

INTRODUCTION AU CALCUL DES TAUX DEMOGRAPHIQUES DES CHEPTELS TROPICAUX

Matthieu Lesnoff
CIRAD (Département EMVT)
Détaché à l'ILRI - Nairobi

Juin 2005

TABLES DES MATIERES

Introduction.....	3
1. Contexte et objectif du document	4
2. Collectes des données sur le terrain	4
3. Liste des taux démographiques.....	5
4. Probabilités et risques instantanés.....	6
5. Calculs des taux en temps discret.....	9
Partie I : Définition et calcul des taux démographiques	10
1. Introduction.....	11
2. Taux de mort naturelle et d'exploitation.....	11
2.1. Probabilités brute.....	11
2.2. Risques instantanés	12
2.2.1. Définition.....	12
2.2.2. Approximation des risques instantanés	13
2.3. Relation entre les probabilités brutes et les risques instantanés.....	14
2.4. Interférences entre taux démographiques	14
2.5. Synthèse des estimations.....	17
3. Taux d'importation	18
4. Taux de reproduction.....	18
4.1. Taux de mise bas et d'avortement.....	18
4.1.1. Introduction	18
4.1.2. Le risque instantané de mise bas	19
4.1.3. Probabilité de mise bas et lien avec le risque instantané	19
4.1.4. Remarque sur l'intervalle moyen entre mises bas	22
4.2. Taux de prolificité et de mortinatalité	22
5. Indicateurs synthétiques de productivité numérique	23
Partie II : Les méthodes de groupements de données démographiques.....	24
1. Introduction.....	25
2. Décomposition de cohortes	25
2.1. Principe.....	25
2.2. Regroupement d'unités.....	27
2.3. Pseudo-cohortes.....	28
3. Autres décompositions	30
Partie III : Analyse statistique des taux démographiques.....	31
1. Introduction.....	32
2. Approximation normale.....	33
2.1. Intervalles de confiance de la probabilité et du risque instantané	33
2.2. Comparaison de deux probabilités	33
2.3. Comparaison de deux risques instantanés.....	34
3. Modèles linéaires généralisés	35
3.1. Définition.....	35
3.2. Régression logistique	36
3.3. Régression log-linéaire	36
4. Exemple numérique.....	37
4.1. Introduction.....	37
4.2. Analyse descriptive.....	37
4.3. Tests.....	38
4.3.1. Approximation normale.....	38
4.3.2. Modèles linéaires généralisés.....	39
Bibliographie.....	42

INTRODUCTION

1. CONTEXTE ET OBJECTIF DU DOCUMENT

Les cheptels domestiques représentent une valeur économique déterminante pour de nombreux pays tropicaux, aussi bien au niveau national qu'au niveau des ménages. Disposer de méthodes et d'indicateurs standardisés pour mieux estimer l'impact d'interventions zootechniques, sanitaires, économiques ou politiques sur la dynamique et la productivité et la dynamique des cheptels permet de mieux cibler les stratégies efficaces pouvant améliorer la situation économique des pays et le bien-être des ménages agricoles.

Les taux démographiques, comme par exemple les taux annuels de reproduction ou de mortalité, font partie de ces indicateurs et synthétisent l'effet des caractéristiques biologiques des animaux aussi bien que des pratiques de gestion mises en œuvre par les éleveurs.

Ils représentent également les données de base nécessaires des modèles mathématiques simulant la dynamique des cheptels dans différents contextes (Itty, 1995; Lesnoff, 1999; Tacher, 1975a, b; Texeira and Paruelo, 2005) et permettant de calculer *ex post* ou *ex ante* les impacts d'interventions sur leur productivité globale (Baptist, 1992; Hary, 2004; Lesnoff et al., 2000; Lesnoff et al., 2001; Upton, 1989).

De nombreuses définitions de taux démographiques ont été présentées en élevage tropical (par exemple CIRAD-IEMVT, 1972; IEMVT, 1989; ILCA, 1990; Landais and Sissokho, 1986; Lhoste et al., 1993; Mukasa-Mugerwa, 1989; Wilson, 1986). Cependant, les bases théoriques du calcul des taux et les possibles écueils méthodologiques ont encore été très peu détaillés dans la littérature. Par ailleurs, la multiplicité des définitions des taux et l'hétérogénéité des termes employés pour ces paramètres constituent un problème important pour les utilisateurs potentiels.

Dans ce contexte, le présent document a pour objectif de présenter les principaux concepts sous-jacents à la définition et au calcul des taux démographiques des cheptels tropicaux.

Le document est centré sur le cas des systèmes d'élevage tropicaux traditionnels, extensifs ou semi-extensifs, et dans lesquels la reproduction n'est pas ou peu contrôlée. Certains des concepts présentés peuvent ne pas être pertinents lorsque la reproduction est totalement contrôlée ou lorsque les animaux sont gérés par lots, comme dans les stations expérimentales, certains ranchs ou en élevage intensif.

2. COLLECTES DES DONNEES SUR LE TERRAIN

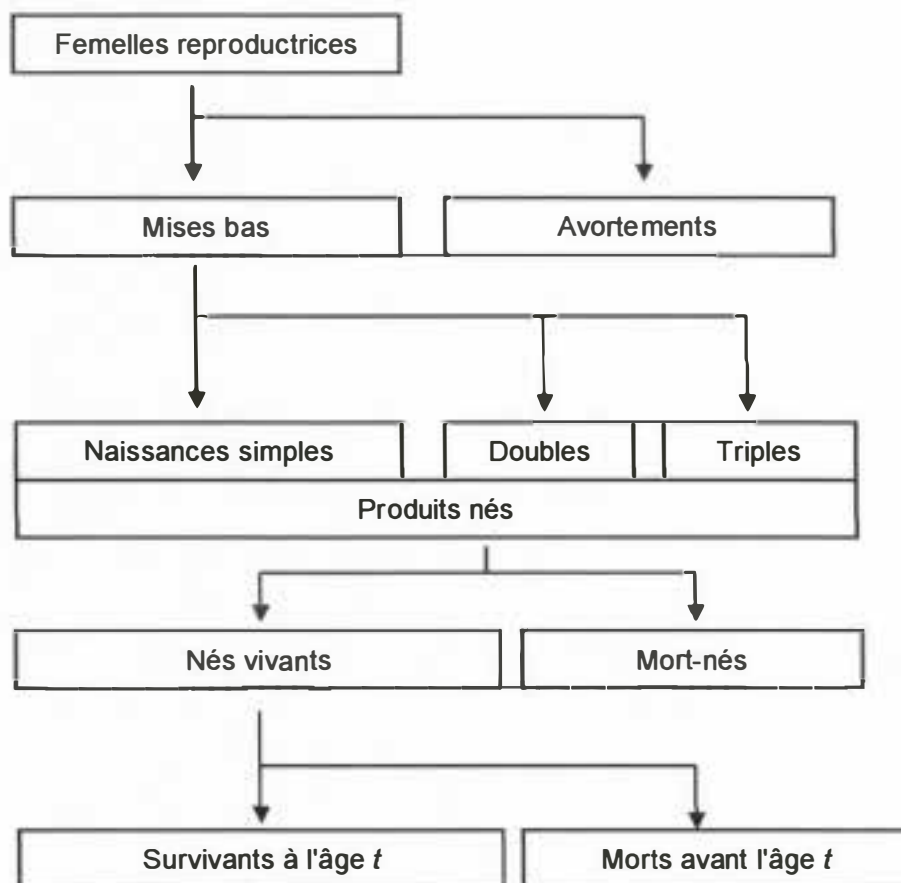
Les taux démographiques doivent être estimés à partir de données collectées sur le terrain. Cette collecte constitue en général la part la plus difficile du travail, en raison d'une part de la dispersion et du déplacement des troupeaux et d'autre part du fait qu'il n'existe pas dans les fermes de registres écrits décrivant la démographie des troupeaux. Deux types de méthodes d'enquête sont habituellement rencontrés dans la littérature. Le premier type correspond aux suivis de troupeaux (Faugère and Faugère, 1986), qui représentent actuellement la méthode "gold-standard" parmi les méthodes de terrain en milieu tropical. Ils consistent à suivre un même échantillon de troupeaux pendant une ou plusieurs années successives. Les animaux des troupeaux échantillonnés sont identifiés individuellement, par exemple avec des boucles auriculaires. Des enquêteurs visitent régulièrement les troupeaux et notent tous les événements démographiques ayant eu lieu entre deux visites successives. Les suivis de troupeaux engendrent des données précises et fiables mais sont difficiles à étendre à grande échelle. Le second type de méthode correspond aux enquêtes rétrospectives faisant appel à la mémoire

des éleveurs et concernant des périodes pouvant remonter sur plusieurs années dans le temps. Plusieurs techniques d'enquêtes sont disponibles (CIRAD-IEMVT, 1989; ILCA, 1990; SEDES, 1975). Ces enquêtes, plus rapides à mettre en œuvre que les suivis de troupeaux, peuvent être reproduites à grandes échelles mais engendrent des données moins fiables. Les concepts décrits dans le présent document peuvent être utiles dans le cadre de ces deux types de méthodes d'enquête.

3. LISTE DES TAUX DEMOGRAPHIQUES

La liste des taux considérés dans le présent document a été établie dans l'objectif de disposer d'un ensemble de paramètres complémentaires pouvant servir de base à la quantification et la modélisation de la dynamique et de la productivité numérique des cheptels. Cette liste n'est bien-sûr pas exhaustive et peut être adaptée en fonction des objectifs spécifiques de l'étude.

La liste considérée s'appuie sur la séquence démographique proposée par Landais et Sissokho, dont une version simplifiée est représentée ci-dessous :



Les taux sont les suivants :

Taux de performance zootechnique	
Taux de mise bas	Probabilité ou risque instantané qu'une femelle mette bas durant une période de temps
Taux d'avortement	Probabilité ou risque instantané qu'une femelle avorte durant une période de temps (un avortement est une gestation qui n'a pas été à son terme)
Taux de prolificité	Effectif moyen de produits nés (vivants ou morts) par mise bas
Taux de mortinatalité	Probabilité qu'un produit né soit un mort-né (la mortinatalité n'est pas comptabilisée dans la mortalité naturelle, qui concerne uniquement les produits nés vivants)
Taux de mort naturelle	Probabilité ou risque instantané qu'un animal décède de mort naturelle durant une période de temps

Taux de gestion démographique	
Taux d'exploitation	Probabilité ou risque instantané qu'un animal soit exploité (abattage, vente, prêt, don, etc.) durant une période de temps
Taux d'importation	Probabilité ou risque instantané qu'un animal soit importé (achat, prêt, don, etc.) durant une période de temps

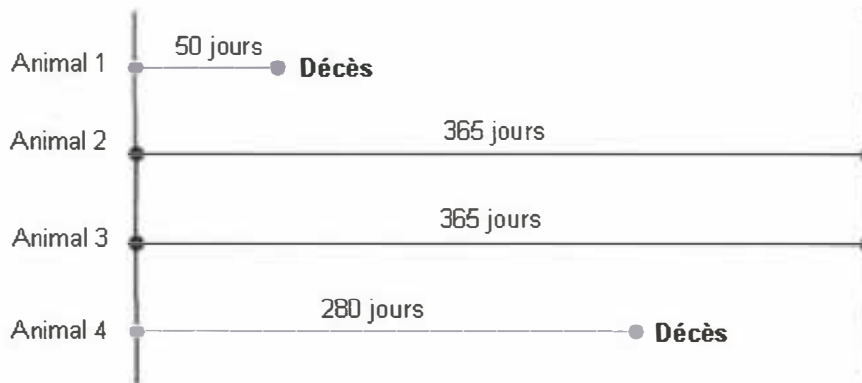
4. PROBABILITES ET RISQUES INSTANTANES

Très utilisé dans la littérature et dans le langage courant, le terme "taux démographique" est cependant ambigu car il peut représenter des paramètres mathématiques distincts. Ceci a été décrit dans certains ouvrages épidémiologiques de référence (par exemple Cox and Oakes, 1984; Kalbfleisch and Prentice, 1980; Lee, 1992) mais très peu dans le domaine de l'élevage tropical et, dans ce cas, toujours dans des documents spécialisés (Landais and Sissokho, 1986; Lesnoff, 2000). Par exemple, le manuel ILCA sur les méthodes de recherche dans les systèmes tropicaux (ILCA, 1990, chap. 5) ne différencie pas clairement les différents types de paramètres.

En particulier, selon leur mode de calcul, certains taux présentés dans la liste précédente (mise bas, avortement, mort naturelle, exploitation et importation) peuvent correspondre soit à une probabilité soit à un risque instantané. Ces deux paramètres ont des significations et des interprétations différentes. La partie I du présent document détaille les principaux points méthodologiques clés sur les probabilités et les risques instantanés, dont une synthèse est la suivante :

a) Pour un groupe d'animaux et un type d'événement démographique donnés, la probabilité d'occurrence (π) et le risque instantané (μ) de cet événement durant une période de temps peuvent être estimées par le rapport du nombre d'événements observés durant la période sur respectivement le nombre d'animaux à risque en début de période et le "temps total à risque", c'est-à-dire le temps de présence des animaux cumulé sur l'ensemble de la période. Prenons

l'exemple de quatre animaux suivis pendant une année et dont deux décèdent en cours d'année :



La probabilité et le risque instantané de mort peuvent être estimés par :

$$\hat{\pi} = \frac{2}{4} = 0.50 = 50 \%,$$

$$\hat{\mu} = \frac{2}{(50 + 365 + 365 + 280)/365} = 0.69 \text{ année}^{-1}.$$

b) Une probabilité prend sa valeur entre 0 et 1 et, bien qu'elle n'ait pas d'unité, peut être exprimée en pourcentage. Un risque instantané peut être supérieur à 1 et devrait être exprimé en unité temps⁻¹ et non en pourcentage.

c) Dans certaines enquêtes, le temps total à risque n'est pas connu et doit être approché. Une approximation possible est le nombre moyen d'animaux présents dans la période. Le rapport du nombre d'événements à la "taille moyenne du troupeau sur l'année", souvent utilisé, correspond à un risque instantané et non à une probabilité comme cela est parfois interprété dans la littérature. Dans l'exemple précédent, cette approximation donne :

$$\hat{\mu} = \frac{2}{4 + \frac{2}{2}} = \frac{2}{3} = 0.67 \text{ année}^{-1}.$$

d) Etant donnée leur nature différente, les probabilités et les risques instantanés ne sont pas analysés avec les mêmes méthodes statistiques. Par exemple, si des modèles de régression sont utilisés, les probabilités doivent être analysées à l'aide de régressions binomiales et les risques instantanés à l'aide de régressions de Poisson (Agresti, 1990).

e) Dans certains cas, les probabilités et les risques instantanés sont liés par des relations fonctionnelles. Par exemple, sous l'hypothèse qu'il n'y ait pas de risques compétitifs (cf. point f), la probabilité et le risque instantané de mort sont liés par la formule

$$\pi = 1 - \exp(-\mu \Delta t),$$

où Δt est la durée de la période considérée. Par exemple, un risque instantané de mort de 0.30 année⁻¹ correspond à une probabilité annuelle de mort de 0.26, ce qui signifie que 26 % des animaux vont mourir en moyenne dans l'année et non 30 %.

f) Le lien entre les probabilités et les risques instantanés de mise bas est plus difficile à représenter car d'une part les femelles ne disparaissent pas du cheptel après les mises bas et d'autre part les femelles de certaines espèces peuvent avoir plusieurs mises bas dans l'année.

g) Les taux démographiques sont soumis au problème des risques compétitifs, bien décrits en épidémiologie humaine (Prentice, 1978) ou en écologie (Anderson and Burnham, 1976; Burnham and Anderson, 1984), mais qui reste peu pris en compte dans l'analyse démographique des cheptels tropicaux (Landais, 1986; Lesnoff, 2000). Deux événements démographiques sont dits compétitifs si l'occurrence de l'un empêche la future occurrence de l'autre. Deux événements compétitifs typiques dans les cheptels sont la mort naturelle et l'exploitation des animaux, mais les autres types d'événements sont aussi impliqués (par exemple la mort naturelle et l'exploitation sont compétitives pour les mises bas). L'une des conséquences de la compétition entre la mortalité et l'exploitation est que la probabilité brute de mort, calculée par le rapport du nombre de décès observés sur le nombre d'animaux à risque, dépend directement du niveau d'exploitation dans le cheptel. Plus le niveau d'exploitation est élevé, plus la probabilité brute de mort diminue. Ceci peut rendre trompeurs certains résultats bruts. Le tableau ci-dessous représente un exemple de résultat expérimental où la diminution de mortalité observée parmi les troupeaux ayant reçu une intervention a été engendrée par une exploitation des animaux plus élevée et non par l'effet de l'intervention elle-même :

	Probabilité brute de mort naturelle (%)	Probabilité brute d'exploitation (%)
Troupeaux contrôles	25	5
Troupeaux avec intervention	20	35

Ce phénomène est un exemple d'interférence entre les taux démographiques. Les interférences concernent tous les paramètres de performance et sont d'autant plus importantes que la période de temps considérée pour le calcul des taux est grande ou que les risques instantanés sont élevés. Des approches spécifiques doivent être mises en œuvre pour contrôler cet effet (Lesnoff, 2000).

5. CALCULS DES TAUX EN TEMPS DISCRET

Les concepts décrits dans le présent document sont basés sur une approche de calcul "en temps discret", c'est-à-dire qu'avant les calculs l'âge des animaux et le temps sont décomposés en une succession de classes d'âge et de périodes de temps. Les données démographiques sont ensuite groupées pour chacune de ces classes d'âge et périodes.

Par exemple, pour une classe d'âge et une période de temps, les données nécessaires pour calculer les taux présentés dans la liste précédente (excepté le taux de prolificité qui est de nature différente des autres taux) sont constituées :

- du nombre d'événements m survenus dans la classe d'âge et la période de temps,
- de l'effectif initial n , c'est-à-dire de l'effectif des animaux présents en début de classe d'âge et de période de temps initiaux,
- du temps total de présence T (appelé temps à risque) de ces animaux dans la classe d'âge et la période de temps.

Différentes procédures sont possibles pour effectuer le groupement des données démographiques. Ces procédures sont décrites dans la partie II du présent document.

La partie III du document introduit l'analyse statistique des taux démographiques. Les méthodes statistiques relatives à l'approche en temps discret correspondent aux méthodes d'analyse des données de comptage, présentées dans de nombreux ouvrages de référence (par exemple Agresti, 1990; Collett, 2003; Kalbfleisch and Prentice, 1980). D'autres approches sont possibles, notamment celles qui considèrent l'âge et le temps comme des variables continues. Ces approches, qui font appel à des méthodes statistiques comme la méthode de Kaplan-Meier ou le modèle de Cox et ses extensions, ne sont pas abordées dans le présent document.

PARTIE I :

**DEFINITION ET CALCUL DES
TAUX DEMOGRAPHIQUES**

1. INTRODUCTION

Dans cette partie, nous introduisons quelques définitions générales (probabilités, risques instantanés et interférences) en prenant l'exemple des taux de mort naturelle et d'exploitation. Les autres taux sont ensuite définis. Des éléments plus théoriques sur les taux peuvent également être trouvés dans Lesnoff (2000).

2. TAUX DE MORT NATURELLE ET D'EXPLOITATION

Les définitions sont illustrées à partir d'un exemple numérique concernant l'étude de la mortalité naturelle de jeunes ovins de la naissance à l'âge d'un an (365 jours). Dans cet exemple, les animaux sont supposés soit rester dans le troupeau soit disparaître du fait d'un décès ou d'une exploitation. Les données sont les suivantes :

Type de donnée	Notation	Valeur
Effectif des animaux suivis	n	697
Effectif des décès (mort naturelle) observés avant l'âge d'1 an parmi les n animaux	m_{mort}	156
Effectif des exploitations observées avant l'âge d'1 an parmi les n animaux (les exploitations représentent ici des censures pour l'étude de la mortalité naturelle)	m_{expl}	255
Effectif des animaux " survivants " à l'âge d'1 an (animaux ayant échappé à la mortalité naturelle et l'exploitation)	m_{surv}	286
Temps total à risque entre les âges de 0 et 1 an (exprimé en année)	T	485.856

2.1. Probabilités brute

La probabilité brute d'un événement correspond à la proportion d'animaux touchés par cet événement. Nous notons respectivement π_{mort} , π_{cens} et π_{surv} les probabilités brutes de mort naturelle, d'exploitation et de survie dans la période de temps considérée :

$$\pi_{\text{mort}} + \pi_{\text{expl}} + \pi_{\text{surv}} = 1.$$

Les probabilités brutes des jeunes ovins entre 0 et 1 an peuvent être estimées par :

$$\hat{\pi}_{\text{mort}} = \frac{m_{\text{mort}}}{n} = \frac{156}{697} = 22.4 \%$$

$$\hat{\pi}_{\text{expl}} = \frac{m_{\text{expl}}}{n} = \frac{255}{697} = 36.6 \%$$

$$\hat{\pi}_{\text{surv}} = \frac{m_{\text{surv}}}{n} = \frac{286}{697} = 41.0 \%$$

Les probabilités brutes sont parfois appelées quotients démographiques dans la littérature francophone (Landais and Sissokho, 1986).

2.2. Risques instantanés

2.2.1. Définition

Soit P la probabilité d'occurrence d'un événement démographique dans le petit intervalle $[t, t + dt)$ pour les animaux ayant survécu jusqu'à t . Le risque instantané correspondant à cet événement à l'âge exact t est la limite du rapport P/dt quand dt tend vers 0.

Un risque instantané est un rapport entre une probabilité (sans unité) et une durée. Il s'exprime donc en unité temps^{-1} , et non en pourcentage comme souvent dans le langage courant.

Nous notons respectivement $\mu_{\text{mort}}(t)$ et $\mu_{\text{expl}}(t)$ les risques instantanés de mort naturelle et d'exploitation à l'âge exact t . Le taux total de disparition du troupeau à l'âge t est :

$$\mu_{\text{tot}}(t) = \mu_{\text{mort}}(t) + \mu_{\text{expl}}(t).$$

Dans notre exemple, les risques instantanés sont supposés constants entre 0 et 1 an. Ils peuvent alors être estimés par :

$$\hat{\mu}_{\text{mort}} = \frac{m_{\text{mort}}}{T} = \frac{156}{485.856} = 0.321 \text{ année}^{-1},$$

$$\hat{\mu}_{\text{expl}} = \frac{m_{\text{expl}}}{T} = \frac{255}{485.856} = 0.525 \text{ année}^{-1},$$

$$\hat{\mu}_{\text{tot}} = \hat{\mu}_{\text{mort}} + \hat{\mu}_{\text{expl}} = 0.321 + 0.525 = 0.846 \text{ année}^{-1}.$$

Un même risque instantané peut être exprimé en différentes unités temps^{-1} , par exemple :

$$\hat{\mu}_{\text{mort}} = 0.321 \text{ année}^{-1},$$

$$= \frac{0.321}{12} = 0.02675 \text{ mois}^{-1},$$

$$= \frac{0.321}{365} = 0.0008795 \text{ jour}^{-1}.$$

2.2.2. Approximation des risques instantanés

Lorsque le temps total à risque T n'est pas connu, les risques instantanés peuvent être approchés en supposant une distribution uniforme des disparitions d'animaux dans la période de temps (un exemple est donné dans Holford, 1976). Sous cette hypothèse, le temps de présence dans la période d'un animal disparu du troupeau (mort ou exploité) est en moyenne une demi-période. Le temps de présence de l'ensemble des n animaux est donc :

$$\begin{aligned} T &= (n - m_{\text{mort}} - m_{\text{expl}}) \times 1 + (m_{\text{mort}} + m_{\text{expl}}) \times \frac{1}{2}, \\ &= n - \frac{m_{\text{mort}} + m_{\text{expl}}}{2}. \end{aligned}$$

Le risque instantané de mort naturelle est ainsi estimé par :

$$\hat{\mu}_{\text{mort}} = \frac{m_{\text{mort}}}{T} = \frac{m_{\text{mort}}}{n - \frac{m_{\text{mort}} + m_{\text{expl}}}{2}},$$

ce qui donne dans notre exemple :

$$\begin{aligned} \hat{\mu}_{\text{mort}} &= \frac{156}{697 - \frac{156 + 255}{2}} = \frac{156}{491.5}, \\ &= 0.317 \text{ année}^{-1}. \end{aligned}$$

L'approximation du temps total à risque T peut aussi s'exprimer en fonction des effectifs des animaux en début et en fin de période, notés respectivement $n_{\text{initial}} = n$ et $n_{\text{final}} = n - m_{\text{mort}} - m_{\text{expl}}$:

$$\begin{aligned} T &= n - \frac{m_{\text{mort}} + m_{\text{expl}}}{2}, \\ &= \frac{2n - (m_{\text{mort}} + m_{\text{expl}})}{2}, \\ &= \frac{n + (n - m_{\text{mort}} - m_{\text{expl}})}{2}, \\ &= \frac{n_{\text{initial}} + n_{\text{final}}}{2}, \end{aligned}$$

d'où se déduit une expression classique :

$$\hat{\mu}_{\text{mort}} = \frac{m_{\text{mort}}}{\frac{n_{\text{initial}} + n_{\text{final}}}{2}},$$

2.3. Relation entre les probabilités brutes et les risques instantanés

En supposant les risques instantanés constants dans la période de temps, les relations entre les probabilités brutes et les risques instantanés sont les suivantes (Anderson and Burnham, 1976)

$$\pi_{\text{mort}} = \frac{\mu_{\text{mort}}}{\mu_{\text{tot}}} [1 - \exp(-\mu_{\text{tot}})],$$

$$\pi_{\text{expl}} = \frac{\mu_{\text{expl}}}{\mu_{\text{tot}}} [1 - \exp(-\mu_{\text{tot}})].$$

Ces relations peuvent être utilisées pour estimer les probabilités brutes à partir des risques instantanés, par exemple pour la mortalité naturelle :

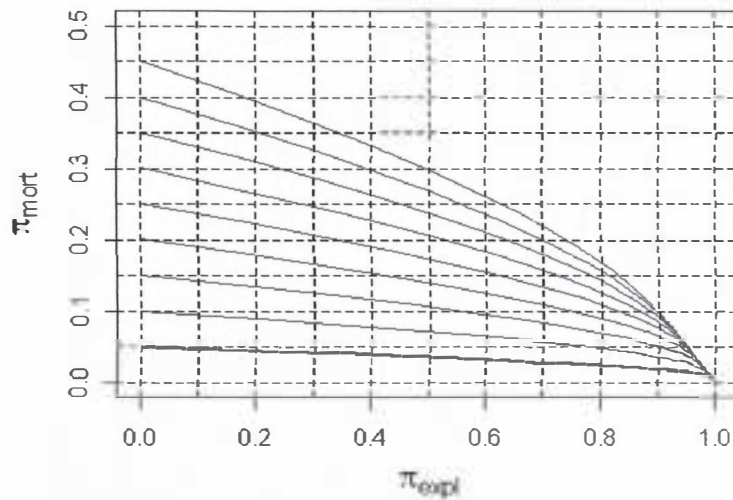
$$\hat{\pi}_{\text{mort}} = \frac{\hat{\mu}_{\text{mort}}}{\hat{\mu}_{\text{tot}}} [1 - \exp(-\hat{\mu}_{\text{tot}})].$$

Dans notre exemple, cela donne :

$$\begin{aligned} \hat{\pi}_{\text{mort}} &= \frac{\hat{\mu}_{\text{mort}}}{\hat{\mu}_{\text{tot}}} [1 - \exp(-\hat{\mu}_{\text{tot}})], \\ &= \frac{0.321}{0.321 + 0.525} [1 - \exp(-(0.321 + 0.525))], \\ &= 21.7 \%. \end{aligned}$$

2.4. Interférences entre taux démographiques

La probabilité brute π_{mort} n'est pas toujours un bon indicateur de la mortalité naturelle car elle correspond à une mortalité naturelle apparente, qui diminue lorsque l'exploitation des animaux augmente :



La théorie des risques compétitifs permet d'expliquer ce phénomène d'interférence. Dans la période de temps considérée, l'effectif des animaux exposés au risque instantané μ_{mort} va diminuer à chaque instant de la fraction exploitée. Dans un cheptel exploité, l'effectif de décès observés sera donc plus faible que s'il n'y avait pas eu d'exploitation et la probabilité brute π_{mort} sous-estimera le niveau réel de mortalité naturelle.

Un indicateur préférable à π_{mort} est la probabilité corrigée $\pi_{\text{mort}, c}$, quantifiant la mortalité qui serait observée s'il n'y avait pas d'exploitation. Nous présentons trois approches pour calculer $\pi_{\text{mort}, c}$.

Approche n°1

Cette approche utilise la relation entre les probabilités brutes et les risques instantanés :

$$\left. \begin{aligned} \pi_{\text{mort}} &= \frac{\mu_{\text{mort}}}{\mu_{\text{mort}} + \mu_{\text{expl}}} \left(1 - \exp(-\mu_{\text{mort}} + \mu_{\text{expl}}) \right) \\ \mu_{\text{expl}} &= 0 \end{aligned} \right\} \Rightarrow \pi_{\text{mort}, c} = 1 - \exp(-\mu_{\text{mort}}).$$

Dans notre exemple, la probabilité corrigée est :

$$\begin{aligned} \hat{\pi}_{\text{mort}, c} &= 1 - \exp(-\hat{\mu}_{\text{mort}}), \\ &= 1 - \exp(-0.321), \\ &= 27.5 \%. \end{aligned}$$

Approche n°2

Cette approche utilise la méthode des tables de vie (décrite par exemple dans Kalbfleisch and Prentice, 1980) dont le principe est de corriger l'effectif d'animaux à risque. Comme pour l'approximation du temps à risque, cette méthode suppose une distribution uniforme des disparitions d'animaux au cours de la période de temps. Le temps de présence d'un animal disparu du troupeau (mort ou exploité) est alors en moyenne une demi-période et la méthode donne un poids 1/2 à cet animal dans l'effectif à risque. L'effectif à risque corrigé est donc :

$$n_c = n - \frac{m_{\text{expl}}}{2}$$

Dans notre exemple, la probabilité corrigée est :

$$\begin{aligned}\hat{\pi}_{\text{mort}, c} &= \frac{m_{\text{mort}}}{n_c} \\ &= \frac{156}{697 - \frac{255}{2}} = \frac{156}{569.5} \\ &= 27.4\%.\end{aligned}$$

Approche n°3

Les interférences deviennent négligeables lorsque les probabilités brutes ont des valeurs faibles (dans ce cas, les probabilités brutes et les risques instantanés ont des valeurs très proches), ce qui survient notamment lorsque la période de temps considérée pour le calcul des taux est petite (par exemple ≤ 1 mois). L'approche n°3 consiste donc à décomposer la période de temps en plusieurs petites périodes successives (pour lesquelles les interférences sont négligeables), de calculer les probabilités brutes à l'échelle de ces petites périodes, puis de reconstituer la probabilité correspondant à la période initiale.

Dans notre exemple, l'approche peut consister à décomposer la période 0-1 an en 12 périodes mensuelles. Le cumul de tous les effectifs initiaux correspondant aux 12 périodes mensuelles a été de 6097 animaux, et l'effectif total des décès a été de 156. La probabilité brute moyenne mensuelle est donc :

$$\hat{\pi}_{\text{mort}, 1 \text{ mois}} = \frac{156}{6097} = 2.56 \%,$$

et la probabilité annuelle corrigée peut être estimée par :

$$\begin{aligned}\hat{\pi}_{\text{mort}, c} &= 1 - \left(1 - \hat{\pi}_{\text{mort}, 1 \text{ mois}}\right)^{12}, \\ &= 1 - (1 - 0.0256)^{12}, \\ &= 26.7 \%.\end{aligned}$$

Comme indiqué précédemment, lorsque la période de temps est petite (ici le mois), la probabilité brute se rapproche de la valeur du risque instantané :

$$\left\{ \begin{array}{l} \pi_{\text{mort, 1 mois}} = 0.0256, \\ \hat{\mu}_{\text{mort}} \text{ (en mois}^{-1}\text{)} = \frac{0.321}{12} = 0.0268. \end{array} \right.$$

2.5. Synthèse des estimations

Les différentes estimations de la mortalité naturelle obtenues dans l'exemple numérique sont synthétisées ci-dessous :

Mortalité naturelle	Valeurs
Probabilités brutes	
$\hat{\pi}_{\text{mort}} = \frac{m_{\text{mort}}}{n}$	22.4 %
$\hat{\pi}_{\text{mort}} = \frac{\hat{\mu}_{\text{mort}}}{\hat{\mu}_{\text{tot}}} \left(1 - \exp(-\hat{\mu}_{\text{tot}}) \right)$	21.7 %
Risques instantanés	
$\hat{\mu}_{\text{mort}} = \frac{m_{\text{mort}}}{T}$	0.321 année ⁻¹
$\hat{\mu}_{\text{mort}} = \frac{m_{\text{mort}}}{n - \frac{m_{\text{mort}} + m_{\text{expl}}}{2}} = \frac{m_{\text{mort}}}{\frac{n_{\text{initial}} + n_{\text{final}}}{2}}$	0.317 année ⁻¹
Probabilités corrigées	
$\hat{\pi}_{\text{mort, c}} = 1 - \exp(-\hat{\mu}_{\text{mort}})$	27.5 %
$\hat{\pi}_{\text{mort, c}} = \frac{m_{\text{mort}}}{n_{\text{c}}} = \frac{n}{n - \frac{m_{\text{expl}}}{2}}$	27.4 %
$\hat{\pi}_{\text{mort, c}} = 1 - \left(1 - \hat{\pi}_{\text{mort, 1 mois}} \right)^{12}$	26.7 %

3. TAUX D'IMPORTATION

Les animaux importés au cours d'une période donnée dans un troupeau enquêté peuvent provenir de deux origines, d'une part les troupeaux enquêtés et d'autre part les autres troupeaux du cheptel (qui représentent la part la plus importante). Il n'est ainsi pas rigoureusement correct de calculer une probabilité ou un risque instantané d'importation à partir de l'effectif n ou du temps à risque T , qui ne correspondent qu'aux troupeaux enquêtés.

Cependant, pour rester homogène avec les autres taux, le même type de définition peut être retenu. Le risque instantané d'importation, ici défini par rapport aux animaux présents dans les troupeaux enquêtés et non par rapport à la population à risque totale qui est inconnue, peut par exemple être estimé par :

$$\hat{\mu}_{\text{imp}} = \frac{m_{\text{imp}}}{T},$$

où m_{imp} représente le nombre d'animaux importés dans les troupeaux enquêtés et T le temps total à risque calculé pour les troupeaux enquêtés.

Ces définitions ont une justification théorique sous certaines hypothèses démographiques (Lesnoff, 1999 #117; Lesnoff, 2000 #338, p. 57).

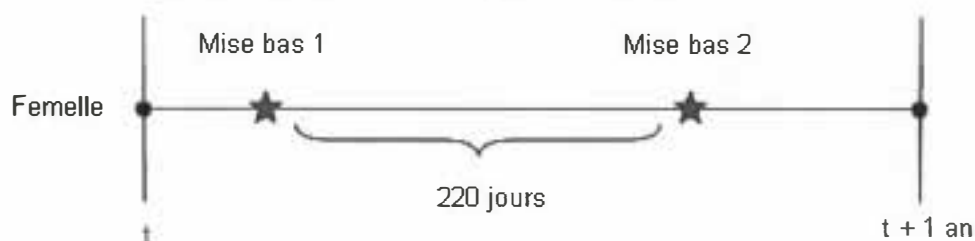
4. TAUX DE REPRODUCTION

4.1. Taux de mise bas et d'avortement

4.1.1. Introduction

Les définitions des taux de mise bas et d'avortement sont identiques, excepté l'événement pris en compte (mise bas *versus* avortement). Seul le taux de mise bas est donc considéré ci-dessous.

Le taux de mise bas se différencie des taux de mort naturelle et d'exploitation du fait que la femelle ne disparaît pas après la mise bas. Deux mises bas peuvent donc être observées chez la même femelle si la période de temps considérée est suffisamment longue, par exemple une année (ceci est en particulier fréquent chez les petits ruminants) :



Cette situation pose problème pour le calcul d'une probabilité car le rapport $m_{\text{mise bas}} / n$ peut être supérieur à 1. Par exemple, dans un troupeau où 25 % des femelles n'ont pas eu de mise bas au cours de l'année, 25 % ont eu une mise bas et les 50 % restant ont eu deux mises bas, le rapport $m_{\text{mise bas}} / n$ donne :

$$\frac{25 \times 0 + 25 \times 1 + 50 \times 2}{100} = 1.25,$$

ce qui n'a pas de sens pour une probabilité.

Les définitions du taux de mise bas doivent tenir compte de cette particularité. Dans ces définitions, nous supposons qu'au sein des femelles présentes les femelles vides ne peuvent être différenciées des femelles en cours de gestation, ce qui est le cas général sur le terrain lorsque la reproduction n'est pas contrôlée.

4.1.2. Le risque instantané de mise bas

Le risque instantané de mise bas est défini en supposant que, pour chaque femelle présente et chaque jour, l'occurrence d'une mise bas est engendrée par une loi de Bernoulli de faible probabilité et que ces lois sont indépendantes et identiques pour toutes les femelles et tous les jours de la période de temps. L'effectif de mises bas observées dans la période de temps suit alors une loi de Poisson de moyenne :

$$\lambda = \mu_{\text{mise bas}} \times T,$$

où $\mu_{\text{mise bas}}$ représente le risque instantané de mise bas et T le temps total de présence des femelles dans la période considérée, incluant les jours postérieurs aux mises bas éventuelles. Le risque de mise bas peut être estimé par :

$$\hat{\mu}_{\text{mise bas}} = \frac{m_{\text{mise bas}}}{T}$$

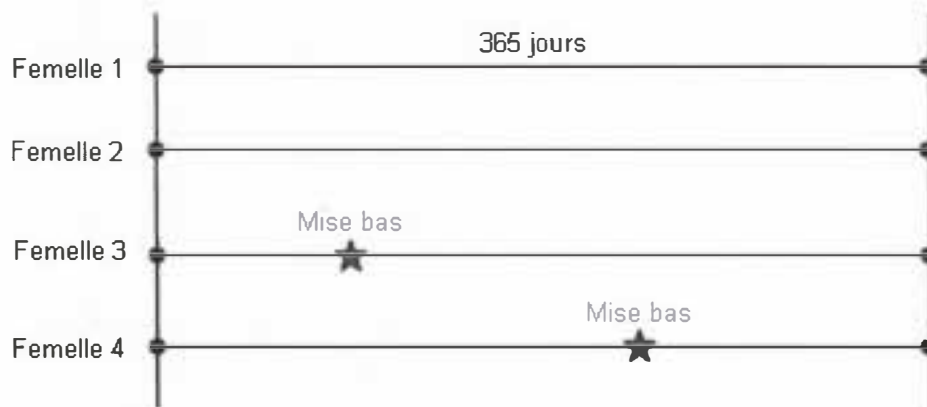
où $m_{\text{mise bas}}$ représente le nombre total de mises bas observées dans la période de temps.

4.1.3. Probabilité de mise bas et lien avec le risque instantané

Nous présentons tout d'abord la situation simplifiée où il n'y a pas de censure (c'est-à-dire où il n'y a aucune disparition d'animaux durant la période considérée) puis la situation plus réaliste avec censures.

Situation sans censure

Considérons le cas où chaque femelle est supposée ne pouvoir mettre bas qu'au plus une fois dans la période de temps considérée. Prenons l'exemple de quatre femelles suivies durant une année, dont deux ont mis bas en court d'année :



Dans cet exemple, la probabilité annuelle de mise bas peut être estimée par :

$$\begin{aligned}\hat{\pi}_{\text{mise bas}} &= \frac{m_{\text{mise bas}}}{n} \\ &= \frac{2}{4} = 0.50 = 50 \text{ \%}.\end{aligned}$$

Dans la situation sans censure, le temps total de présence des femelles dans la période, s'il est mesuré en unité "période", est égal à l'effectif initial des femelles :

$$T = n,$$

et donc :

$$\hat{\pi}_{\text{mise bas}} = \frac{m_{\text{mise bas}}}{n} = \frac{m_{\text{mise bas}}}{T} = \hat{\mu}_{\text{mise bas}}.$$

Dans l'exemple, le risque instantané est en effet :

$$\hat{\mu}_{\text{mise bas}} = \frac{2}{4} = 0.50 \text{ année}^{-1}.$$

Dans le cas où une femelle peut avoir plusieurs mises bas dans la période de temps considérée, la probabilité $\pi_{\text{mise bas}}$ n'a plus de sens, à la différence de $\mu_{\text{mise bas}}$. Cependant, il est toujours possible de définir des sous-périodes assurant la condition "au plus une mise bas par femelle" et de calculer les probabilités de mise bas correspondant à ces sous périodes à partir du risque instantané de mise bas calculé sur la période globale. Par exemple, même si les femelles d'une espèce peuvent avoir deux mises bas par an, on peut calculer une probabilité de mise bas mensuelle à partir du risque instantané annuel :

$$\left. \begin{aligned} \hat{\pi}_{\text{mise bas, 1 mois}} &= \hat{\mu}_{\text{mise bas, 1 mois}} \\ \hat{\mu}_{\text{mise bas, 1 mois}} &= \frac{\hat{\mu}_{\text{mise bas}}}{12} \end{aligned} \right\} \Rightarrow \hat{\pi}_{\text{mise bas, 1 mois}} = \frac{\hat{\mu}_{\text{mise bas}}}{12}$$

Par exemple, si le taux de mise bas annuel est :

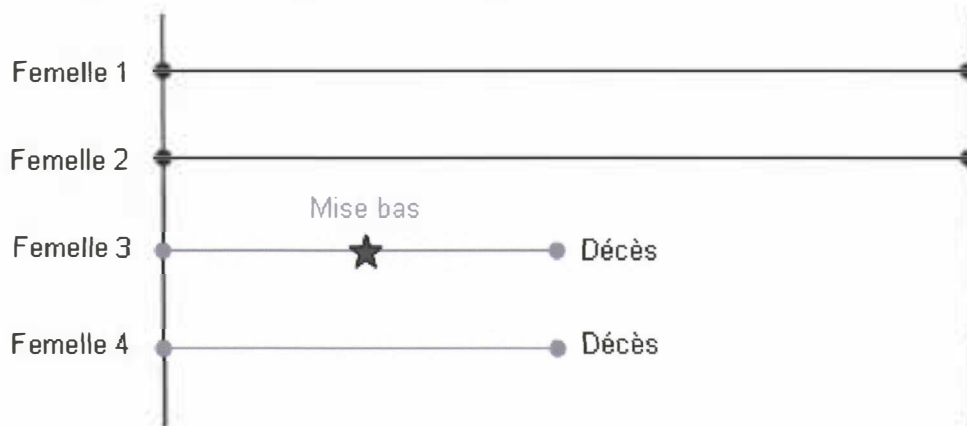
$$\hat{\mu}_{\text{mise bas}} = 1.34 \text{ année}^{-1},$$

la probabilité de mise bas mensuelle (par rapport a toutes les femelles présentes au début de chaque mois, qu'elles soient vides ou pleines) est :

$$\hat{\pi}_{\text{mise bas, 1 mois}} = \frac{1.34}{12} = 0.11 = 11 \%$$

Situation avec censures

Nous reprenons l'exemple précédent en supposant maintenant que les femelles n°3 et n°4 ont été censurées pour cause de décès en milieu d'année :



Comme pour la mortalité naturelle, la probabilité brute $\pi_{\text{mise bas}}$ représente une probabilité apparente et diminue lorsque le niveau des censures augmente. Il est donc préférable d'estimer la probabilité corrigée :

$$\begin{aligned} \hat{\pi}_{\text{mise bas}} &= \frac{m_{\text{mise bas}}}{n_c} = \frac{m_{\text{mise bas}}}{n - \frac{m_{\text{cens}}}{2}}, \\ &= \frac{1}{4 - \frac{2}{2}} = 0.33 = 33 \%. \end{aligned}$$

Si les censures sont réparties de manières uniformes dans la période, alors le temps de présence dans la période d'un animal censuré est en moyenne une demi-période temps. Dans ce cas, $T = n_c$ et :

$$\hat{\pi}_{\text{mise bas, c}} = \frac{m_{\text{mise bas}}}{n_c} = \frac{m_{\text{mise bas}}}{T} = \hat{\mu}_{\text{mise bas}}$$

Dans l'exemple, le risque instantané de mise base est bien égal à la probabilité corrigée :

$$\hat{\mu}_{\text{mise bas}} = \frac{m_{\text{mise bas}}}{T} = \frac{1}{1 + 1 + \frac{1}{2} + \frac{1}{2}} = 0.33 \text{ année}^{-1}$$

Cependant, dans la majorité des cas, la répartition uniforme des censures ne sera pas exactement observée dans les données, et la probabilité corrigée ne sera qu'approximativement égale au risque instantané :

$$\hat{\pi}_{\text{mise bas, c}} \approx \hat{\mu}_{\text{mise bas}}$$

Comme dans la situation sans censure, la probabilité $\pi_{\text{mise bas, c}}$ n'a plus de sens lorsqu'une femelle peut avoir plusieurs mises bas dans la période de temps considérée. Cependant, il est à nouveau toujours possible de définir des sous-périodes assurant la condition "au plus une mise bas par femelle" et de calculer les probabilités de mise bas correspondant à ces sous-périodes à partir du risque instantané calculé sur la période globale.

4.1.4. Remarque sur l'intervalle moyen entre mises bas

L'intervalle moyen entre mises bas (*IMB*), paramètre souvent rencontré dans la littérature, est très difficile à estimer directement d'après les intervalles observés dans les données. Cette estimation directe ne prend en compte que les femelles pour lesquelles deux mises bas successives ont été observées. Ceci favorise les femelles ayant présenté une bonne reproduction et engendre en général une sous estimation de l'*IMB* réel du cheptel. Mettre en place des suivis de très longue durée est nécessaire pour limiter ce problème, ce qui est rarement possible dans les conditions d'élevage traditionnel.

Il est préférable d'estimer indirectement l'*IMB* à partir du taux de mise bas par les formules :

$$\widehat{IMB} = \frac{1}{\hat{\mu}_{\text{mise bas}}}$$

ou :

$$\widehat{IMB} = \frac{1}{\hat{\pi}_{\text{mise bas, c}}}$$

4.2. Taux de prolificité et de mortalité

Le taux de prolificité n'est ni une probabilité ni un risque instantané. Il est défini par l'effectif moyen de produits nés (vivants ou morts) par mise bas et peut être estimé par :

$$\widehat{prol} = \frac{m_{\text{produits vivants ou morts}}}{m_{\text{mise bas}}}$$

Le taux de mortinatalité est défini par la probabilité qu'un produit né soit un mort-né et peut être estimé par :

$$\pi_{\text{mortinat}} = \frac{m_{\text{produits morts}}}{m_{\text{produits vivants ou morts}}}$$

Dans le présent document, la mortinatalité n'est pas comptabilisée dans la mortalité naturelle, qui concerne uniquement les produits nés vivants.

5. INDICATEURS SYNTHETIQUES DE PRODUCTIVITE NUMERIQUE

Les taux démographiques peuvent être utilisés pour construire des indicateurs synthétiques de la productivité numérique d'un cheptel (des revues d'indicateurs sont présentées dans Bosman et al., 1997; James and Carles, 1996; Knipscheer et al., 1984; Peacock, 1987; Upton, 1989; Wilson, 1983).

L'un de ces indicateurs est le "taux annuel de fécondité" défini comme le nombre moyen de nés vivants produits par femelle et par an (Landais and Sissokho, 1986). Il peut être calculé par :

$$FEC = M \times prol \times (1 - \pi_{\text{mortinat}}),$$

où $prol \times (1 - \pi_{\text{mortinat}})$ représente le nombre moyen de produits nés vivants obtenus par mise bas et M le nombre moyen de mise bas par femelle et par an. Le nombre M correspond aux mises bas d'une femelle présente toute l'année et peut donc être calculé par le risque instantané de mise bas (en unité année⁻¹), ce qui donne :

$$FEC = \mu_{\text{mise bas}} \times prol \times (1 - \pi_{\text{mortinat}}),$$

Un autre indicateur est le taux annuel de productivité numérique à 3, 6, 9 ou 12 mois, défini par le nombre moyen de produits (à 3, 6, 9 ou 12 mois) obtenus en moyenne par femelle et par an. Il se déduit directement du taux annuel de fécondité, par exemple pour la productivité 6 mois :

$$PN_{6 \text{ mois}} = FEC \times (1 - \pi_{\text{mort, c. 0-6mois}}).$$

D'autres indicateurs utilisant les taux démographiques peuvent également être construits à partir de modèles démographiques (Baptist, 1992; Hary, 2004; Lesnoff, 2000; Lesnoff et al., 2000; Lesnoff et al., 2001; Upton, 1989). Plus complexes à calculer, ils ne sont pas abordés dans le présent document.

PARTIE II :

**LES METHODES DE
GROUPEMENTS DE DONNEES
DEMOGRAPHIQUES**

1. INTRODUCTION

Le groupement de données démographiques pour le calcul des taux démographiques s'effectue en deux étapes. La première étape consiste à positionner les événements (mise bas, décès, etc.) touchant chaque animal au cours de sa vie dans différentes "unités" correspond au croisement d'une classe d'âge et d'une période de temps. La seconde étape consiste à sommer les données individuelles au sein de chaque unité. Par exemple, pour le calcul du taux de mort naturelle, les données sont de type $\{m_{\text{mort}}, n, T\}$ où :

- m_{mort} représente le nombre de décès survenus dans l'unité,
- n l'effectif des animaux à risque en début d'unité,
- T le temps total de présence de ces animaux dans l'unité.

Il existe deux grandes approches pour construire les unités, chacune dépendant du fait que la décomposition concerne des cohortes d'animaux ou non. Nous présentons ces deux approches en prenant l'exemple de décompositions annuelles, c'est-à-dire engendrant des unités constituées du croisement de classes d'âge et de périodes de temps annuelles. La même démarche peut être utilisée pour construire des unités plus petites, par exemple des unités mensuelles ou hebdomadaires.

2. DECOMPOSITION DE COHORTES

2.1. Principe

Une cohorte est définie par l'ensemble des animaux nés au cours d'une même période de temps. Deux décompositions sont possibles :

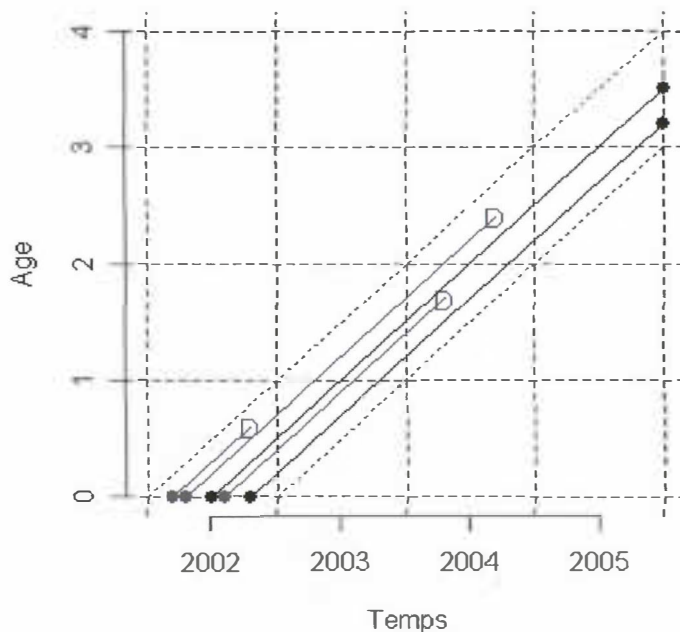
- une décomposition " horizontale ", c'est-à-dire par rapport à l'âge exact des animaux,
- une décomposition " verticale ", c'est-à-dire par rapport à l'âge révolu des animaux.

Considérons l'exemple d'une cohorte annuelle de 5 animaux nés entre le 1^{er} janvier et le 31 décembre 2002 et observés de leur naissance jusqu'au 31 décembre 2005. Trois animaux sont supposés être décédés avant cette date.

La carrière de cette cohorte peut être représentée dans un graphique âge = $f(\text{temps})$ appelé diagramme de Lexis. Les unités des décompositions horizontale et verticale sont des parallélogrammes, sauf la première unité verticale qui correspond à un triangle. Par convention dans ce document, les unités de décomposition sont numérotées de 0 à L .

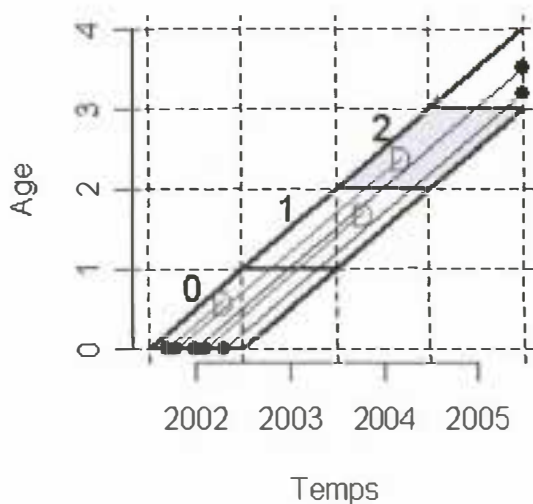
Dans l'exemple, les unités vont de 0 à 2 pour la décomposition horizontale (l'unité 3 est incomplète donc elle n'est pas considérée) et de 0 à 3 pour la décomposition verticale.

Diagramme de Lexis :

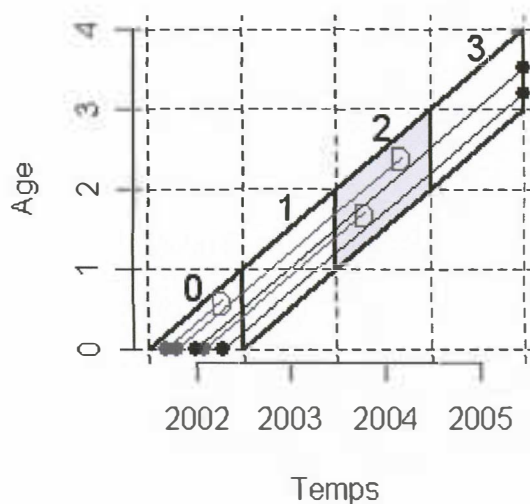


Décompositions des données :

Décomposition horizontale



Décomposition verticale



Le groupement des données par unité correspond au tableau suivant :

Unité	Type de décomposition					
	Horizontale			Verticale		
	n	T (année)	m_{mort}	n	T (année)	m_{mort}
0	5	4.6	1	5	2.2	1
1	4	3.6	1	4	4.0	0
2	3	2.3	1	4	2.9	2
3	-	-	-	2	2.0	0

Les taux démographiques des unités horizontales et verticales ont des interprétations différentes. Par exemple, l'unité horizontale n°2 représente les événements survenus entre les âges de 2 ans exacts (730 jours) et de 3 ans exacts (1095 jours), avec une probabilité brute de mort de :

$$\hat{\pi}_{\text{mort}} = \frac{m_{\text{mort}}}{n} = \frac{1}{3} = 33 \%$$

alors que l'unité verticale n°2 représente les événements survenus entre les âges d'1 an révolu et de 2 ans révolus, avec une probabilité brute de mort de :

$$\hat{\pi}_{\text{mort}} = \frac{m_{\text{mort}}}{n} = \frac{2}{4} = 50 \%$$

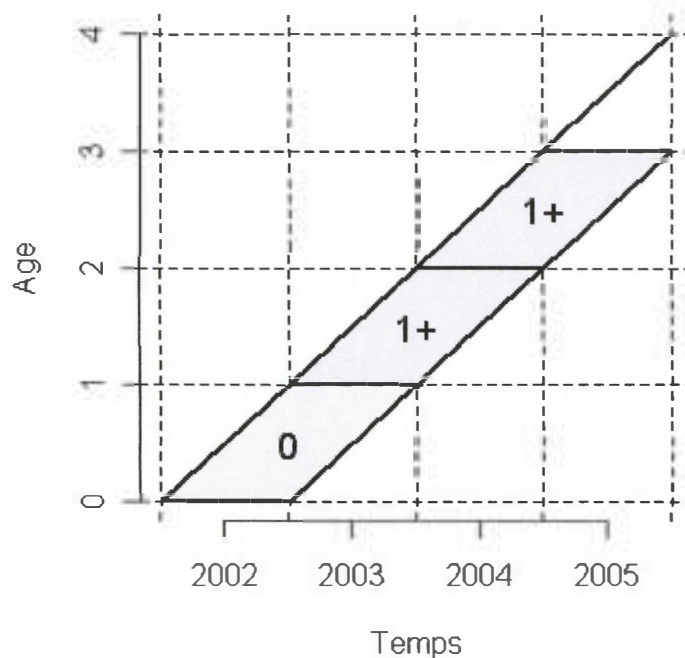
De la même manière, la probabilité de mort calculée dans l'unité horizontale n°0 représente la mortalité entre les âges exacts 0 et 1 an, alors que celle calculée dans l'unité verticale n°0 représente la mortalité des animaux "nés dans l'année" (et non la mortalité entre 0 et 1 an).

2.2. Regroupement d'unités

Selon les taux souhaités, il est possible de regrouper plusieurs unités construites d'après une première décomposition. Il suffit de sommer les données $\{m, n, T\}$ des unités qui composent le regroupement.

Par exemple, si dans l'exemple précédent l'objectif est d'étudier les mortalités annuelles entre 0 et 1 an exact puis des animaux d'âge supérieur à 1 an, les unités horizontales n°1 et n°2 peuvent être regroupées en unité "1+" :

Décomposition horizontale



La somme des données par groupe d'unités donne le tableau suivant :

Unité	n	T	m_{mort}
0	5	4.6	1
1 ⁺	7	5.9	2

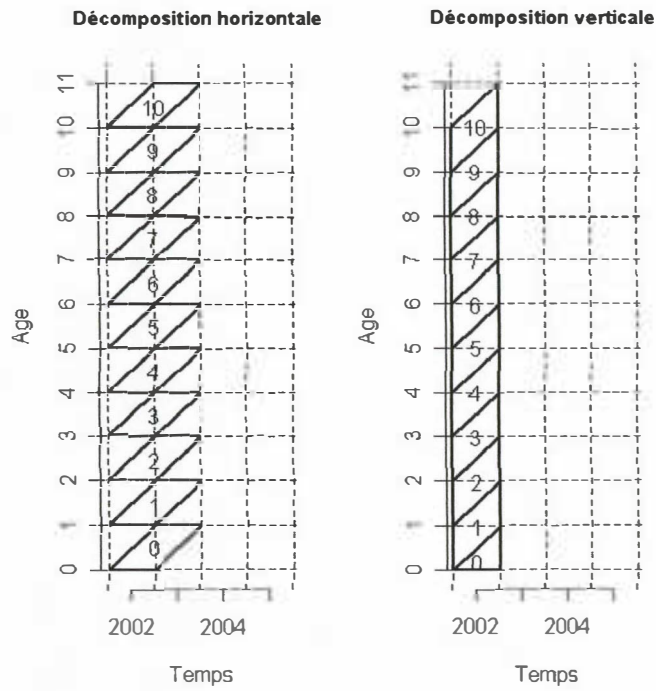
et la probabilité brute de mort pour l'unité n°1⁺ est de :

$$\hat{\pi}_{mort} = \frac{m_{mort}}{n} = \frac{2}{7} = 29 \%$$

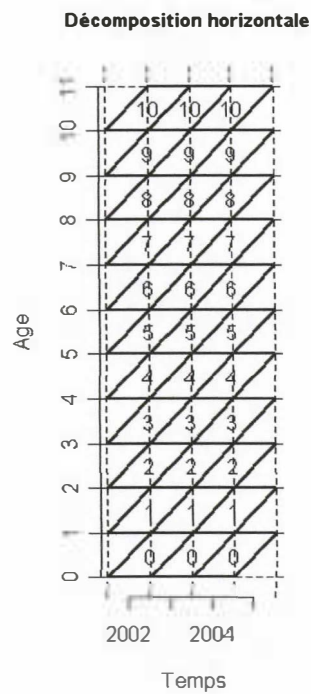
2.3. Pseudo-cohortes

En milieu tropical, il est rare de disposer de suivis suffisamment longs pour observer la carrière démographique complète de cohortes annuelles. Pour les bovins par exemple, ceci demanderait un suivi d'au moins 15 années successives. La démarche la plus courante est de décomposer plusieurs cohortes simultanément et de constituer des colonnes d'unités appelées "pseudo-cohortes".

Par exemple, les deux graphiques suivants représentent une pseudo-cohorte " 2002 " constituée d'unités provenant de 11 cohortes annuelles différentes :

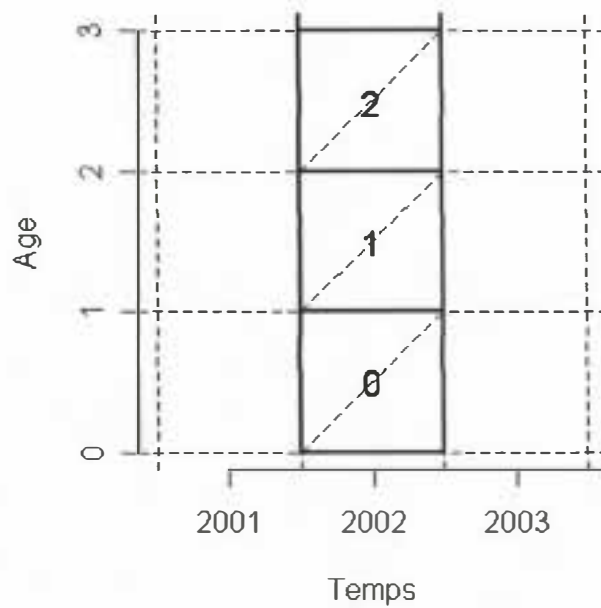


Cette démarche peut être appliquée sur plusieurs années successives, par exemple pour une décomposition horizontale :



3. AUTRES DECOMPOSITIONS

D'autres types de décompositions sont possibles, en particulier celles qui ne tiennent pas compte de la notion de cohortes. Par exemple, la décomposition croisant perpendiculairement l'âge des animaux et le temps engendre une succession de carrés dont chacun regroupe deux demi-cohortes différentes :



Cette décomposition ne permet pas d'estimer directement des probabilités car la notion d'effectif à risque n n'a pas de réelle signification ici.

PARTIE III :

**ANALYSE STATISTIQUE DES
TAUX DEMOGRAPHIQUES**

1. INTRODUCTION

Nous introduisons deux méthodes permettant de tester l'égalité de deux probabilités ou de deux risques instantanés : les tests utilisant l'approximation normale et les modèles linéaires généralisés. Ces méthodes, ainsi que de nombreuses autres méthodes disponibles, sont décrites en détail dans les ouvrages spécialisés (par exemple Agresti, 1990; Aitkin et al., 1989; Collett, 2003).

Les deux méthodes sont basées sur les hypothèses suivantes :

- tous les animaux sont indépendants entre eux (en particulier, il n'y a pas d'effet "troupeau" ou "village" pouvant engendrer des corrélations entre les animaux et des problèmes de surdispersion des données),
- pour la probabilité π , conditionnellement à l'effectif initial n , le nombre d'événements considérés suit une loi binomiale de moyenne π ,
- Pour le risque instantané μ , conditionnellement au temps à risque T , le nombre d'événements considérés suit une loi de Poisson de moyenne λ avec $\lambda = \mu \times T$.

Les estimateurs de la probabilité π et du risque instantané μ sont :

$$\hat{\pi} = \frac{m}{n},$$

$$\hat{\mu} = \frac{m}{T}.$$

et ont pour variances :

$$\text{Var}[\hat{\pi}] = \frac{\pi(1-\pi)}{n},$$

$$\text{Var}[\hat{\mu}] = \frac{\mu}{T}.$$

Les estimations de ces variances sont obtenues par :

$$\widehat{\text{Var}}[\hat{\pi}] = \frac{\hat{\pi}(1-\hat{\pi})}{n},$$

$$\widehat{\text{Var}}[\hat{\mu}] = \frac{\hat{\mu}}{T}.$$

Les écart-types correspondant sont (s.e. pour *standard errors*) :

$$\widehat{\text{s.e.}}[\widehat{\pi}] = \sqrt{\widehat{\text{Var}}[\widehat{\pi}]},$$

$$\widehat{\text{s.e.}}[\widehat{\mu}] = \sqrt{\widehat{\text{Var}}[\widehat{\mu}]}.$$

2. APPROXIMATION NORMALE

2.1. Intervalles de confiance de la probabilité et du risque instantané

Les intervalles de confiance de la probabilité π et du risque instantané μ sont souvent calculés à partir de l'approximation des lois binomiales et de Poisson par la loi normale, ce qui donne pour un risque d'erreur $\alpha\%$:

$$\widehat{IC}_\alpha[\pi] = \widehat{\pi} \pm z_{1-\alpha/2} \widehat{\text{s.e.}}[\widehat{\pi}],$$

$$\widehat{IC}_\alpha[\mu] = \widehat{\mu} \pm z_{1-\alpha/2} \widehat{\text{s.e.}}[\widehat{\mu}],$$

où $z_{1-\alpha/2}$ est le quantile d'ordre $1 - \alpha/2$ de la loi normale centrée réduite ($z = 1.96$ lorsque $\alpha = 5\%$).

2.2. Comparaison de deux probabilités

Considérons deux échantillons indépendants de tailles n_1 et n_2 et pour lesquels les probabilités estimées ont été :

$$\widehat{\pi}_1 = \frac{m_1}{n_1},$$

$$\widehat{\pi}_2 = \frac{m_2}{n_2}.$$

Le test de l'hypothèse nulle :

$$H_0 : \pi_1 = \pi_2$$

est construit en supposant que la différence des probabilités estimées suit approximativement une loi normale :

$$\widehat{\pi}_1 - \widehat{\pi}_2 \sim N(\pi_1 - \pi_2, \text{Var}[\widehat{\pi}_1 - \widehat{\pi}_2])$$

Comme précédemment, cela permet d'estimer l'intervalle de confiance de la différence entre les deux probabilités :

$$\widehat{IC}_\alpha[\pi_1 - \pi_2] = (\widehat{\pi}_1 - \widehat{\pi}_2) \pm z_{1-\alpha/2} \widehat{s.e.}[\widehat{\pi}_1 - \widehat{\pi}_2],$$

avec :

$$\begin{aligned} \widehat{s.e.}[\widehat{\pi}_1 - \widehat{\pi}_2] &= \sqrt{\widehat{Var}[\widehat{\pi}_1 - \widehat{\pi}_2]}, \\ &= \sqrt{\widehat{Var}[\widehat{\pi}_1] + \widehat{Var}[\widehat{\pi}_2]}, \\ &= \sqrt{\frac{\widehat{\pi}_1(1 - \widehat{\pi}_1)}{n_1} + \frac{\widehat{\pi}_2(1 - \widehat{\pi}_2)}{n_2}}. \end{aligned}$$

L'hypothèse nulle H_0 peut être rejetée au risque α si cet intervalle de confiance ne contient pas la valeur zéro. La valeur-p (*p-value*) du test peut également être calculée par :

$$p = P\left(|Z| > \frac{\widehat{\pi}_1 - \widehat{\pi}_2}{\widehat{s.e.}[\widehat{\pi}_1 - \widehat{\pi}_2]}\right),$$

où Z suit une loi normale centrée réduite.

2.3. Comparaison de deux risques instantanés

La comparaison de deux risques instantanés suit le même principe que celui utilisé pour les probabilités. L'intervalle de confiance de la différence entre deux risques instantanés μ_1 et μ_2 , dont les estimations proviennent de deux échantillons indépendants, est :

$$\widehat{IC}_\alpha[\mu_1 - \mu_2] = (\widehat{\mu}_1 - \widehat{\mu}_2) \pm z_{1-\alpha/2} \widehat{s.e.}[\widehat{\mu}_1 - \widehat{\mu}_2],$$

avec :

$$\begin{aligned} \widehat{\text{s.e.}}[\widehat{\mu}_1 - \widehat{\mu}_2] &= \sqrt{\widehat{\text{Var}}[\widehat{\mu}_1 - \widehat{\mu}_2]}, \\ &= \sqrt{\widehat{\text{Var}}[\widehat{\mu}_1] + \widehat{\text{Var}}[\widehat{\mu}_2]}, \\ &= \sqrt{\frac{\widehat{\mu}_1}{T_1} + \frac{\widehat{\mu}_2}{T_2}}. \end{aligned}$$

La valeur-p du test est calculée par :

$$p = P\left(|Z| > \frac{\widehat{\mu}_1 - \widehat{\mu}_2}{\widehat{\text{s.e.}}[\widehat{\mu}_1 - \widehat{\mu}_2]}\right),$$

où Z suit une loi normale centrée réduite.

3. MODELES LINEAIRES GENERALISES

3.1. Définition

Les modèles linéaires généralisés sont des modèles de régression dont l'écriture générique est la suivante :

$$g(E[y_i]) = x_i^T \beta,$$

où :

- g représente une fonction de lien, par exemple le lien *logit* ou le lien *log*,
- $x_i^T \beta$ représente le prédicteur linéaire (c'est-à-dire les covariables) pour l'observation n°i,
- la réponse y_i est une variable aléatoire supposée suivre une loi appartenant à la famille exponentielle, par exemple une loi binomiale ou de Poisson, et $E[y_i]$ représente son espérance (c'est-à-dire sa moyenne).

Dans l'échelle de la réponse y_i , le modèle s'écrit :

$$E[y_i] = g^{-1}(x_i^T \beta).$$

Les paramètres β sont estimés à l'aide de la méthode du maximum de vraisemblance. Les tests sur ces paramètres sont effectués par des tests de différences de déviances (appelés aussi tests de rapports de vraisemblances).

3.2. Régression logistique

L'un des modèles possibles pour comparer des probabilités est la régression logistique, correspondant à la fonction de lien *logit*. Pour l'observation n°i, la variable y_i représente la proportion d'animaux touchés par l'événement considéré :

$$y_i = \frac{m_i}{n_i}.$$

La moyenne de la variable y_i est la probabilité :

$$\pi_i = E[y_i].$$

Le modèle est :

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = x_i^T \beta,$$

et dans l'échelle de la réponse :

$$\pi_i = \frac{\exp(x_i^T \beta)}{1 + \exp(x_i^T \beta)}.$$

3.3. Régression log-linéaire

Les risques instantanés peuvent être comparés à l'aide du modèle de régression log-linéaire, correspondant à la fonction de lien *log* (Agresti, 1990 p. 190-192; Holford, 1980; Laird and Olivier, 1981).

Pour l'observation n°i, la variable y_i représente le nombre d'événements observés pour le temps à risque T_i . Conditionnellement à T_i , la moyenne de la variable y_i est :

$$\lambda_i = E[y_i].$$

Le risque instantané est :

$$\mu_i = \frac{\lambda_i}{T_i} = \frac{E[y_i]}{T_i}.$$

Le modèle est ajusté conditionnellement aux temps T_i qui sont considérés comme des valeurs fixes et connues. Les termes $\log(T_i)$ sont appelés *offsets*. Le modèle est :

$$\log \lambda_i = x_i^T \beta + \log T_i$$

ou (puisque $\mu_i = \lambda_i / T_i$) :

$$\log \mu_i = x_i^T \beta,$$

et dans l'échelle de la réponse :

$$\lambda_i = T_i \exp(x_i^T \beta)$$

ou :

$$\mu_i = \exp(x_i^T \beta).$$

4. EXEMPLE NUMERIQUE

4.1. Introduction

Nous reprenons l'exemple de la mortalité des jeunes ovins entre 0 et 1 an présenté dans la partie I. Les données proviennent d'une étude sur l'effet d'un traitement vétérinaire sur la mortalité des animaux. Deux groupes de troupeaux ont été considérés : un groupe contrôle (CTRL) et un groupe ayant reçu une prophylaxie (TREATED). Les données sont les suivantes :

Groupe	n	T (année)	m_{mort}	m_{expl}
CTRL	387	265.150	106	140
TREATED	310	220.706	50	115

Il s'agit de comparer les probabilités et les risques instantanés des deux groupes. Nous présentons les calculs avec le logiciel gratuit R téléchargeable à l'adresse <http://www.r-project.org>. Ces calculs peuvent être reproduits avec n'importe quel logiciel statistique.

4.2. Analyse descriptive

Sous le logiciel R, nous stockons les données dans un tableau nommé *mydata* :

```
mydata <- data.frame(
  group = c("CTRL", "TREATED"),
  n = c(387, 310),
  trisk = c(265.150, 220.706),
  m.death = c(106, 50),
  m.offtake = c(140, 115)
)
mydata
  group   n  trisk m.death m.offtake
1  CTRL 387 265.150    106      140
2 TREATED 310 220.706     50      115
```

Pour tenir compte des risques compétitifs (dus ici à l'exploitation des animaux dans les deux groupes), nous calculons l'effectif corrigé n_c :

```
mydata$n.c <- round(mydata$n - 0.5 * mydata$m.offtake)
```

```
mydata
  group   n   trisk m.death m.offtake n.c
1  CTRL 387 265.150   106     140 317
2 TREATED 310 220.706    50     115 252
```

Nous calculons ensuite la probabilité brute, la probabilité corrigée et le risque instantané de mort dans les deux groupes :

```
mydata$p.death <- mydata$m.death / mydata$n
mydata$p.death.c <- mydata$m.death / mydata$n.c
mydata$mu.death <- mydata$m.death / mydata$trisk
mydata
  group   n   trisk m.death m.offtake n.c p.death p.death.c mu.death
1  CTRL 387 265.150   106     140 317  0.274   0.334   0.400
2 TREATED 310 220.706    50     115 252  0.161   0.198   0.227
```

Ce tableau montre clairement que la probabilité brute (27 % et 16 % dans les groupes CTRL et TREATED) sous-estime le niveau réel de mortalité et qu'un meilleur indicateur est la probabilité corrigée (respectivement 33 % et 20 %). Les risques instantanés ont été respectivement de 0.40 année⁻¹ et 0.23 année⁻¹.

On peut vérifier que le calcul de la probabilité corrigée d'après le risque instantané donne des résultats très proches que celui utilisant l'effectif corrigé n_c :

```
cbind(mydata$p.death.c, 1 - exp(-mydata$mu.death))
  [,1] [,2]
[1,] 0.334 0.330
[2,] 0.198 0.203
```

Il s'agit maintenant de comparer les mortalités entre les deux groupes.

4.3. Tests

4.3.1. Approximation normale

Pour comparer les probabilités, nous calculons la différence :

$$\pi_{CTRL,c} - \pi_{TREATED,c}$$

et son écart-type :

$$s.e. \left[\hat{\pi}_{CTRL,c} - \hat{\pi}_{TREATED,c} \right]$$

```
p <- mydata$p.death.c
n <- mydata$n.c
dif <- p[1] - p[2]
se <- sum(p * (1 - p) / n)^0.5
c(dif = dif, se = se)
  dif     se
0.136  0.036
```

Nous en déduisons l'intervalle de confiance au risque 5 % :

$$\widehat{IC}_{5\%} [\pi_{CTRL,c} - \pi_{TREATED,c}] = (\widehat{\pi}_{CTRL,c} - \widehat{\pi}_{TREATED,c}) \pm 1.96 \times \widehat{s.e.} [\widehat{\pi}_{CTRL,c} - \widehat{\pi}_{TREATED,c}],$$

```
c(lower = dif - 1.96 * se, upper = dif + 1.96 * se)
```

```
lower    upper
0.064    0.207
```

La différence estimée entre les deux probabilités est de 13 % avec un intervalle de confiance (risque 5 %) de [6 %, 21 %] dont la borne inférieure est bien supérieure à la valeur zéro. Le traitement vétérinaire a donc eu un effet significatif sur la mortalité des jeunes ovins.

Nous appliquons les mêmes procédures de calcul pour comparer les risques instantanés :

```
mu <- mydata$mu.death
trisk <- mydata$trisk
dif <- mu[1] - mu[2]
se <- sum(mu / trisk)^0.5
```

```
c(dif = dif, se = se)
```

```
  dif      se
0.173  0.050
```

```
c(lower = dif - 1.96 * se, upper = dif + 1.96 * se)
```

```
lower    upper
0.074    0.271
```

La différence estimée entre les deux risques instantanés est de 0.17 année⁻¹ avec un intervalle de confiance (risque 5 %) de [0.07 année⁻¹, 0.27 année⁻¹].

4.3.2. Modèles linéaires généralisés

Régression logistique

Le modèle de régression logistique utilisé pour comparer les deux probabilités de mort est défini par :

$$\left\{ \begin{array}{l} \log\left(\frac{\pi_{CTRL}}{1 - \pi_{CTRL}}\right) = \beta_0, \\ \log\left(\frac{\pi_{TREATED}}{1 - \pi_{TREATED}}\right) = \beta_0 + \beta_1. \end{array} \right.$$

Le test de l'hypothèse nulle :

$$H_0 : \pi_{CTRL} = \pi_{TREATED}$$

revient au test de l'hypothèse nulle :

$$H_0 : \beta_1 = 0.$$

Sous R, le modèle peut être estimé par :

```
f1 <- glm(
  formula = cbind(m.death, n.c - m.death) ~ group,
  family = binomial(link = logit),
  data = mydata
)
summary(f1)
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.6884      0.1191  -5.783 7.36e-09 ***
groupTREATED  -0.7078      0.1978  -3.579 0.000346 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Les paramètres estimés sont :

$$\begin{cases} \hat{\beta}_0 = -0.6884 & (\text{s.e.} = 0.1191), \\ \hat{\beta}_1 = -0.7078 & (\text{s.e.} = 0.1978). \end{cases}$$

La valeur-p associée au paramètre β_1 est très faible ($p = 0.000346$) ce qui confirme l'effet significatif du traitement sur la mortalité.

Une autre procédure de test est le test de différence de déviances qui compare les modèles avec et sans effet groupe :

```
anova(f1, test = "Chisq")
Analysis of Deviance Table
Model: binomial, link: logit
Response: cbind(m.death, n.c - m.death)
Terms added sequentially (first to last)
      Df Deviance Resid. Df Resid. Dev P(>|Chi|)
NULL                                1    13.3128
group  1    13.3128                0 -1.110e-14    0.0003
```

La valeur-p du test est à nouveau très faible ($p = 0.0003$).

Régression log-linéaire

Le modèle de régression log-linéaire utilisé pour comparer les deux risques instantanés de mort est défini par :

$$\begin{cases} \log \lambda_{\text{CTRL}} = \beta_0 + \log T_{\text{CTRL}}, \\ \log \lambda_{\text{TREATED}} = \beta_0 + \beta_1 + \log T_{\text{TREATED}}. \end{cases}$$

X-Authentication-Warning: localhost.localdomain: mail set sender to owner-codiremvt@messmpl.cirad.fr using -f
Date: Mon, 22 May 2006 14:42:48 +0200
From: Emmanuel Camus <emmanuel.camus@cirad.fr>
Organization: CIRAD
User-Agent: Mozilla/5.0 (Windows; U; Windows NT 5.1; fr-FR; rv:1.0.2) Gecko/20030208 Netscape/7.02
X-Accept-Language: fr-fr, fr
To: codiremvt@cirad.fr
Subject: [Fwd: stratégie scientifique du département EMVT]

Comme précédemment, il s'agit de tester l'hypothèse nulle :

$$H_0 : \beta_1 = 0.$$

Sous R, l'estimation du modèle et le test de différence de déviance sont réalisés par :

```
fit <- anova(codiremvt@cirad.cirad.fr  
X-Spam-Checker-Version: SpamAssassin 2.63 (2004-01-11) on queyras.cirad.fr  
X-Spam-Offset = log(trisk)  
X-Spam-Status: No, bits=-104.7 required=5.0 tests=BAYES_00,HTML_MESSAGE,  
HTML_TITLE_EMPTY,USER_IN_WHITELIST autolearn=no version=2.63  
data = mydata  
)
```

summary(f2)

---Message d'origine -----

Estimate Std. Error z value Pr(>|z|)

Sujet: stratégie scientifique du département EMVT 16 ***

Date: Mon, 22 May 2006 14:02:48 +0200 -3.31 0.000931 ***

De: Emmanuel Camus <emmanuel.camus@cirad.fr>

Société: CIRAD: 0 '***' 0.001 '*' 0.01 '*' 0.05 '.' 0.1 ' ' 1

A: college@cirad.fr

anova(f2, test = "Chisq")

Analysis of Deviance Table

22/05/2006

Model: poisson, link: log

Response: m.death

Terms added sequentially (first to last)

Nous vous recommandons de consulter le site de présentation de la stratégie

scientifique du département EMVT, 11.5951

en vue du 11.5951 23 mai 2006. 4.885e-15 0.0007

Vous en souhaitant bonne réception,

L'interprétation des résultats est identique à celle des résultats du modèle de régression

logistique.

Secrétariat de M. E. CAMUS

Attachment Converted: "c:\eudora\attach\EPRD 2007 =windows-1252Qstrat=E9gie scientifique EMVT=2E21 mai ==windows-1252Q06=2Edoc=.doc"

BIBLIOGRAPHIE

- Agresti, A., 1990, *Categorical data analysis*. Wiley.
- Aitkin, M., Anderson, D., Francis, B., Hinde, J., 1989, *Statistical modelling in GLIM*. Clarendon Press, Oxford.
- Anderson, D.R., Burnham, K.P., 1976, Population of the Mallard. VI. The effect of exploitation on survival. United States Department of the Interior Fish and Wildlife Service, Resource Publication 128, Washington D.C., USA.
- Baptist, R., 1992, Derivation of steady-state herd productivity. *Agric. Syst.* 39, 253-272.
- Bosman, H.G., Moll, H.A.J., Udo, H.M.J., 1997, Measuring and interpreting the benefits of goat keeping in tropical farm systems. *Agric. Syst.* 53, 349-372.
- Burnham, K.P., Anderson, D.R., 1984, Tests of compensatory vs. additive hypotheses of mortality in mallards. *Ecology* 65, 105-112.
- CIRAD-IEMVT, 1972, Définition des paramètres zootechniques concernant la reproduction, l'économie des troupeaux, l'économie du bétail de boucherie en élevage bovin, ovin, caprin. CIRAD-IEMVT, Maisons-Alfort.
- CIRAD-IEMVT, 1989, Les enquête sur la productivité du bétail. Fiche Technique n°5. CIRAD-IEMVT, 4 p.
- Collett, D., 2003, *Modelling binary data*, 2 Edition. Chapman & Hall/CRC, New York.
- Cox, D.R., Oakes, D., 1984, *Analysis of survival data*. Chapman and Hall, New York, 201 p. p.
- Faugère, O., Faugère, B., 1986, Suivi de troupeaux et contrôle de performances individuelles des petits ruminants en milieu traditionnel africain. Aspects méthodologiques. *Rev. Elev. Méd. vét. Pays trop.* 39, 29-40.
- Hary, I., 2004, Derivation of steady-state herd productivity using age-structured population models and mathematical programming. *Agric. Syst.* 81, 133-152.
- Holford, T.R., 1976, Life tables with concomitant information. *Biometrics* 32, 587-597.
- Holford, T.R., 1980, The analysis of rates and of survivorship using log-linear models. *Biometrics* 36, 299-305.
- IEMVT, 1989, *Manuel d'élevage du mouton en zone tropicale humide d'Afrique*, Vol 11. Ministère de la Coopération et du Développement, La Documentation française, 208 p.
- ILCA, 1990, *Livestock systems research manual. Working Paper 1, Vol. 1*. International Livestock Centre for Africa (ILCA), Addis Ababa, Ethiopia, 399 p.
- Itty, P., 1995, Application of a bio-economic herd simulation model to African cattle production systems: implications for village milk production. *Q. J. Int. Agric.* 34, 372-385.
- James, A.D., Carles, A.B., 1996, Measuring the productivity of grazing and foraging livestock. *Agric. Syst.* 52, 271-291.
- Kalbfleisch, J.D., Prentice, R.L., 1980, *The statistical analysis of failure time data*. Wiley, New York, USA.
- Knipscheer, H.C., Kusnadi, O., DeBoer, A.J., 1984, Some efficiency measures for analysis of the productive potential of Indonesian goats. *Agric. Syst.* 15, 125-135.
- Laird, N., Olivier, D., 1981, Covariance analysis of censored survival data using log-linear analysis techniques. *Journal of American Statistical Association* 76, 231-240.
- Landais, E., 1986, Contributions à l'étude des interférences entre phénomènes démographiques et propositions pour l'estimation sans biais des paramètres démographiques usuels, In: *Méthodes pour la recherche sur les systèmes d'élevage en Afrique intertropicale. Etudes et synthèses de l'IEMVT*. CIRAD - IEMVT, Maisons Alfort, France, pp. 497-517.
- Landais, E., Sissokho, M.M., 1986, Bases méthodologiques du contrôle des performances animales pour l'analyse zootechnique et démographique : collecte des données et choix des variables, In: *Méthodes pour la recherche sur les systèmes d'élevage en Afrique intertropicale. Etudes et synthèses de l'IEMVT*. CIRAD-IEMVT, Maisons Alfort, France, pp. 433-485.
- Lee, E.T., 1992, *Statistical methods for survival data analysis*, 2 Edition. Wiley, New York, USA, 482 p.

- Lesnoff, M., 1999, Dynamics of a sheep population in a sahelian area (Ndiagne district in Senegal): A periodic matrix model. *Agric. Syst.* 61, 207-221.
- Lesnoff, M., 2000. Etude de la dynamique et de la productivité des populations domestiques tropicales par les modèles matriciels en temps discret : les populations d'ovins au Sénégal. Université de Montpellier II - Sciences et Techniques du Languedoc, Montpellier.
- Lesnoff, M., Lancelot, R., Tillard, E., Dohoo, I.R., 2000, A steady-state approach of benefit-cost analysis with a periodic Leslie-matrix model. Presentation and application to the evaluation of a sheep-diseases preventive scheme in Kolda, Senegal. *Prev. Vet. Med.* 46, 113-128.
- Lesnoff, M., Lancelot, R., Tillard, E., Faye, B., 2001, Analyse comparative de la productivité des cheptels de petits ruminants en élevage extensif tropical : une nouvelle approche par les modèles matriciels en temps discret. *Rev. Elev. Méd. vét. Pays trop.* 54, 69-80.
- Lhoste, P., Dollé, V., Rousseau, J., Soltner, D., 1993, Manuel de zootechnie des régions chaudes - Les systèmes d'élevage. Ministère de la Coopération, Collection Précis d'Élevage, Paris.
- Mukasa-Mugerwa, E., 1989, A review of reproductive performance of female *Bos Indicus* (zebu) cattle. ILCA monograph n°6, International Livestock Centre for Africa, Addis Ababa.
- Peacock, C.P., 1987, Measures for assessing the productivity of sheep and goats. *Agric. Syst.* 23, 197-210.
- Prentice, R.L., 1978, The analysis of failure times in the presence of competing risks. *Biometrics* 34, 541-554.
- SEDES, 1975, Etude de la structure et de la dynamique des troupeaux bovins : méthodologie pratique. Ministère de la Coopération, Paris.
- Tacher, G., 1975a, Notes sur la démographie bovine au Sahel : I. Représentation et projection théorique du troupeau par un modèle mathématique. *Rev. Elev. Méd. vét. Pays trop.* 28, 547-569.
- Tacher, G., 1975b, Notes sur la démographie bovine au Sahel : II. Représentation et projection théoriques du troupeau par un modèle mathématique après une catastrophe. *Rev. Elev. Méd. vét. Pays trop.* 28, 571-595.
- Texeira, M., Paruelo, J.M., 2005, Demography, population dynamics and sustainability of the Patagonian sheep flocks. *Agric. Syst.* In Press.
- Upton, M., 1989, Livestock productivity assessment and herd growth models. *Agric. Syst.* 29, 149-164.
- Wilson, R.T., 1983, Studies on the livestock of Southern Darfur, Sudan. VIII. A comparison of productivity indices for goats and sheep. *Trop. Anim. Hlth Prod.* 15, 63-68.
- Wilson, R.T., 1986, Livestock production in central Mali: long-term studies on cattle and small ruminants in the agropastoral system. Research Report 14. ILCA, Addis Ababa, Ethiopia.