. Distribution, moyenne et écart-type des variables nombre d'enfants du couple (à gauche), date de mariage des parents (au centre) et sexe de l'enfant (à droite) pour le groupe des « inclus ».

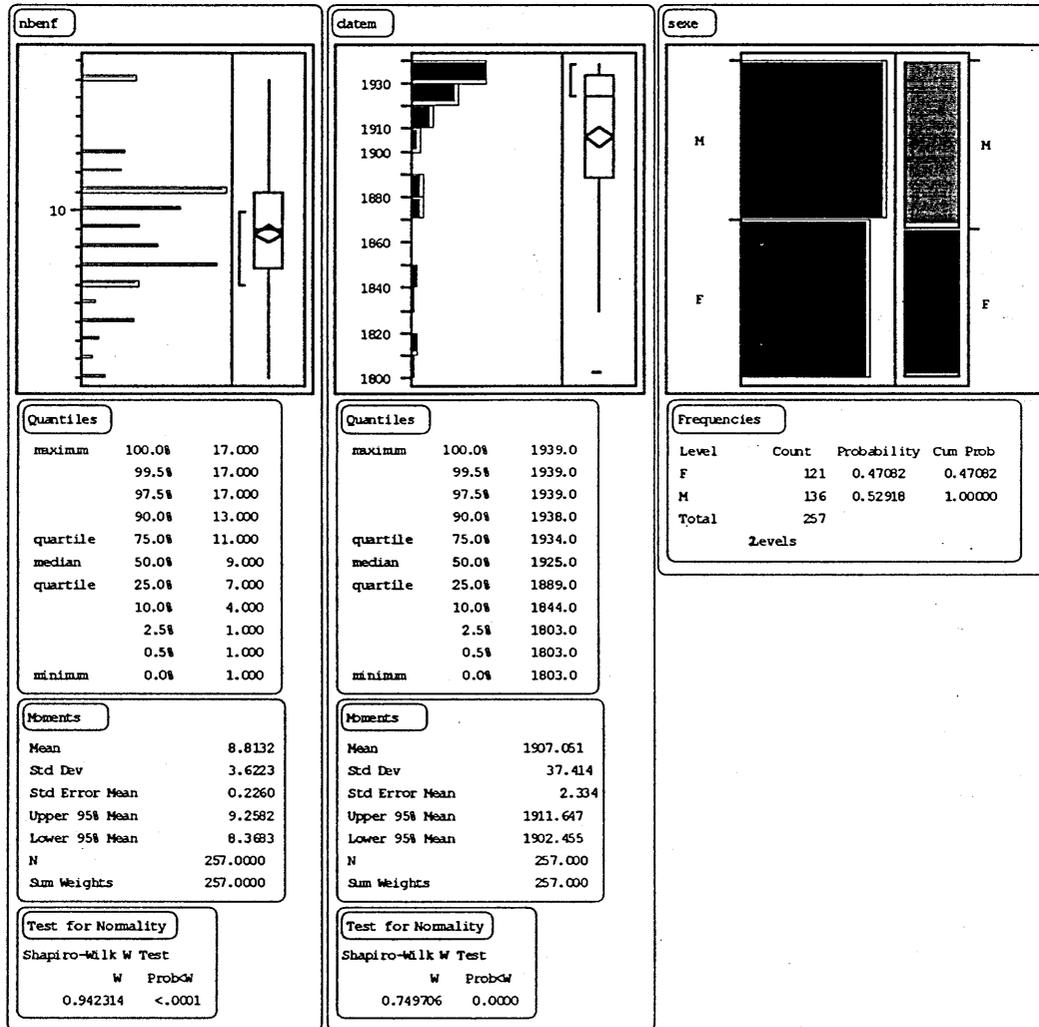


Figure J.3. Distribution, moyenne et écart-type des variables nombre d'enfants du couple (à gauche), date de mariage des parents (au centre) et sexe de l'enfant (à droite) pour le groupe des « couples ».

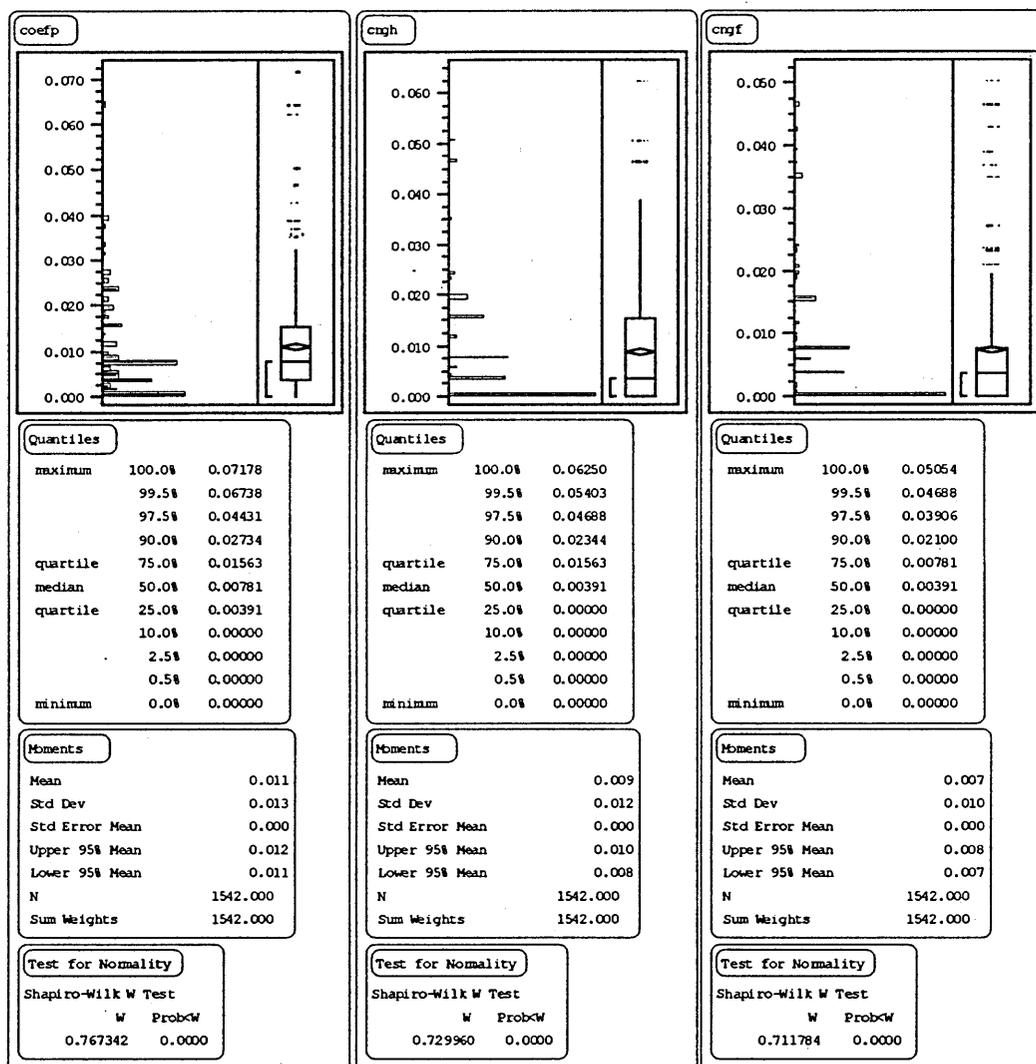


Figure J.4. Distribution, moyenne et écart-type des variables apparemment total des parents (à gauche), consanguinité totale du père (au centre) et consanguinité totale de la mère (à droite) pour le groupe des « exclus ».

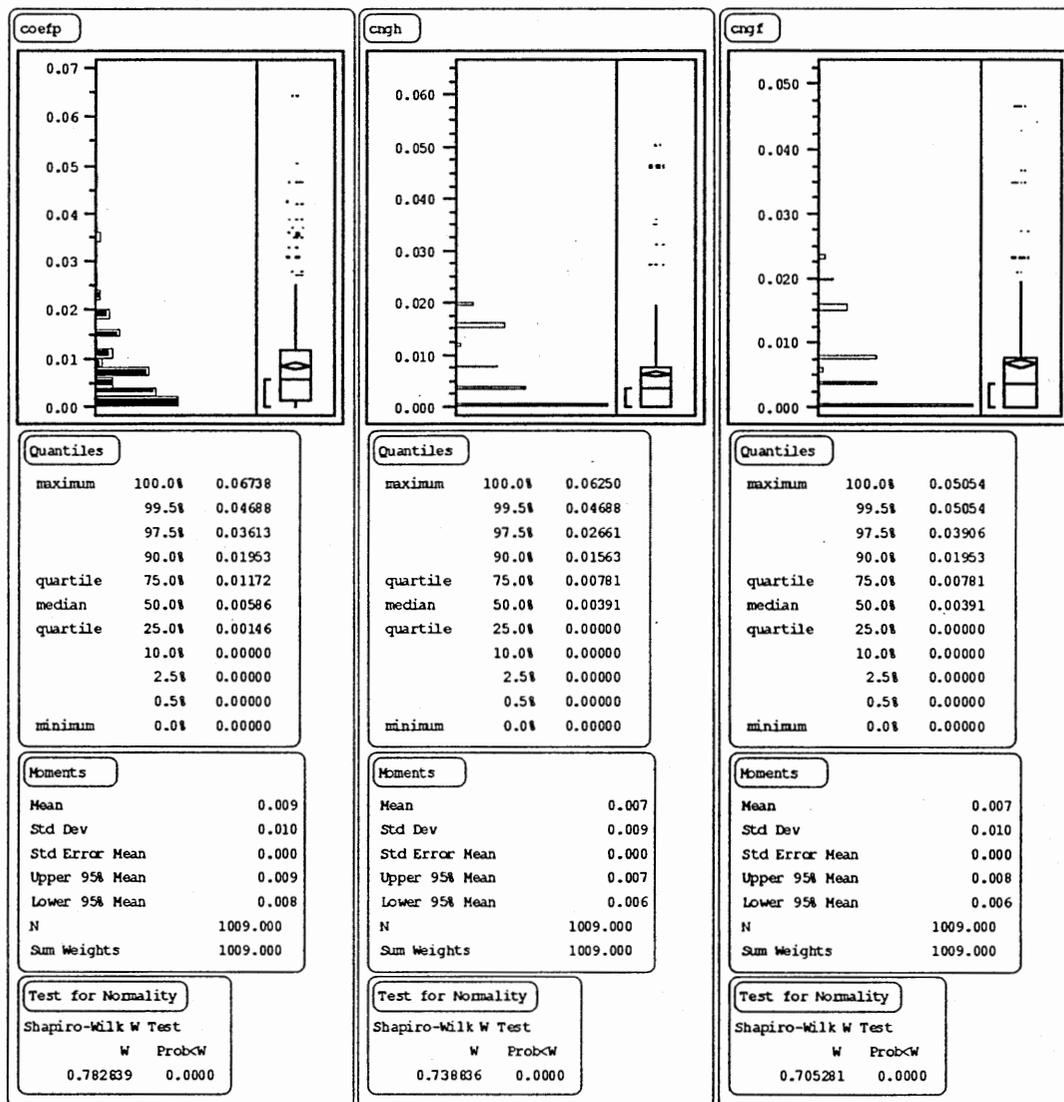


Figure J.5. Distribution, moyenne et écart-type des variables apparemment total des parents (à gauche), consanguinité totale du père (au centre) et consanguinité totale de la mère (à droite) pour le groupe des « inclus ».

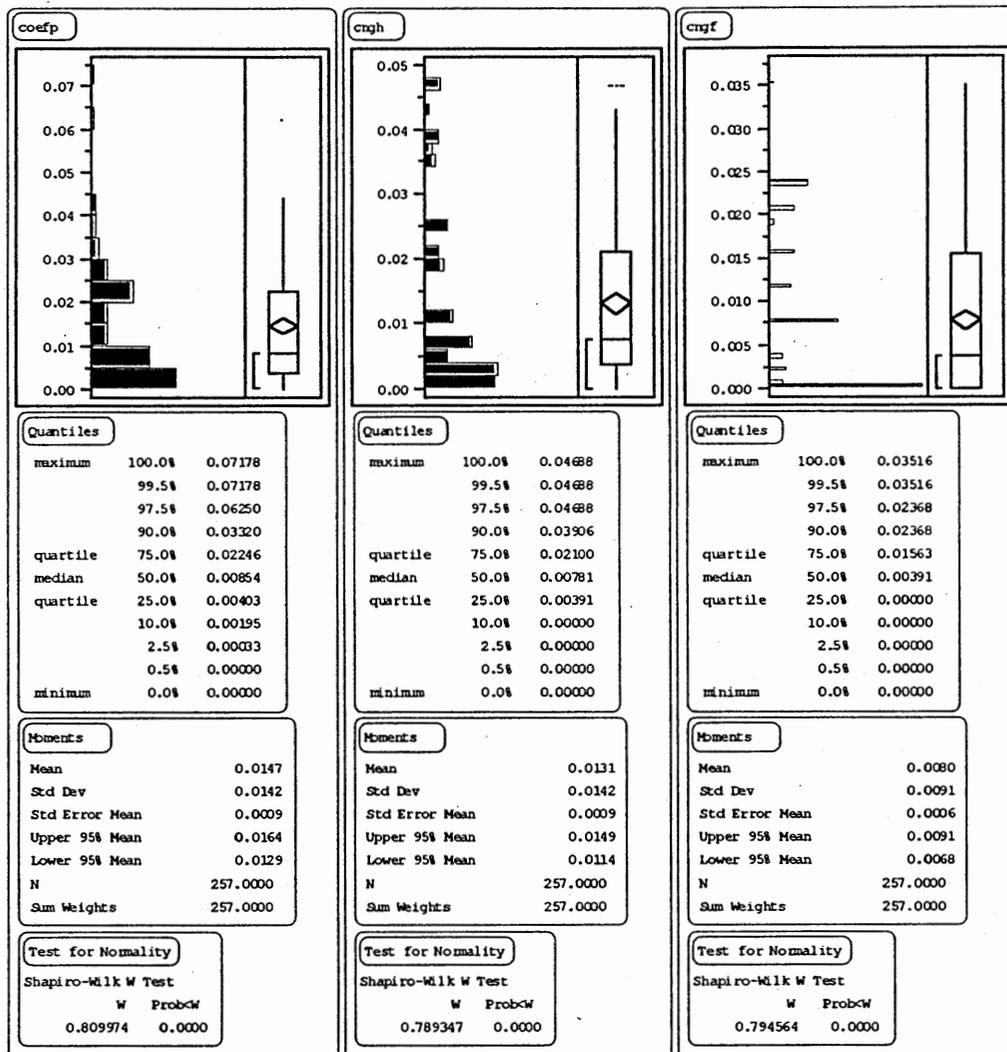


Figure J.6. Distribution, moyenne et écart-type des variables apparemment total des parents (à gauche), consanguinité totale du père (au centre) et consanguinité totale de la mère (à droite) pour le groupe des « couples ».

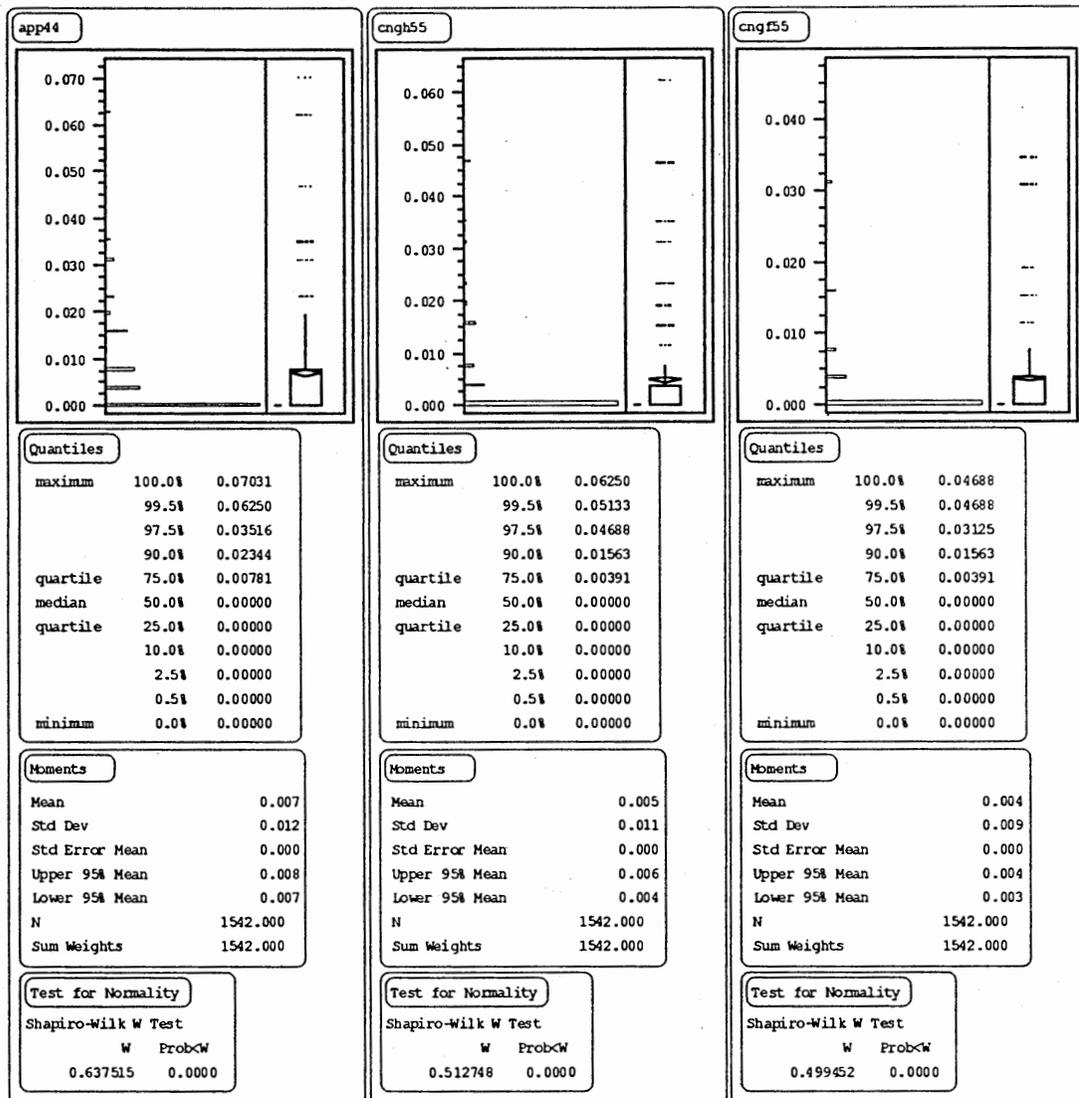


Figure J.7. Distribution, moyenne et écart-type des variables apparemment proche des parents (à gauche), consanguinité proche du père (au centre) et consanguinité proche de la mère (à droite) pour le groupe des « exclus ».

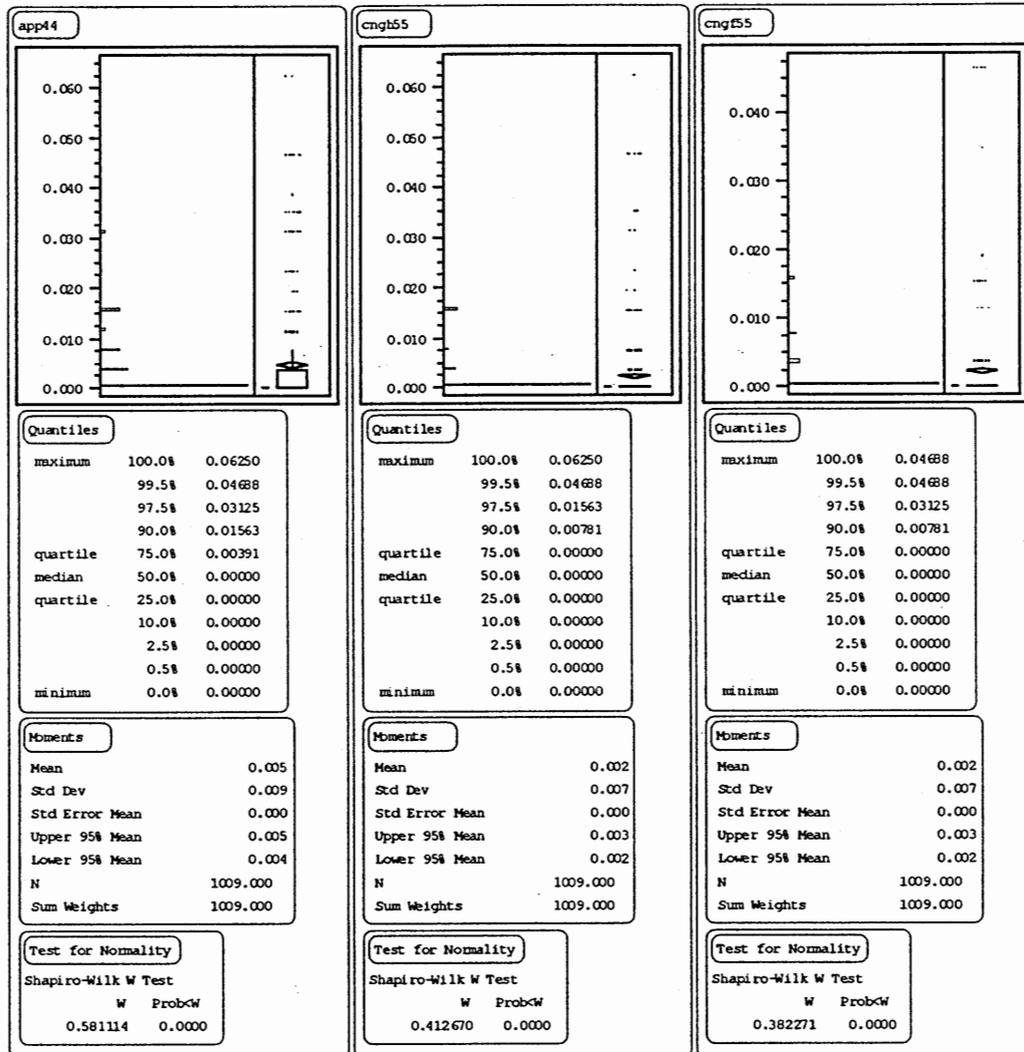


Figure J.8. Distribution, moyenne et écart-type des variables apparemment proche des parents (à gauche), consanguinité proche du père (au centre) et consanguinité proche de la mère (à droite) pour le groupe des « inclus ».

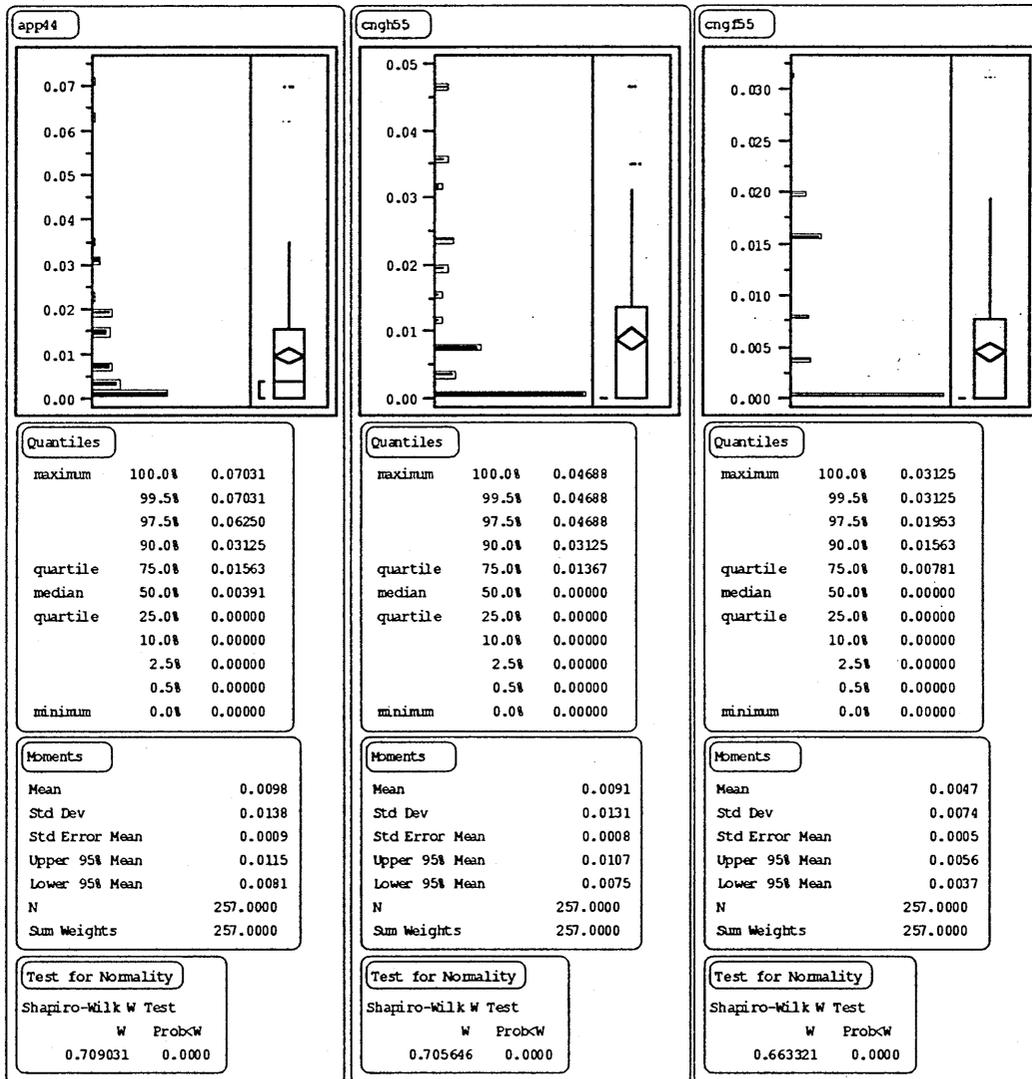


Figure J.9. Distribution, moyenne et écart-type des variables apparemment proche des parents (à gauche), consanguinité proche du père (au centre) et consanguinité proche de la mère (à droite) pour le groupe des « couples ».

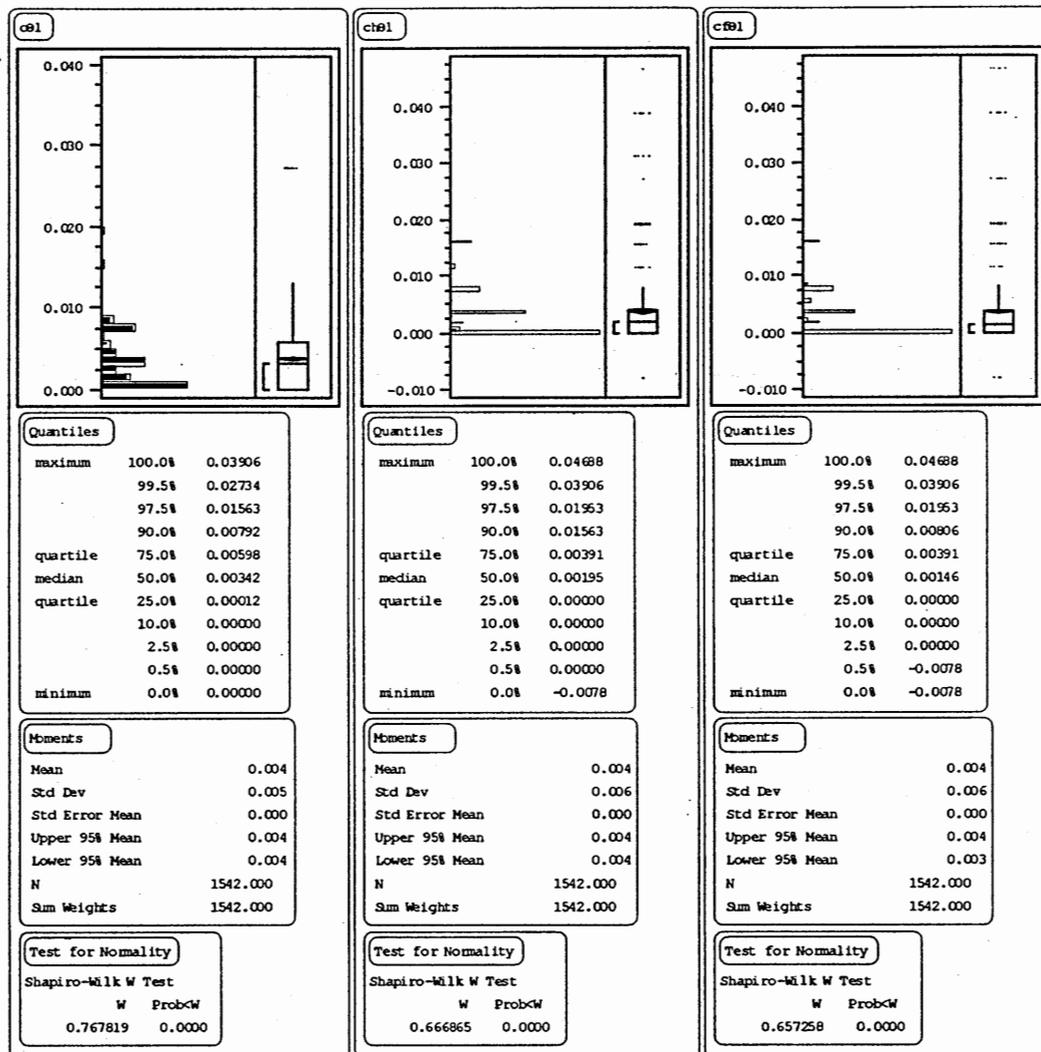


Figure J.10. Distribution, moyenne et écart-type des variables apparemment éloigné des parents (à gauche), consanguinité éloignée du père (au centre) et consanguinité éloignée de la mère (à droite) pour le groupe des « exclus ».

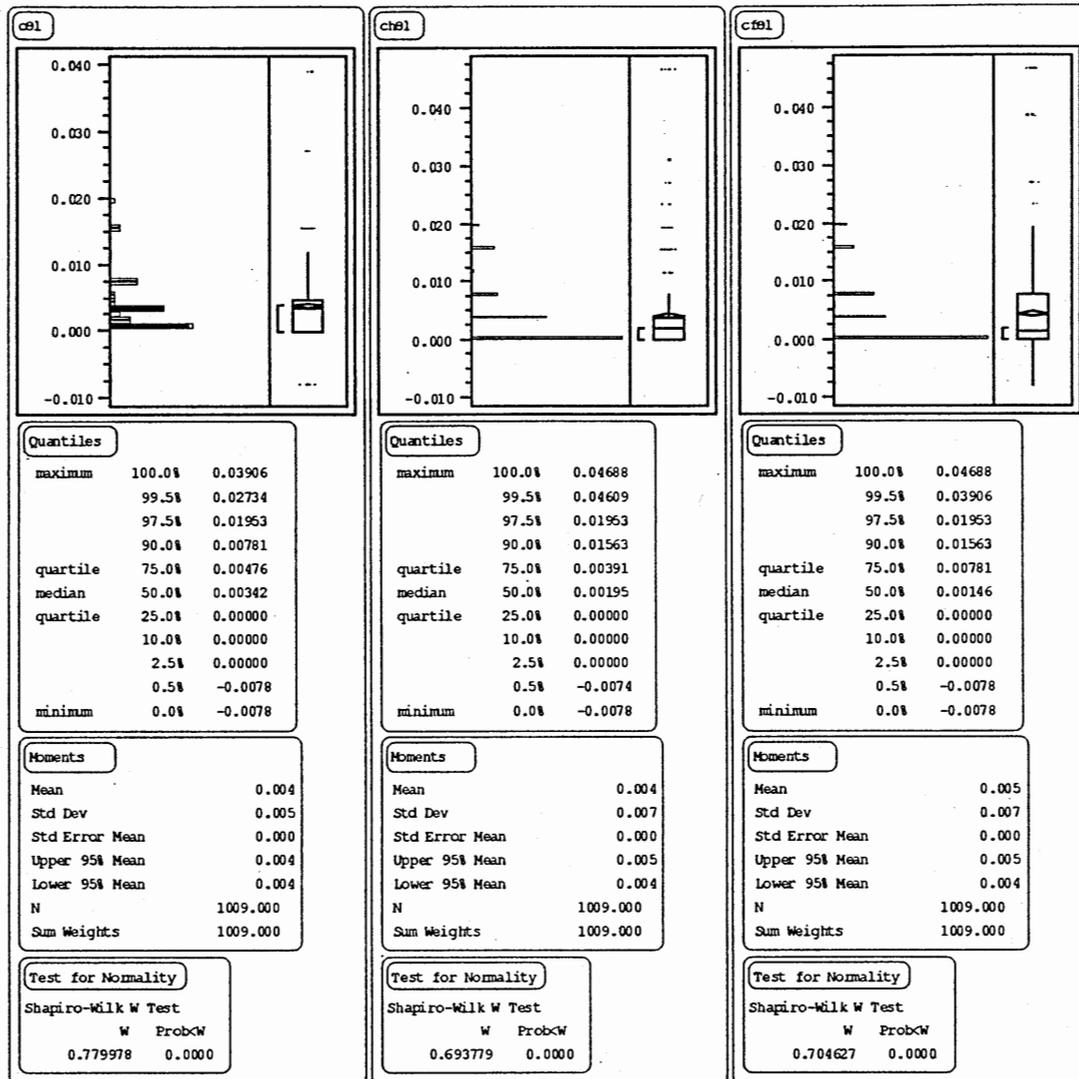


Figure J.11. Distribution, moyenne et écart-type des variables apparemment éloigné des parents (à gauche), consanguinité éloignée du père (au centre) et consanguinité éloignée de la mère (à droite) pour le groupe des « inclus ».

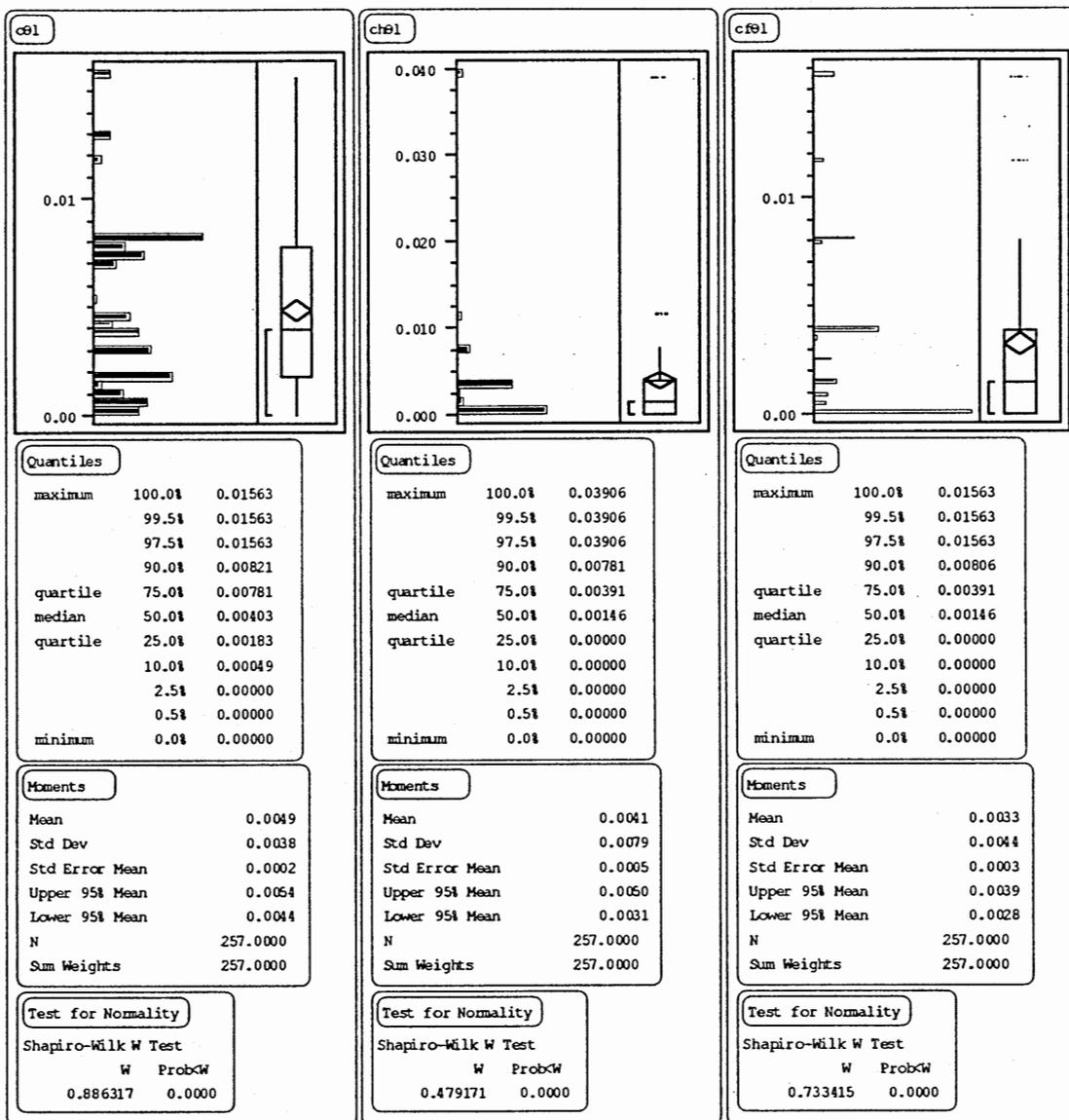


Figure J.12. Distribution, moyenne et écart-type des variables apparemment éloigné des parents (à gauche), consanguinité éloignée du père (au centre) et consanguinité éloignée de la mère (à droite) pour le groupe des « couples ».

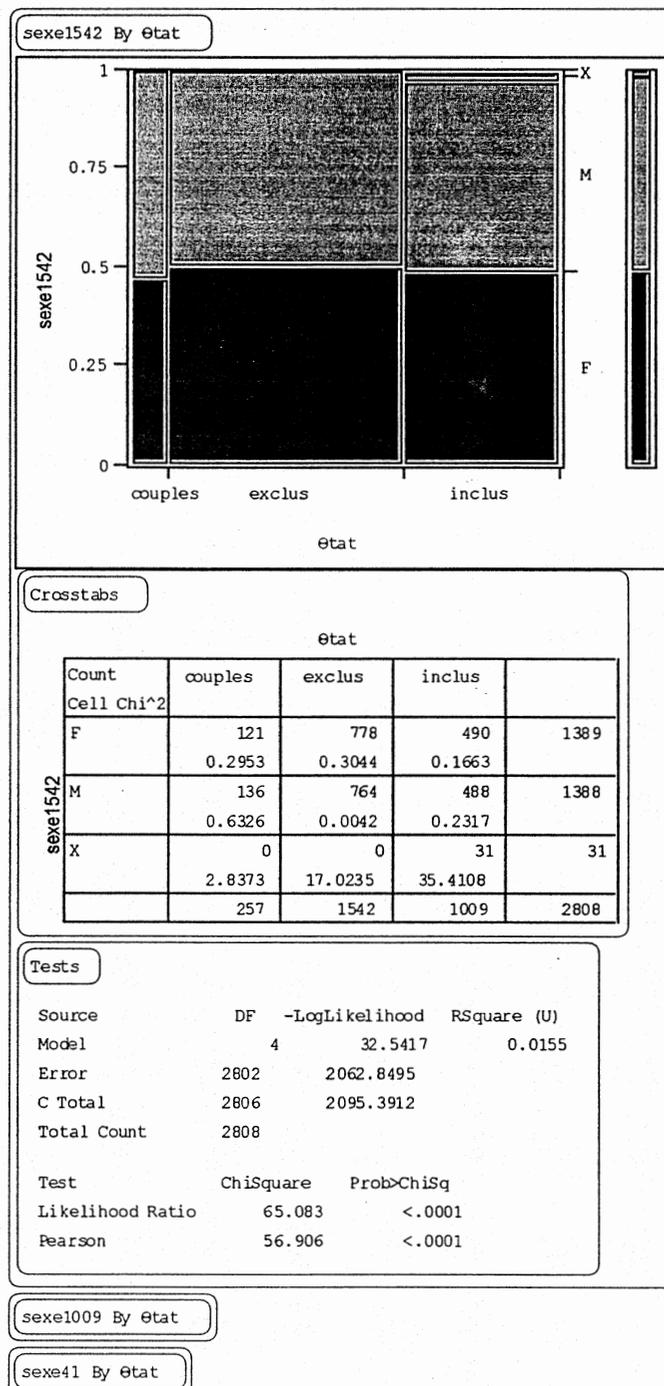


Figure J.13. Chi-carré entre les groupes des « exclus », des « inclus » et de « couples » pour la variable « sexe de l'enfant ».

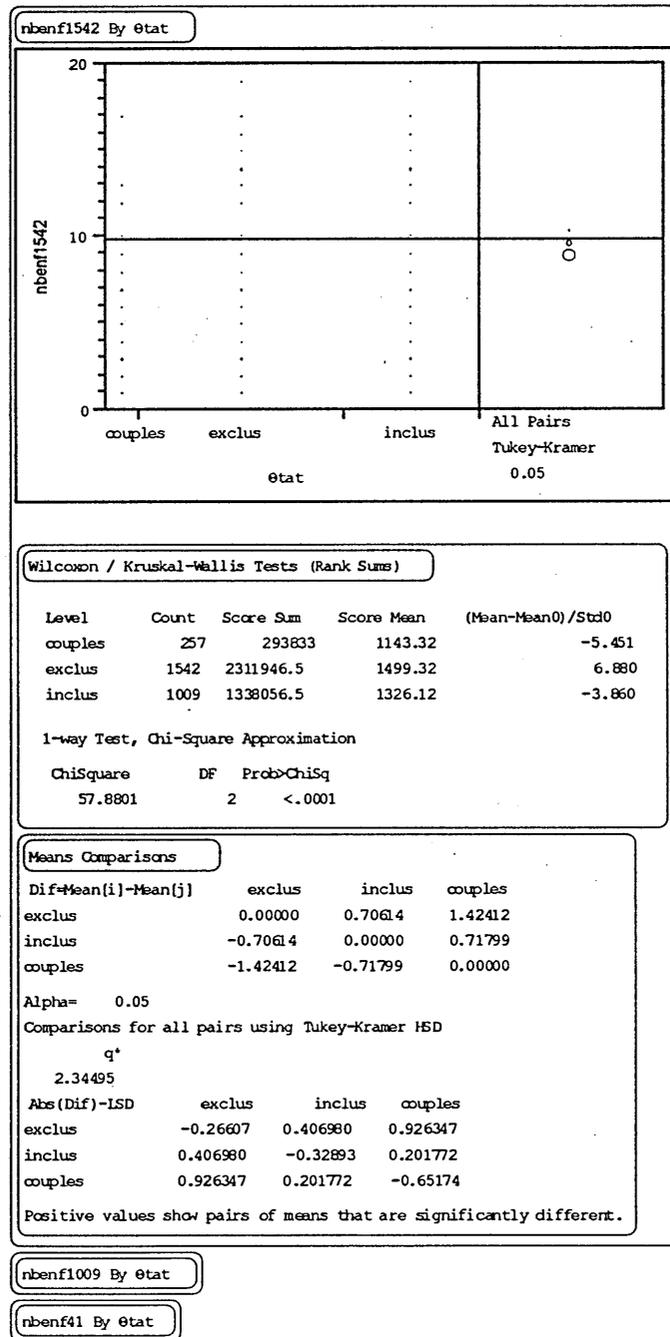


figure J.14. Tests de comparaison de moyennes entre les groupes « couples », « exclus » et « inclus » et test de Tuckey pour la variable « nombre d'enfants du couple ».

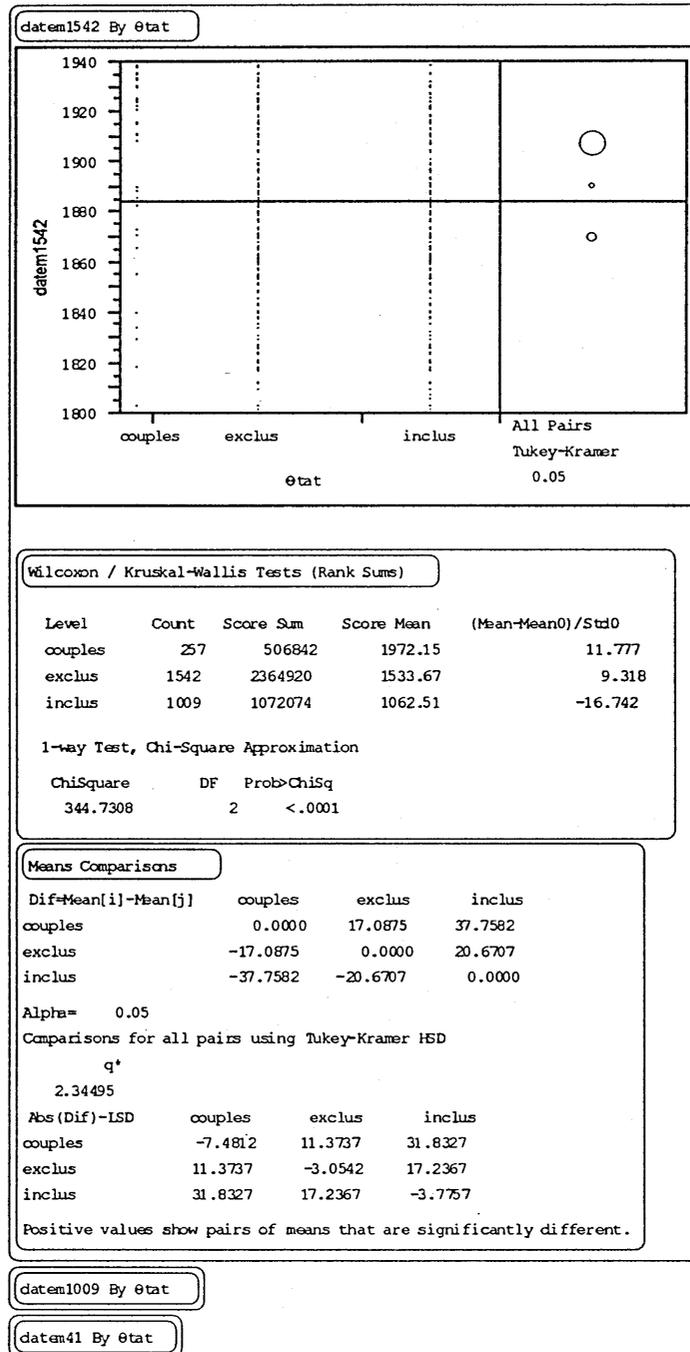


figure J.15. Tests de comparaison de moyennes entre les groupes « couples », « exclus » et « inclus » et test de Tuckey pour la variable « date de mariage des parents ».

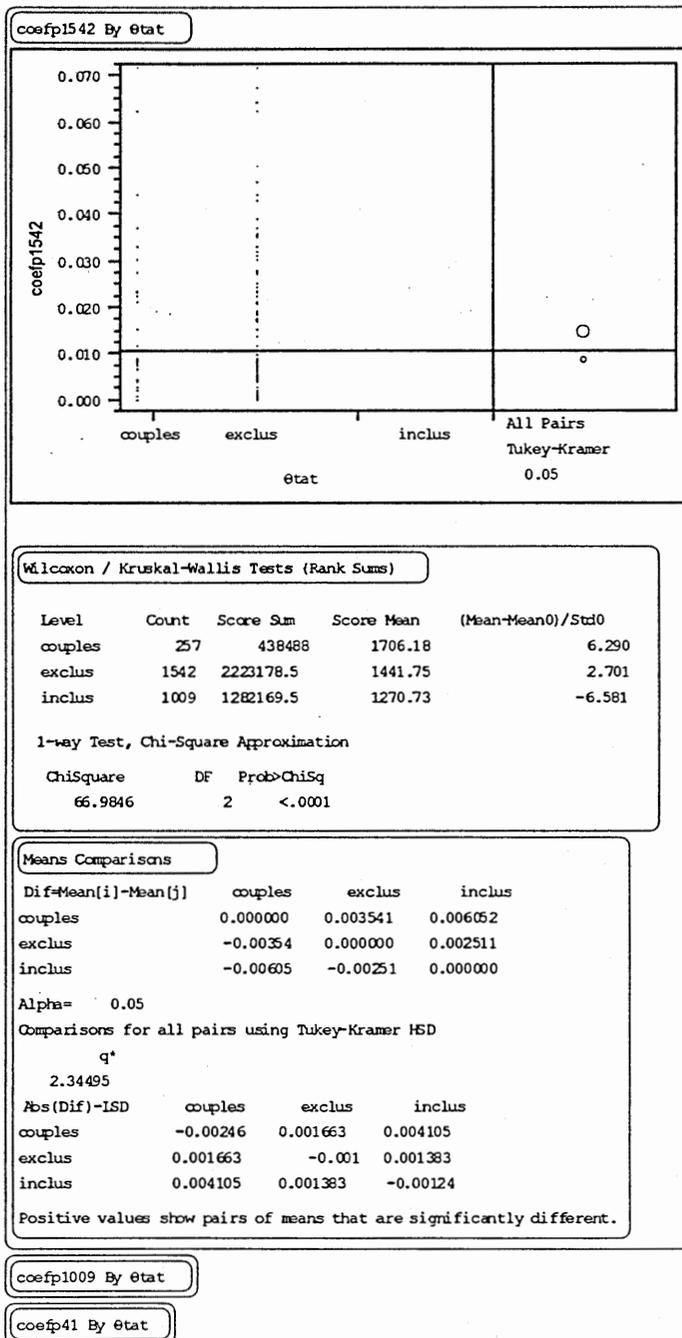


figure J.16. Tests de comparaison de moyennes entre les groupes « couples », « exclus » et « inclus » et test de Tuckey pour la variable « apparemment total des parents ».

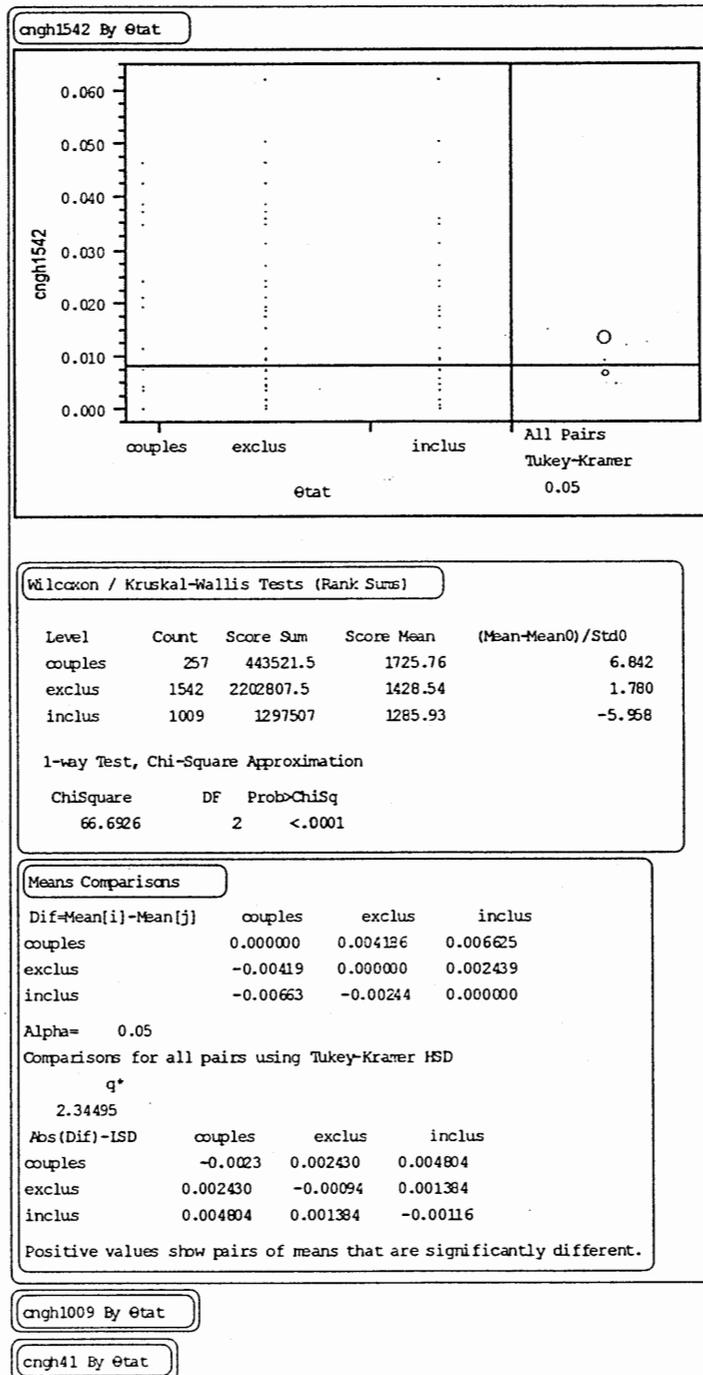


figure J.17. Tests de comparaison de moyennes entre les groupes « couples », « exclus » et « inclus » et test de Tuckey pour la variable « consanguinité paternelle totale ».

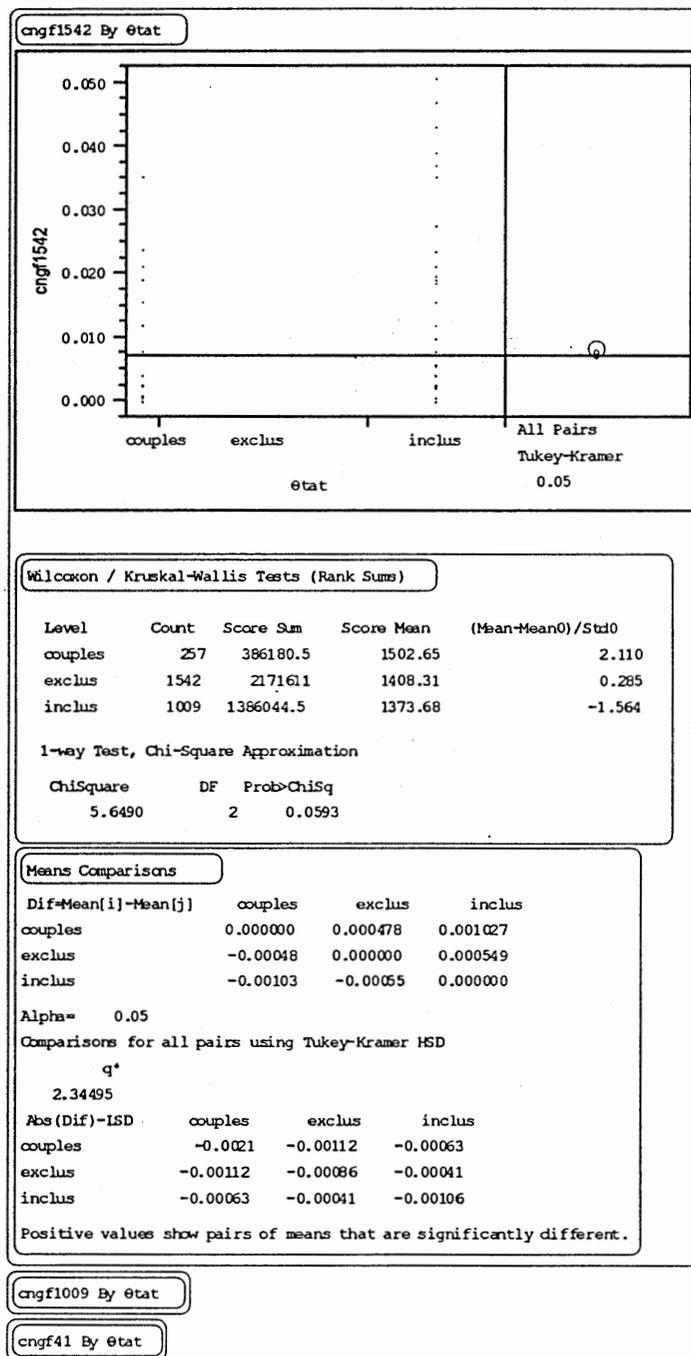


figure J.18. Tests de comparaison de moyennes entre les groupes « couples », « exclus » et « inclus » et test de Tuckey pour la variable « consanguinité maternelle totale ».

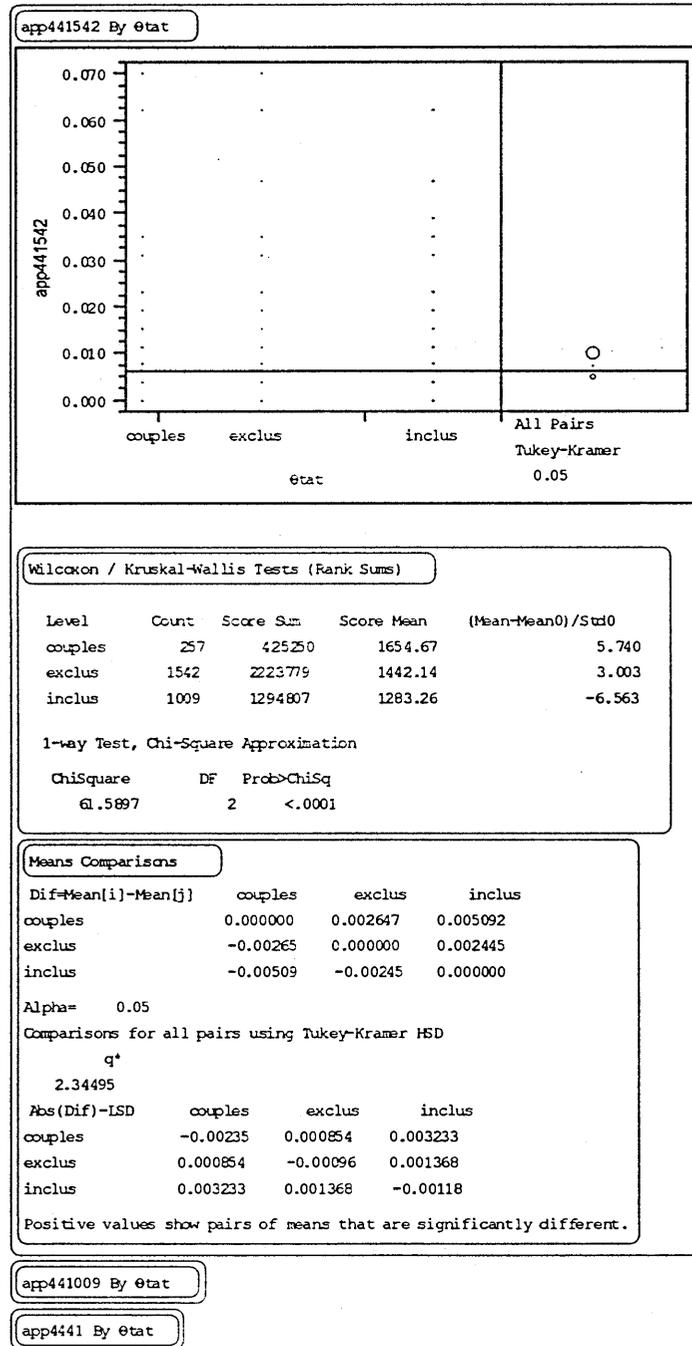


figure J.19. Tests de comparaison de moyennes entre les groupes « couples », « exclus » et « inclus » et test de Tuckey pour la variable « apparemment proche des parents ».

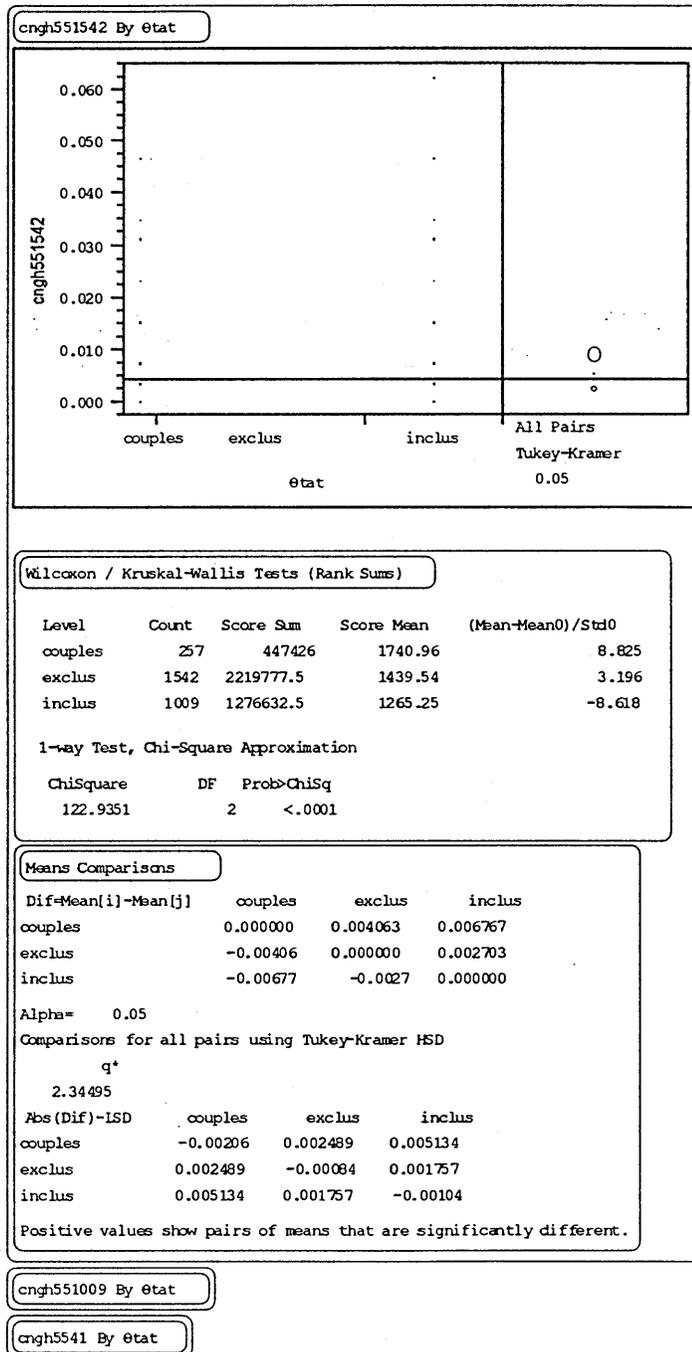


figure J.20. Tests de comparaison de moyennes entre les groupes « couples », « exclus » et « inclus » et test de Tuckey pour la variable « consanguinité paternelle proche ».

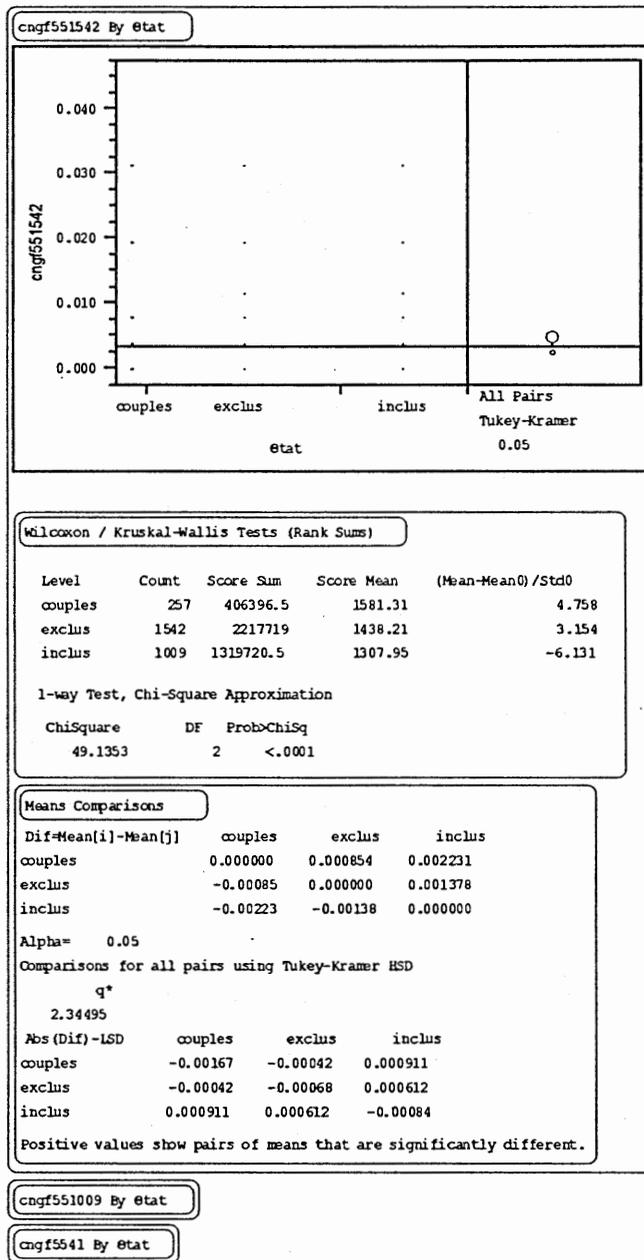


figure J.21. Tests de comparaison de moyennes entre les groupes « couples », « exclus » et « inclus » et test de Tuckey pour la variable « consanguinité maternelle proche ».

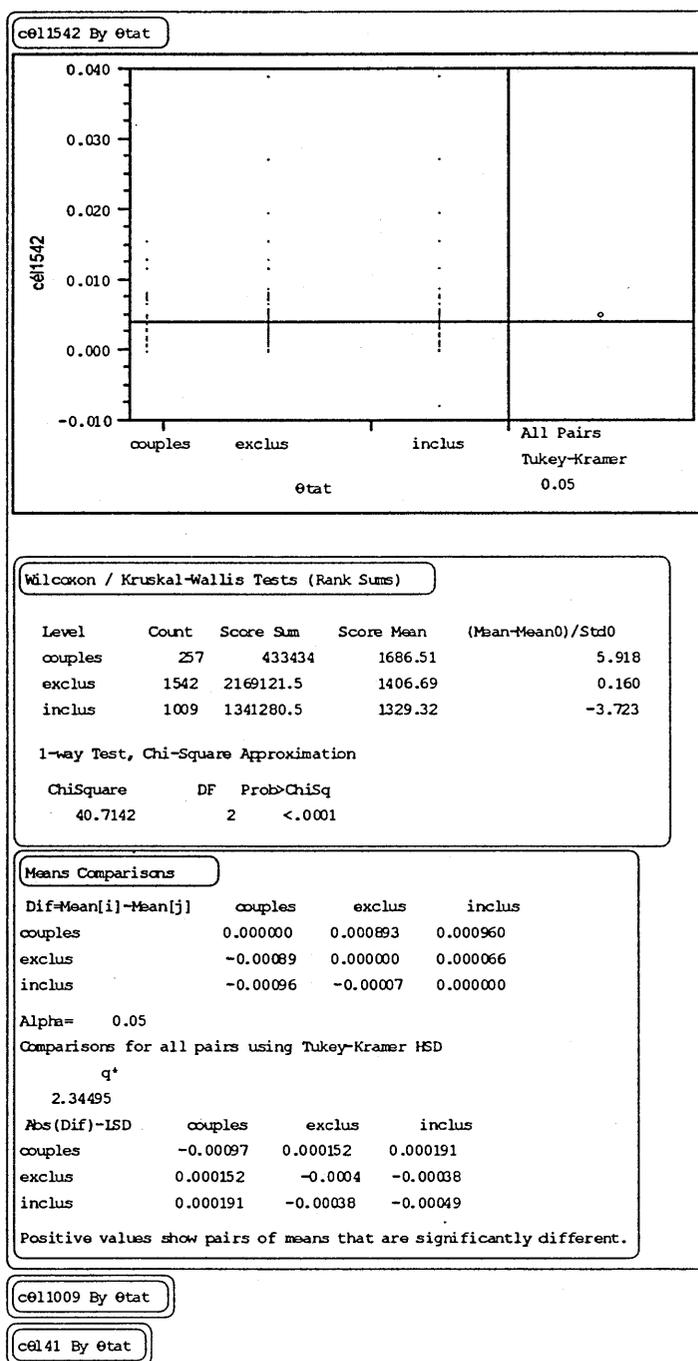


Figure J.22. Tests de comparaison de moyennes entre les groupes « couples », « exclus » et « inclus » et test de Tuckey pour la variable « apparemment  loign  des parents ».

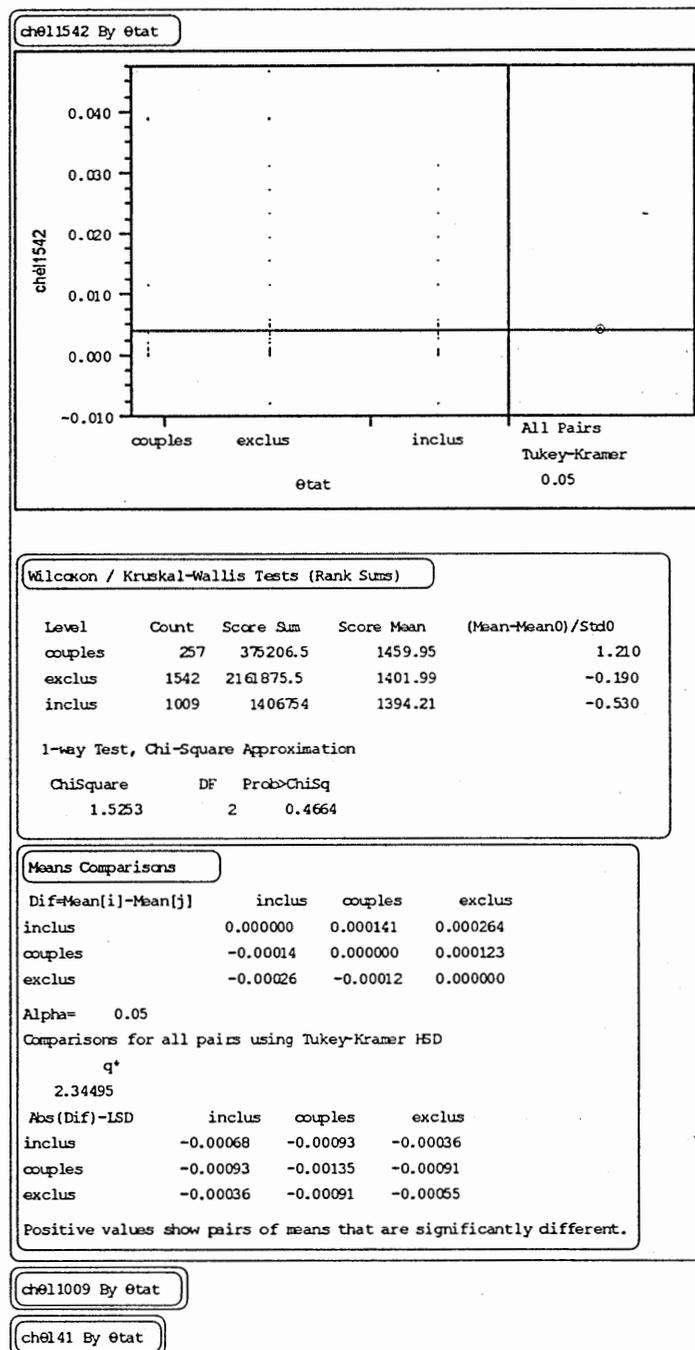


Figure J.23. Tests de comparaison de moyennes entre les groupes « couples », « exclus » et « inclus » et test de Tuckey pour la variable « consanguinité paternelle éloignée ».

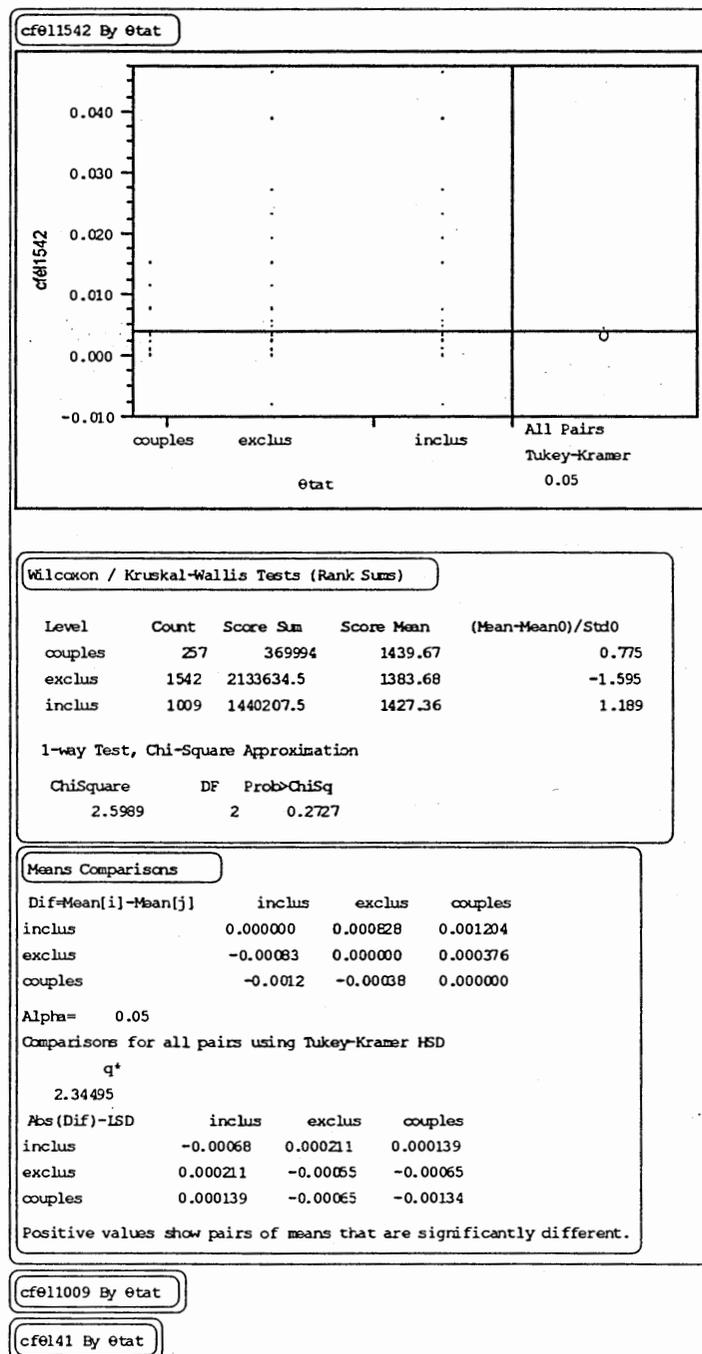


figure J.24. Tests de comparaison de moyennes entre les groupes « couples », « exclus » et « inclus » et test de Tuckey pour la variable « consanguinité maternelle éloignée ».

**BIBLIOGRAPHIE**

Alberts, Susan C. et Carole Ober. 1993. « Genetic Variability in the MHC: a Review of Non-Pathogen-Mediated Selective Mechanisms ». *Yearbook of Physical Anthropology*, vol. 36, p. 71-89.

Adinolfi, M., P.E. Polani et J.A. Crolla. 1985. « Is the Sex Ratio at Birth Affected by Immune Selection? ». *Experimental and Clinical Immunogenetics*, vol. 2, p. 54-64.

Allain, Katia. 1996. *La mortalité infantile et la consanguinité: un lien certain...*, Rapport de stage, Montréal, Université du Québec à Montréal, 54 p.

Astolfi P., M. Martinetti, F. Gigli-Berzolari et M. Cuccia. 1990. « The Effect of Parental and Maternal-Fetal Histocompatibility at MHC on Sex Ratio in Offspring », *Tissue Antigens*, vol. 35, p. 172-177.

Beaudry, Jean-R. 1985. *Génétiq ue générale*, Paris, Décarie Maloine, 501p.

Beer, Alan E., J.F. Quebbeman, J.W.T. Ayers, R.F. Haines. 1981. « Major Histocompatibility Complex Antigens, Maternal and Paternal Immune Responses, and Chronic Habitual Abortions in Humans ». *American Journal of Obstetrics and Gynecology*, vol. 141, no 8, p.987-999.

Bittles, Alan H. 1994. « The Role and Significance of Consanguinity as a Demographic Variable ». *Populations and Development Review*, vol. 20, p 561-84.

Bodmer, W.F. 1972. « Evolutionary Significance of the HLA System ». *Nature*, vol. 327, p. 139-145.

Boisvert, Mireille et Francine M. Mayer. 1994. « Mortalité infantile et consanguinité dans une population endogame du Québec ». *Populations*, vol 3, p. 685-724.

Boisvert, Mireille. 1992. *Mortalité infantile et consanguinité dans une population isolée du Québec*. Mémoire de maîtrise, Montréal, Université du Québec à Montréal, 1992, 251p.

Brittain, Ann W. 1992. « Birth Spacing and Child Mortality in a Caribbean Population », *Human Biology*, vol.64, no. 2, p. 223-241.

De Brakeleer, M. et M. Ross. 1991. « Inbreeding in Saguenay-Lac-Saint-Jean (Québec,Canada): A Study of Catholic Church Dispensations 1842-1971 », *Human Heredity*, vol. 41, p 379-384.

De Brakeleer, Marc. 1991. « Les gènes délétères ». In *Histoire d'un génome*, Québec, P.U.Q., p. 345-364.

Dronamraju K.R. 1963. « Genetic Studies of the Andras Pradesh Population in: *The Genetics of Migrants and Isolate Populations*, p.154-159. (Goldschmidt edit.), New York, Williams and Williams.

Edmond M. et M. De Braekeleer. 1993a. « Inbreeding Effects on Prereproductive Mortality: a Case-Control Study in Saguenay-Lac-Saint-Jean (Québec,Canada) Based on a Population Registry 1838-1971 », *Annals of Human Biology*, vol. 20, no. 6,p. 535-543.

Edmond M. et M. De Braekeleer. 1993b. « Inbreeding Effects on Fertility and Sterility: a Case-Control Study in Saguenay-Lac-Saint-Jean (Québec, Canada) based on a population registry 1838-1971 », *Annals of Human Biology*, vol.20, no. 6, p 545-555.

Futuyma, Douglas J. 1986. *Evolutionary Biology*, Sunderland, Massachussets, Sinauer Associated inc., Second Edition, 600p.

Gill, Thomas J. 1997. « Presidential Adress, Genetic Factors in Reproduction and Their Evolutionary Significance ». *American Journal of Reproductive Immunology*, vol. 37, p. 7-16.

Grant, J.C. et A.H. Bittles. 1997. « The Comparative Role of Consanguinity and Childhood Mortality in Pakistan », *Annals of Human Genetics*, vol. 61, p 143-149.

Griffiths, Anthony J.F., J.F. Miller, D.T. Suzuki, R.C. Lewontin, W.M Gelbart. 1997. *Introduction à l'analyse génétique*. Paris, De Boeck Université s.a., 915 p.

Haldane, J.B.S. 1990. *Causes of Evolution*. Princeton, Princeton University Press, J., 327p.

Henry, Louis. 1958. « Intervals between Confinements in Absence of Birth Control ». *Eugenics Quarterly*, vol. 5, p. 200-211.

Hosmer, David W. et Stanley Lemeshow. 1989. *Applied Logistic Regression*, New York : John Wiley and sons, 307 p.

Hugues, Al et M. Nei. 1989. « Nucleotide Substitution at Major Histocompatibility Complex Class II loci: Evidence for Overdominant Selection ». *Proc. Natl. Acad. Sci., USA*, vol. 86, p 958-962.

Hull Peter. 1969. « Maternal-Foetal Incompatibility Associated with the H-3 Locus in the Mouse ». *Heredity*, vol. 24, p 203-209.

Jacquard, Albert. 1974. *Génétiq ue des populations humaines*. P.U.F., Paris, p. 47.

Jakobi, L. et A. Jacquard. 1971. « Consanguinité proche, consanguinité éloignée, Essai de mesure dans un village breton ». *Génétiq ue et Populations*, INED, Travaux et Documents, no. 60, pp263-268, P.U.F.

Khlat, Myriam et Muin Houry. 1991. « Inbreeding and Diseases: Demographic, Genetic, and Epidemiologic Perspectives ». *Epidemiologic Reviews*, vol.13, p 28-40.

Kleimbaum, D.g., L. Kupper et K.A. Muller. 1987. « Applied Regression Analysis and other Multivariate Methods ». 2<sup>e</sup> édition, Boston, Duxbury Press, 282 p.

Kouladjian, Kevork. 1986. « Une mesure d'entropie généalogique ». Document III-C-43, Programme de recherche en génétique humaine, SOREP, 4 p.

Labelle, Éric .1996. « *Logiciel de gestion de registre de population humaine, d'analyse généalogique, démographique et génétique.* » Première version. EDYPH, Université du Québec à Montréal, 39 p.

- Laberge, Claude. 1967. « La consanguinité des Canadiens français ». *Populations*, vol. 22, no. 5, p.861-896.
- Lepage, Marie-Claire et Madeleine Levasseur. 1989. « Mortalité et morbidité périnatales et infantiles ». *La périnatalité au Québec 4*, Ministère de la Santé et des Services Sociaux, Québec, p. 43-47.
- Le Querrec, Jacques. 1981. « L'Île-aux-Coudres, vers un divorce ethno-écologique? ». *Anthropologie et Sociétés*, vol. 5, no. 1, p. 165-169.
- Liazoghli, Dalinda. 2000. *Les effets de l'apparentement biologique sur la performance reproductive des couples dans une population endogame humaine des Antilles françaises : existe-t-il une compensation reproductive?* Mémoire de maîtrise, Montréal, Université du Québec à Montréal, 80 p.
- Macklin, Madge T. 1951. « Sex Ratios in Partial Sex-Linkage Excess of Affected Females from Consanguineous Matings ». *American Journal of Human Genetics*, p. 14-30.
- Maheu, Sylvie. 1997. *Facteurs influençant la mortalité infantile à l'Île-aux-Coudres entre 1800 et 1940*. Rapport de statistiques, Montréal, Université du Québec à Montréal, 36 p.
- Malécot, G. 1948. *Les mathématiques de l'hérédité*. Paris, Masson.
- Marieb, Elaine N. *Anatomie et physiologie humaines*. 2<sup>e</sup> éd. Adaptation française de René Lachaine. Saint-Laurent: Éditions de Renouveau Pédagogique Inc. (ERPI), 1194p.

Martin, Yves. 1957. « L'Île-aux-Coudres: Population et économie », Cahiers de géographie 2, p. 167-195.

Nei M. et Al Hugues. 1991. « Polymorphism Due to Balancing Selection at the Major Histocompatibility Complex Loci », In EC Dudley (ed.): *The Unity of Evolutionary Biology*, p. 878-889. Portland, OR :Discorides Press.

Ober, Carole. 1995. « Current Topic : HLA and Reproduction : Lessons from Studies in the Hutterites ». *Placenta*, vol. 16, p. 569-577.

Ober, Carole et A. Aldrich. 1997. « HLA-G Polymorphism : Neutral Evolution or Novel Function? ». *Reprod. Evolution*, vol. 61, p.497-504.

Ober, Carole. 1998. « HLA and Pregnancy: The Paradox of the Fetal Allograph ». *American Journal of Human Genetics*, vol. 62, p 1-5.

Ober, Carole, Terry Hyslop, Sherman Elias, Lowell R. Weitkamp et Walter W. Hauck. 1998. « Human Leukocyte Antigen Matching and Fetal Loss : Results of a 10 Year Prospective Study ». *Human Reproduction*, vol. 13, no 1, p. 33-38.

Péron , Yves et Claude Strohmerger. 1985. *Indices démographiques et indicateurs de santé des populations, présentation et interprétation*. Ottawa, Publication autorisée par le Ministère des Approvisionnements et Services Canada, p. 197.

Philippe, Pierre. 1969. *Structure de la consanguinité à l'Île-aux-Coudres*, Mémoire de maîtrise, Montréal, Université du Québec à Montréal, 111 p.

Philippe, Pierre et Jacques Gomila. 1971. « Structure de population et mariages consanguins à l'Isle-aux-Coudres (Québec) », *Population*, vol. 4, p. 708-716.

Philippe, Pierre et Jacques Gomila. 1972. « Inbreeding Effects in a French Canadian Isolate I. Evolution of Inbreeding. ». *Z. Morph. Anthropol.*, vol. 64, no. 1, p. 54-59.

Philippe, Pierre. 1973. « Analyse statistique des intervalles protogénésiques et intergénéésiques à l'Isle-aux-Coudres, Étude de démographie historique » *Population*, vol. 1, p 81-93.

Pirot, Philippe. 1992. *Étude sur l'interaction entre la consanguinité et la mortalité infantile dans une population endogame de la Caraïbe : Saint-Barthélémy*. Rapport de stage, Montréal, Université du Québec à Montréal, 107 p.

Plassart, Gaël. 1997. *Rôle différentiel de la consanguinité proche et de la consanguinité éloignée sur l'évolution d'une population humaine des Caraïbes: Saint Barthélémy*. Rapport de stage, Montréal, Université du Québec à Montréal, 21 p.

Rochette, Louise, Denis Drouin et Geneviève Tremblay. 1986. « Pour mieux connaître la population. ». *DSC Hôpital Enfant Jésus*, 208 p.

Sall, John et Ann Lehman. 1996. *JMP Start Statistics, A Guide to Statistics and Data Analysis Using JMP and JMP In Software*. États-Unis : SAS Institute Inc., 521 p.

Shami, S.A., L.H. Schmitt et A.H. Bittles. 1989. « Consanguinity Related Prenatal and Postnatal Mortality of the Populations of Seven Pakistani Punjab Cities ». *Journal of Medicals Genetics*, vol. 26, p 267-271.

Shull, William J., H. Nagano, M. Yamamoto et I. Komatsu. 1970a. « The Effect of Parental Consanguinity and Inbreeding in Hirado, Japan. I. Stillbirth and Prereproductive Mortality ». *American Journal of Human Genetics*, vol. 22, p 239-262.

Shull, William J., T. Furosho, M. Yamamoto, H. Nagano et I. Komatsu. 1970b. « The Effect of Parental Consanguinity and Inbreeding in Hirado, Japan. IV. Fertility and Reproductive Compensation ». *Humangenetik*, vol. 9, p. 294-315.

Shull, William J. et James V. Neel. 1972. « The Effect of Parental Consanguinity and Inbreeding in Hirado, Japan. V. Summary and Interpretation ». *American Journal of Human Genetics*. vol. 24), p. 425-453.

SPSS Inc. 1997. « *SPSS BASE 7.5 for Windows User's Guide*. Chicago : SPSS Inc.

Stokes, M.E., C. S. Davis et G.G. Kosh. 1995. *Categorical Data Analysis Using the SAS System*. SAS Institute.

Sutter, Jean. 1968. « Fréquence de l'endogamie et ses facteurs au XIX<sup>e</sup> siècle », *Populations*, vol. 2, p. 303-324.

Vandeschrick, Christophe. 1995. *Analyse démographique*. Coll. « Population et développement », no 1. Préf. de Guillaume Wunsch. Louvain-la-Neuve : Bruylant-Academia, 182 p.

Wallace, Bruce. 1989. « One Selectionist's Perspective ». *The Quarterly Reviews of Biology*, vol. 64, no. 2, p. 127-143.

Wright, S. 1922. *Coefficients of Inbreeding and Relationship*, American Naturalist.