ÉCOLE NATIONALE VÉTÉRINAIRE D'ALFORT

Année 2015

GUIDE PRATIQUE DE MODÉLISATION DE LA CROISSANCE PONDÉRALE DE JEUNES MAMMIFÈRES SAUVAGES À TRAVERS DEUX EXEMPLES

THÈSE

Pour le

DOCTORAT VÉTÉRINAIRE

Présentée et soutenue publiquement devant

LA FACULTÉ DE MÉDECINE DE CRÉTEIL

le 19 Février 2015

par

Laurie Marie VIROLLE

Née le 8 Septembre 1988 à Limoges (Haute-Vienne)

JURY

Président : Pr.
Professeur à la Faculté de Médecine de CRÉTEIL

Membres

Directeur : M. Loïc DESQUILBET
Maître de conférences à l'ENVA

Assesseur : Mme Valérie CHETBOUL
Professeur de cardiologie à l'ENVA

Octobre 2014

LISTE DES MEMBRES DU CORPS ENSEIGNANT

Directeur : M. le Professeur GOGNY Marc

Directeurs honoraires : MM. les Professeurs : COTARD Jean-Pierre, MIALOT Jean-Paul, MORAILLON Robert, PARODI André-Laurent, PILET Charles, TOMA Bernard.
Professeurs honoraires : Mme et MM. : BENET Jean-Jacques, BRUGERE Henri, BRUGERE-PICOUX Jeanne, BUSSIERAS Jean, CERF Olivier, CLERC Bernard,CRESPEAU François, DEPUTTE Bertrand, MOUTHON Gilbert, MILHAUD Guy, POUCHELON Jean-Louis, ROZIER Jacques.

DEPARTEMENT D'ELEVAGE ET DE PATHOLOGIE DES EQUIDES ET DES CARNIVORES (DEPEC)

Chef du département par intérim : M. GRANDJEAN Dominique, Professeur - Adjoint : M. BLOT Stéphane, Professeur

UNITE DE CARDIOLOGIE
- Mme CHETBOUL Valérie, Professeur *
- Mme GKOUNI Vassiliki, Praticien hospitalier
- Mme SECHI-TREHIOU Emilie, Praticien hospitalier

UNITE DE CLINIQUE EQUINE
- M. AUDIGIE Fabrice, Professeur
- Mme BERTONI Lélia, Maître de conférences contractuel
- Mme BOURZAC Céline, Maître de conférences contractuel
- M. DENOIX Jean-Marie, Professeur
- Mme GIRAUDET Aude, Praticien hospitalier *
- Mme MESPOULHES-RIVIERE Céline, Praticien hospitalier
- Mme TRACHSEL Dagmar, Maître de conférences contractuel

UNITE D'IMAGERIE MEDICALE
- Mme PEY Pascaline, Maître de conférences contractuel
- Mme STAMBOULI Fouzia, Praticien hospitalier

UNITE DE MEDECINE
- Mme BENCHEKROUN Ghita, Maître de conférences
- M. BLOT Stéphane, Professeur*
- M. CAMPOS Miguel, Maître de conférences associé
- Mme FREICHE-LEGROS Valérie, Praticien hospitalier
- Mme MAUREY-GUENEC Christelle, Maître de conférences

UNITE DE MEDECINE DE L'ELEVAGE ET DU SPORT
- Mme CLERO Delphine, Maître de conférences contractuel
- M. FONTBONNE Alain, Maître de conférences
- M. GRANDJEAN Dominique, Professeur *
- Mme MAENHOUDT Cindy, Praticien hospitalier
- M. NUDELMANN Nicolas, Maître de conférences
- Mme YAGUIYAN-COLLIARD Laurence, Maître de conférences contractuel

DISCIPLINE : NUTRITION-ALIMENTATION
- M. PARAGON Bernard, Professeur

DISCIPLINE : OPHTALMOLOGIE
- Mme CHAHORY Sabine, Maître de conférences

UNITE DE PARASITOLOGIE ET MALADIES PARASITAIRES
- M. BLAGA Radu Gheorghe, Maître de conférences (rattaché au DPASP)
- M. CHERMETTE René, Professeur (rattaché au DSBP)
- Mme COCHET-FAIVRE Noëlle, Praticien hospitalier
- M. GUILLOT Jacques, Professeur *
- Mme MARIGNAC Geneviève, Maître de conférences
- M. POLACK Bruno, Maître de conférences
- Mme RISCO CASTILLO Véronica, Maître de conférences (rattachée au DSBP)

UNITE DE PATHOLOGIE CHIRURGICALE
- M. FAYOLLE Pascal, Professeur
- M. MAILHAC Jean-Marie, Maître de conférences
- M. MANASSERO Mathieu, Maître de conférences
- M. MOISSONNIER Pierre, Professeur*
- Mme RAVARY-PLUMIOEN Bérangère, Maître de conférences (rattachée au DPASP)
- Mme VIATEAU-DUVAL Véronique, Professeur
- M. ZILBERSTEIN Luca, Maître de conférences

DISCIPLINE : URGENCE SOINS INTENSIFS
- Mme STEBLAJ Barbara, Praticien Hospitalier

DISCIPLINE : NOUVEAUX ANIMAUX DE COMPAGNIE
- M. PIGNON Charly, Praticien hospitalier

DEPARTEMENT DES PRODUCTIONS ANIMALES ET DE LA SANTE PUBLIQUE (DPASP)

Chef du département : M. MILLEMANN Yves, Professeur - Adjoint : Mme DUFOUR Barbara, Professeur

UNITE D'HYGIENE QUALITE ET SECURITE DES ALIMENTS
- M. AUGUSTIN Jean-Christophe, Professeur
- M. BOLNOT François, Maître de conférences *
- M. CARLIER Vincent, Professeur

UNITE DES MALADIES CONTAGIEUSES
- Mme DUFOUR Barbara, Professeur*
- Mme HADDAD/HOANG-XUAN Nadia, Professeur
- Mme PRAUD Anne, Maître de conférences
- Mme RIVIERE Julie, Maître de conférences contractuel

UNITE DE PATHOLOGIE DES ANIMAUX DE PRODUCTION
- M. ADJOU Karim, Maître de conférences *
- M. BELBIS Guillaume, Assistant d'enseignement et de recherche contractuel
- M. MILLEMANN Yves, Professeur
- Mme ROUANNE Sophie, Praticien hospitalier

UNITE DE REPRODUCTION ANIMALE
- Mme CONSTANT Fabienne, Maître de conférences
- M. DESBOIS Christophe, Maître de conférences (rattaché au DEPEC)
- Mme MASSE-MOREL Gaëlle, Maître de conférences contractuel
- M. MAUFFRE Vincent, Assistant d'enseignement et de recherche contractuel
- Mme EL BAY Sarah, Praticien hospitalier

UNITE DE ZOOTECHNIE, ECONOMIE RURALE
- M. ARNE Pascal, Maître de conférences
- M. BOSSE Philippe, Professeur*
- M. COURREAU Jean-François, Professeur
- Mme DE PAULA-REIS Alline, Maître de conférences contractuel
- Mme GRIMARD-BALLIF Bénédicte, Professeur
- Mme LEROY-BARASSIN Isabelle, Maître de conférences
- M. PONTER Andrew, Professeur
- Mme WOLGUST Valérie, Praticien hospitalier

DEPARTEMENT DES SCIENCES BIOLOGIQUES ET PHARMACEUTIQUES (DSBP)

Chef du département : Mme COMBRISSON Hélène, Professeur - Adjoint : Mme LE PODER Sophie, Maître de conférences

UNITE D'ANATOMIE DES ANIMAUX DOMESTIQUES
- M. CHATEAU Henry, Maître de conférences*
- Mme CREVIER-DENOIX Nathalie, Professeur
- M. DEGUEURCE Christophe, Professeur
- Mme ROBERT Céline, Maître de conférences

DISCIPLINE : ANGLAIS
- Mme CONAN Muriel, Professeur certifié

UNITE DE BIOCHIMIE
- M. BELLIER Sylvain, Maître de conférences*
- Mme LAGRANGE Isabelle, Praticien hospitalier
- M. MICHAUX Jean-Michel, Maître de conférences

DISCIPLINE : BIOSTATISTIQUES
- M. DESQUILBET Loïc, Maître de conférences

DISCIPLINE : EDUCATION PHYSIQUE ET SPORTIVE
- M. PHILIPS Pascal, Professeur certifié

DISCIPLINE : ETHOLOGIE
- Mme GILBERT Caroline, Maître de conférences

UNITE DE GENETIQUE MEDICALE ET MOLECULAIRE
- Mme ABITBOL Marie, Maître de conférences
- M. PANTHIER Jean-Jacques, Professeur*

UNITE D'HISTOLOGIE, ANATOMIE PATHOLOGIQUE
- Mme CORDONNIER-LEFORT Nathalie, Maître de conférences*
- M. FONTAINE Jean-Jacques, Professeur
- Mme LALOY Eve, Maître de conférences contractuel
- M. REYES GOMEZ Edouard, Maître de conférences

UNITE DE PATHOLOGIE GENERALE MICROBIOLOGIE, IMMUNOLOGIE
- M. BOULOUIS Henri-Jean, Professeur
- Mme LE ROUX Delphine, Maître de conférences
- Mme QUINTIN-COLONNA Françoise, Professeur*

UNITE DE PHARMACIE ET TOXICOLOGIE
- Mme ENRIQUEZ Brigitte, Professeur
- M. PERROT Sébastien, Maître de conférences
- M. TISSIER Renaud, Professeur*

UNITE DE PHYSIOLOGIE ET THERAPEUTIQUE
- Mme COMBRISSON Hélène, Professeur
- Mme PILOT-STORCK Fanny, Maître de conférences
- M. TIRET Laurent, Professeur *

DISCIPLINE : VIROLOGIE
- Mme LE PODER Sophie, Maître de conférences *

DISCIPLINE : SCIENCES DE GESTION ET DE MANAGEMENT
- Mme FOURNEL Christelle, Maître de conférences contractuel

*** responsable d'unité**

REMERCIEMENTS

Au Professeur
de la Faculté de Médecine de Créteil, qui nous a fait l'honneur de présider ce jury, hommage respectueux.

Au Docteur Loïc Desquilbet,
qui a accepté la direction de cette thèse, l'obligeant (à nouveau) à corriger mes interprétations, mes équations et mes programmes. Pour sa disponibilité au cours de ces trois dernières années, sa patience presque sans limite face à mes lacunes (nombreuses), sa bonne humeur, son soutien et ses encouragements. Je n'aurais pas pu rêver un meilleur encadrant pour ce travail, mes plus sincères et profonds remerciements.

Au Professeur Valérie Chetboul,
qui a bien voulu être l'assesseur de cette thèse quelque peu éloignée de son domaine de prédilection. Pour son intérêt pour la faune sauvage, ses qualités humaines, son ouverture d'esprit, et ses conseils dépassant largement le cadre de cette thèse, mes sincères et profonds remerciements.

Au Professeur Christophe Degueurce,
qui m'a permis de m'intéresser à l'histoire de cette École (et de la médecine vétérinaire) en devenant guide au musée et en réalisant les inventaires de la collection. Pour le partage de son savoir, sa gentillesse, et cette magnifique expérience, un grand merci.

A Isabelle Bidart, Véronique Le Bechec et Florence Guinault, les « dames de la Scolarité »
qui nous encadrent dans toutes nos démarches depuis la première année, un grand merci pour votre patience et votre écoute.

Au Docteur Fernando Nájera,
pour avoir partagé avec moi son amour de la nature, de la faune sauvage et des panthères nébuleuses. Je suis ravie que nos chemins se soient croisés et que nous ayons pu travailler ensemble sur cet article.

Au Docteur Pierre-Yves Henry et à son équipe,
pour m'avoir proposé un stage de Master qui m'a orientée vers ce sujet de thèse.

À Julie Landes,
pour m'avoir aidée à mieux connaitre les microcèbes, et pour sa disponibilité à tout moment.

À tous mes maîtres de stages,
qui m'ont donné de leur temps, ont partagé leur savoir, et m'ont laissée faire mes premiers pas de vétérinaire. Merci de m'avoir montré que je faisais partie de cette grande famille.

À ma famille, qui m'a toujours encouragée, soutenue, portée et supportée !

Maman, Papa, vous êtes les meilleurs parents de la Terre … sans vous, je n'en serais pas là !

Jojo, mon p'tit frère chéri, allez au boulot ! La prochaine thèse, c'est la tienne !

Mamie, merci de m'entourer de douceur et d'amour depuis 26 ans.

Papy, j'aimerais tant que tu sois là pour pouvoir tout te raconter, tu me manques.

À Camille, pour ta patience, ton soutien et tout le reste… merci d'être entré dans ma vie.

À mes amis vétos, qui ont rendu ces cinq années inoubliables :

À mes deux blondes irremplaçables,

Nath, ma coloc jusqu'au Québec, on a tant partagé… pourvu que ça dure !

Chamo, le hasard fait tellement bien les choses ! Une pensée pour la lampe à lave…

À mes Fisty, et tous nos moments partagés passés et à venir, merci d'être là, merci d'être vous.

Sandra, ma coloc au cœur gros comme ça, tu es unique, ne change rien !

Robinou, mon jambon en chef amateur d'escalade, merci pour tous ces fous rires !

Clo, squatteuse des midis et poulpy en chef, merci pour tes précieux conseils en tous genres.

Niac, petit parasite de canapé, ma porte sera toujours ouverte !

Nono, vilain breton, pour tous les délires et les cafés partagés au grisby.

Laure, belle gosse toujours de bon conseil, pour les moments d'anthologie au volley ou au ski !

Alex, ma co-thésarde de choc, je suis heureuse de partager cette journée avec toi.

Chach, notre pro du frisbee, pour nos discussions et tous ces bons moments.

Aux plus « grands »,

Stef, pour nos aventures en Thaïlande, avec Yaboumba… et celles à venir ! La vie est belle !

Apo, Rémi-chou et Virginie, pour tous nos souvenirs et craquages au Chuva.

Aux petits poulots, qui ont intérêt à assurer la relève ! Hein, Lauriane ?

À mes amies, ça commence à faire un bout de temps que vous me supportez…

Morue, depuis le premier jour de prépa tu as toujours été là, que cette amitié dure toute la vie !

Marie, aux kit-kats et à ton soutien permanent… le temps passe, la complicité reste.

Julie, ça valait le coup de refaire un an de prépa, merci de partager ma passion pour la Nature !

À mes copines de Master, vous êtes des filles en or !

Soline, pour nos rédactions de mémoires tardives et nos levers difficiles.

Audrey, pour tes conseils toujours avisés et nos délires inoubliables en coloc, ou en cours.

Manon, la fille la plus organisée du monde, tu m'épateras toujours.

À Norin, Muriel, Aude et Christelle pour tous ces souvenirs inoubliables à la Ménagerie et votre soutien pendant la rédaction de cette thèse.

À First, le plus beau des chiens guides d'aveugle, qui sera toujours un peu mon loulou.

À Swakopmund, chat magique qui comprend tout, sauf comment doit s'asseoir un vrai chat.

À Havane et ses 27 kilos d'amour débordant, qui mettent tant de bonheur dans mon quotidien.

À tous les autres quatre-pattes qui ont partagé ma vie.

TABLE DES MATIÈRES

INTRODUCTION……9

- PARTIE A - THÉORIE ET ÉTAPES DE CONSTRUCTION D'UN MODÈLE EN ADÉQUATION AVEC DES DONNÉES DE POIDS DE JEUNES PANTHÈRES NÉBULEUSES COLLECTÉES DANS LES 90 PREMIERS JOURS DE VIE 11

I. OBTENTION ET GESTION DES DONNÉES : DU RECUEIL À LA MISE EN FORME..... 13

A. PRÉSENTATION DE L'ESPÈCE ET DES ANIMAUX ÉTUDIÉS.....13

1. La panthère nébuleuse, prise en compte des particularités de l'espèce 13

2. Jeunes panthères nébuleuses utilisées dans l'étude.....15

B. COLLECTE DES DONNÉES 16

C. ÉLABORATION D'UNE BASE DE DONNÉES.....17

II. REPRÉSENTATIONS GRAPHIQUES DES DONNÉES 20

A. NUAGES DE POINTS 20

B. COURBES INDIVIDUELLES.....21

C. MOYENNES ± ÉCARTS TYPES OU INTERVALLES DE CONFIANCE À 95 %.....21

III. CHOIX DU MODÈLE..... 23

A. MODÈLES UTILISANT DES ÉQUATIONS DE DROITES : LES MODÈLES LINÉAIRES 24

1. Théorie mathématique sur les modèles linéaires.....24

2. Modélisation linéaire de la croissance pondérale chez les panthères nébuleuses 27

a. Modèle sans rupture de pente..... 28

b. Modèle à une rupture de pente 29

B. MODÈLES PLUS COMPLEXES : LES MODÈLES NON-LINÉAIRES 31

1. Modèle autorisant des points d'inflexion : le modèle polynomial.....31

2. Modèles de croissance curviligne à un point d'inflexion ou moins 32

a. Modèle exponentiel..... 32

b. Modèle de Gompertz..... 32

c. Modèle logistique 32

C. ADÉQUATION DU MODÈLE AUX DONNÉES 33

1. Une méthode largement utilisée : le critère d'Akaike.....33

2. Cas concret de l'étude sur les panthères nébuleuses : combien de ruptures de pentes choisir ?.....33

IV. INTRODUCTION DE VARIABLES EXPLICATIVES À EFFETS FIXES DANS LE MODÈLE ET ANALYSE DES RÉSULTATS 35

A. ANALYSE DE L'ASSOCIATION BRUTE ENTRE LE SEXE DE LA PANTHÈRE ET SA CROISSANCE 35

1. Description de l'échantillon et du modèle 35

2. Lecture des résultats 36

B. ANALYSE DE L'ASSOCIATION BRUTE ENTRE LA DURÉE DES PROBLÈMES DIGESTIFS ET LA CROISSANCE DES PANTHÈRES 37

1. Description de l'échantillon 37

2. Lecture des résultats 38

C. NOTION DE MODÈLE AJUSTE, PRISE EN COMPTE DE FACTEURS DE CONFUSION 39

V. CONTRAINTES IMPOSÉES PAR LA NON-INDEPENDANCE DES DONNÉES 42

A. NOTION DE NON-INDEPENDANCE 42

B. PRISE EN COMPTE DE LA NON INDÉPENDANCE DES DONNÉES : INTRODUCTION D'EFFETS ALÉATOIRES DANS LE MODÈLE 43

C. CAS PARTICULIER DES DONNÉES RÉPÉTÉES 44

VI. LOGICIELS DE MODÉLISATION ET PROCÉDURES 45

- PARTIE B - RETOUR D'EXPÉRIENCE SUR L'ANALYSE DE LA CROISSANCE PONDÉRALE DE JEUNES MICROCÈBES MURINS À L'AIDE DE DEUX MODÈLES 47

DONNÉES BIBLIOGRAPHIQUES SUR L'ESPÈCE MICROCEBUS MURINUS 49

PARTIE EXPÉRIMENTALE 51

I. CONTEXTE ET OBJECTIFS DE L'ETUDE 51

II. MATÉRIEL ET MÉTHODES 52

A. MATÉRIELS ANIMAUX ET RECUEIL DE DONNÉES 52

B. BASE DE DONNÉES 52

C. MÉTHODES STATISTIQUES 54

1. Construction de deux modèles adaptés à la croissance des microcèbes 54

a. Visualisation des données brutes en nuage de points 54

b. Modèle linéaire 55

c. Modèle non linéaire 56

2. Prise en compte de la non-indépendance des données dans les deux types de modèles 58

3. Modèle linéaire ou non linéaire : lequel choisir ? 61

III. RÉSULTATS 63

A. DESCRIPTION DE L'ÉCHANTILLON 63

B. EFFET BRUT DU NOMBRE DE PETITS DANS LA PORTÉE SUR LA CROISSANCE DU MICROCÈBE 64

1. Analyse après création d'une variable binaire (utilisation des deux modèles) 64
   a. Présentation de la variable et des équations des modèles 64
   b. Résultats de l'analyse avec le modèle linéaire 65
   c. Résultats de l'analyse avec le modèle non linéaire 66

2. Analyse à l'aide de la variable quantitative correspondant au nombre d'autres petits dans la portée dans le modèle linéaire 68
   a. Equation du modèle utilisée 68
   b. Vérification de la linéarité 68
   c. Résultats de l'analyse 69

C. EFFET BRUT DU SEXE SUR LA CROISSANCE DU MICROCÈBE 71

D. EFFET BRUT DE LA PARITÉ DE LA MÈRE SUR LA CROISSANCE DU MICROCÈBE 72

E. AJUSTEMENT : PRISE EN COMPTE DE FACTEURS DE CONFUSION POTENTIELS 72

IV. DISCUSSION 75

CONCLUSION 79

TABLE DES ANNEXES

ANNEXE 1: ANALYSE DE L'EFFET BRUT DU SEXE DES JEUNES PANTHÈRES SUR LEUR CROISSANCE PONDÉRALE 86

ANNEXE 2 : CAPTURE D'ÉCRAN DES RÉSULTATS DE L'ANALYSE DE L'EFFET BRUT DU SEXE SUR LA CROISSANCE DES JEUNES PANTHÈRES FOURNIS PAR LE LOGICIEL 87

ANNEXE 3 : NOTIONS DE VARIANCE, COVARIANCE ET DÉTERMINATION DE LA SIGNIFICATIVITÉ D'UNE SOMME 88

ANNEXE 4 : ANALYSE DE L'EFFET BRUT DE LA DURÉE DE LA MALADIE DES JEUNES PANTHÈRES SUR LEUR CROISSANCE PONDÉRALE 90

ANNEXE 5 : ANALYSE AJUSTÉE SUR LES DEUX VARIABLES « SEXE » ET « LONG » DE LA CROISSANCE PONDÉRALE DES JEUNES PANTHÈRES 91

ANNEXE 6 : PROGRAMMES UTILISÉS POUR LA CRÉATION DU TABLEAU DE DONÉEES ET LA REPRÉSENTATION EN NUAGE DE POINTS 92

ANNEXE 7 : MACRO PERMETTANT DE TESTER LES DIFFERENTS AGES DE RUPTURE DE PENTE 93

ANNEXE 8 : EXEMPLE DE PROGRAMME PERMETTANT DE CONNAITRE LES POURCENTAGES DE MICROCÈBES DANS LES DIFFÉRENTES CATÉGORIES 94

ANNEXE 9 : PROGRAMMES PERMETTANT L'ANALYSE BRUTE DE L'EFFET DU NOMBRE D'AUTRES PETITS SUR LA CROISSANCE PONDÉRALE DES MICROCÈBES 95

ANNEXE 10 : PROGRAMME PERMETTANT L'ANALYSE BRUTE DE L'EFFET DU SEXE DU JEUNE MICROCÈBE SUR SA CROISSANCE PONDÉRALE 100

ANNEXE 11 : PROGRAMME PERMETTANT L'ANALYSE BRUTE DE L'EFFET DE LA PARITÉ (PRIMIPARE *VS* MULTIPARE) DE LA MÈRE DU JEUNE MICROCÈBE SUR LA CROISSANCE PONDÉRALE DE CE DERNIER 101

ANNEXE 12 : PROGRAMME PERMETTANT L'ANALYSE AJUSTÉE DE LA CROISSANCE PONDÉRALE DES JEUNES MICROCÈBES 102

TABLE DES FIGURES

Figure 1 : Dessin d'une panthère nébuleuse (source : Handbook of the Mammals of the World, Volume 1 (Wilson et Mittermeier 2009)). 13

Figures 2a et 2b : Panthères nébuleuses mâles. (Photographie personnelle). 14

Figure 3 : Jeune femelle panthère nébuleuse de 4 mois et demi. (Photographie personnelle). 16

Figure 4 : Pesée d'une panthère nébuleuse femelle de 4 mois et demi. (Photographie personnelle). 17

Figure 5 : Capture d'écran d'un extrait de la base de données sous le logiciel SAS. 19

Figure 6 : Graphique en nuage de points des poids des 49 jeunes panthères dans leurs 90 premiers jours de vie. 20

Figure 7 : Graphique représentant les courbes individuelles de croissance de trois panthères nébuleuses durant leurs 90 premiers jours de vie. 21

Figure 8 : Graphique représentant les moyennes de poids et leurs écarts types pour tous les temps chez les mâles et les femelles panthères nébuleuses au cours des 90 premiers jours de vie. 22

Figure 9 : Graphique des courbes de croissance schématiques pour un modèle linéaire sans rupture de pente pour chacun des deux sexes, présentant les coefficients du modèle. 29

Figure 10 : Graphique des courbes de croissance schématiques pour un modèle linéaire à une rupture de pente pour chacun des deux sexes, présentant les coefficients du modèle. 30

Figure 11 : Dessin d'un microcèbe murin (source : Handbook of the Mammals of the World, Volume 3 (Wilson, Mittermeier 2009)) 49

Figure 12 : Photographie d'un microcèbe adulte de l'élevage de Brunoy (91) (Crédit photographie : Eric Gueton, MNHN) 51

Figure 13 : Jeunes microcèbes âgés de moins d'une semaine manipulés avec précaution pour la pesée. (Crédit photographie : Eric Gueton, MNHN) 52

Figure 14 : Graphique en nuage de points des poids de 150 microcèbes murins de l'étude au cours de leurs 100 premiers jours de vie. 55

Figure 15 : Graphique schématisant la modélisation des données avec une équation linéaire autorisant une rupture de pente à 46 jours. 56

Figure 16 : Graphique schématisant la modélisation non linéaire des données avec une équation de type Gompertz. 57

Figure 17: Allure de croissance des petits uniques ou non uniques en analyse brute avec le modèle linéaire. 65

Figure 18 : Graphique représentant l'allure de croissance des petits uniques ou non uniques en analyse brute avec le modèle de Gompertz. 67

Figure 19 : Graphique représentant les valeurs des coefficients α_i et λ_i en fonction des 3 classes de la variable quantitative "nb_autres_petits". 69

Figure 20 : Graphique représentant l'allure de croissance des petits microcèbes en fonction pour différentes tailles de portée en analyse brute avec le modèle linéaire. 70

TABLE DES TABLEAUX

Tableau 1 : Composition de l'échantillon de panthères nébuleuses pour l'analyse de l'association brute entre le sexe et la croissance. 35

Tableau 2 : Résultats de l'analyse statistique de l'association brute entre le sexe et la croissance chez les panthères nébuleuses. 36

Tableau 3 : Composition de l'échantillon de panthères nébuleuses pour l'analyse de l'association brute entre la durée de la maladie et la croissance. 38

Tableau 4 : Résultats de l'analyse statistique de l'association brute entre la durée de la maladie et la croissance chez les panthères nébuleuses. 38

Tableau 5 : Résultats de l'analyse statistique ajustée sur le sexe et la durée de la maladie. 40

Tableau 6 : Comparaison de la prise en compte des effets aléatoires de l'individu, la portée et la mère sur les paramètres des deux modèles. 59

Tableau 7 : Présentation des avantages et inconvénients des deux types de modèles (modèle linéaire *versus* modèle deGompertz). 62

Tableau 8 : Effectifs utilisés lors de l'étude de la dynamique de croissance des microcèbes. 63

Tableau 9 : Pourcentages et effectifs de l'échantillon de 645 microcèbes suivant le nombre de petits dans la portée. 64

Tableau 10 : Estimations des paramètres du modèle linéaire données par le logiciel lors de l'analyse avec la variable binaire « petit_unique ». 65

Tableau 11 : Valeurs des GMQ pour les petits non uniques et les petits uniques lors de l'analyse brute. 66

Tableau 12 : Estimations des paramètres du modèle non linéaire données par le logiciel lors de l'analyse avec la variable binaire « petit_unique ». 66

Tableau 13 : Valeurs des estimations des paramètres du modèle non linéaire pour les deux groupes (petit unique versus petit non unique). 67

Tableau 14 : Estimations des paramètres du modèle linéaire données par le logiciel lors de l'analyse avec la variable quantitative correspondant au nombre d'autres petits dans la portée. 69

Tableau 15 : Valeurs des GMQ selon la composition de la portée, lors de l'analyse brute. 70

Tableau 16 : Estimations des paramètres du modèle linéaire données par le logiciel lors de l'analyse avec la variable binaire « male». 71

Tableau 17 : Valeurs des GMQ obtenus lors de l'analyse brute chez les mâles et les femelles. 71

Tableau 18 : Estimations des paramètres du modèle linéaire données par le logiciel lors de l'analyse avec la variable binaire « multipare». 72

Tableau 19 : Estimations des paramètres du modèle linéaire données par le logiciel lors de l'analyse ajustée sur les variables d'intérêt principal. 73

Tableau 20 : Valeurs des GMQ obtenus pour les différentes catégories lors de l'analyse ajustée. 74

INTRODUCTION

La croissance pondérale de jeunes mammifères sauvages n'a été que très peu étudiée en comparaison avec celle des animaux de rente (mammifères ou oiseaux). Or, il est pertinent de s'y intéresser afin de mieux comprendre les espèces, c'est-à-dire obtenir un modèle qui donne une allure de la croissance dans telle ou telle espèce, en fonction de différents paramètres propres à l'individu (par exemple son sexe, la composition de la portée, la parité de sa mère) ou dépendant davantage de ses conditions de vie (milieu de vie, petit élevé avec les parents ou à la main par l'Homme, saison, …). La modélisation de la croissance peut également trouver un intérêt dans la problématique plus vaste qui s'interroge sur les conséquences de la dynamique de croissance sur les « performances » futures de l'individu en termes de reproduction ou de longévité par exemple.

L'objectif de cette modélisation est donc de trouver un modèle adapté aux données de poids que l'on a recueillies au cours de la croissance du jeune mammifère sauvage.

Les études longitudinales, c'est-à-dire dans lesquelles les variables sont mesurées de façon répétée dans le temps pour un même individu, sont largement répandues et fournissent souvent une meilleure vue d'ensemble que les études transversales (qui se focalisent sur une situation à un instant donné). Elles permettent de s'intéresser à des variables individuelles changeant au cours du temps (telles que le poids, la taille, la quantité sérique d'une hormone …), ainsi qu'à la relation entre ces variables et les caractéristiques de l'individu (sexe, espèce, mode de vie, …). Les modèles à effets mixtes, qui seront présentés dans ce document, sont devenus très populaires pour répondre à de telles questions.

Conduire une analyse de données longitudinales avec des modèles à effets mixtes requiert de maitriser la modélisation, en se confrontant parfois à des problèmes de non-convergence ou de non-adéquation du modèle aux données. La validité du modèle ainsi que la confiance que l'on peut lui accorder doivent systématiquement être discutées (Cheng *et al.,* 2010).

Cette citation de Cheng *et al.* comporte un certain nombre de mots « barbares » que nous serons amenés à définir et à réutiliser. Le terme de « modélisation » suffit à lui seul à effrayer une majorité de personnes qui souhaitent effectuer des études longitudinales. Mais pas de panique, il n'y a besoin ni d'être un « crack » en mathématiques ni d'être un programmeur informatique pour comprendre ce qui suit !

Le but de la première partie de cette thèse est de présenter, pas à pas, la marche à suivre dans le cadre d'une étude longitudinale concernant la croissance pondérale, les questions à se poser et les écueils à éviter, en s'appuyant sur un exemple assez simple de modélisation linéaire mixte de la croissance pondérale de jeunes panthères nébuleuses (*Neofelis nebulosa*) dans leurs 90 premiers jours de vie.

La seconde partie présentera un retour d'expérience sur l'analyse de la croissance pondérale de microcèbes murins (*Microcebus murinus*) à l'aide de deux modèles différents, dans leurs 100 premiers jours de vie.

- PARTIE A -

THÉORIE ET ÉTAPES DE CONSTRUCTION D'UN MODÈLE EN ADÉQUATION AVEC DES DONNÉES DE POIDS DE JEUNES PANTHÈRES NÉBULEUSES COLLECTÉES DANS LES 90 PREMIERS JOURS DE VIE

Cette partie théorique s'appuie sur un exemple concret et présente de façon aussi pédagogique que possible la marche à suivre dans le cadre d'une étude longitudinale comme celle de la croissance pondérale de jeunes animaux et permet de répondre aux questions suivantes :

- Comment recueillir et gérer les données de poids ?
- Quelles sont les équations à disposition pour modéliser la croissance pondérale ?
- Qu'est-ce que la non-indépendance des données et comment la prendre en compte ?
- Comment vérifier l'adéquation du modèle aux données ?
- Quels sont les logiciels à disposition pour la modélisation et quelles procédures utiliser ?

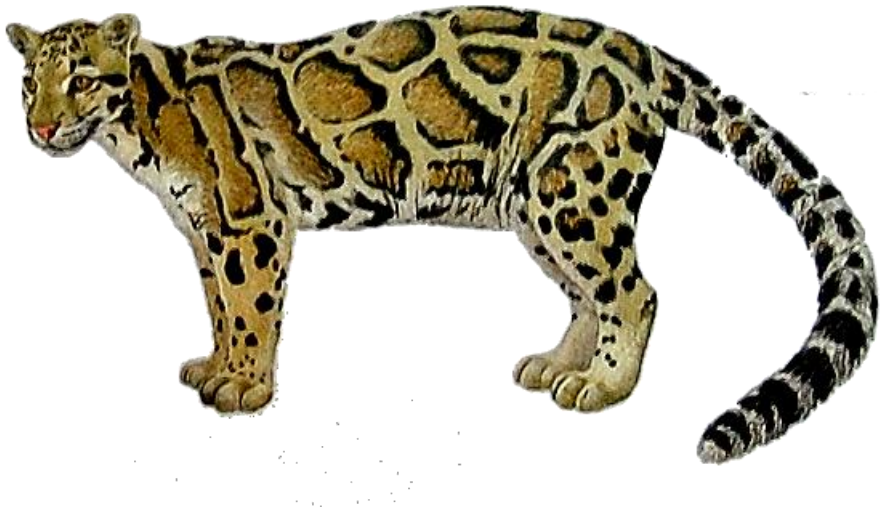
Les quelques résultats présentés ici sont issus d'une analyse partielle et simplifiée des données pondérales de jeunes panthères nébuleuses, dont le but unique est d'illustrer les notions théoriques introduites au fil de l'explication. Les analyses statistiques que j'ai réalisées par ailleurs sur ces mêmes données ont conduit à la rédaction d'un article qui a été récemment accepté pour publication (Nájera, in press), et dont je suis co-auteur. Les objectifs de cette étude étaient de créer des courbes de croissance pour des jeunes panthères nébuleuses nées en captivité et élevées à la main, d'analyser leur croissance, et d'estimer des différences ponctuelles de moyennes pondérales à différents temps.

I. Obtention et gestion des données : du recueil à la mise en forme

A. Présentation de l'espèce et des animaux étudiés

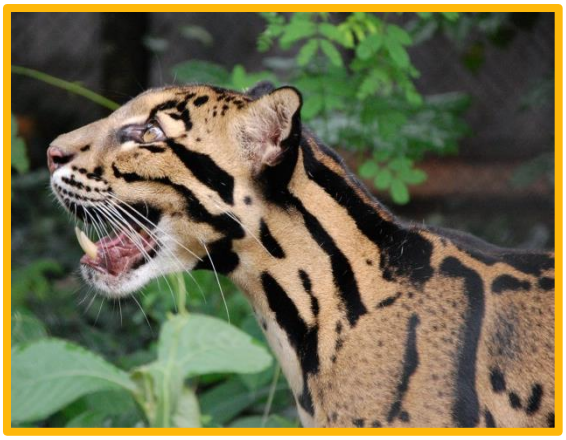
1. La panthère nébuleuse, prise en compte des particularités de l'espèce

Figure 1 : Dessin d'une panthère nébuleuse (source : Handbook of the Mammals of the World, Volume 1 (Wilson et Mittermeier, 2009)).



La panthère nébuleuse (« clouded leopard » en anglais), aussi appelée panthère longibande, a pour nom latin *Neofelis nebulosa*. Il s'agit de l'un des félins sauvages les plus charismatiques et les moins bien connus de l'Asie du Sud Est. Elle doit ses noms aux motifs de sa fourrure : certains ont retenu les taches en forme de nuages (Figure 1), d'autres les longues bandes noires, présentes en particulier au niveau du cou et de la face (Figures 2a et 2b).

Figures 2a et 2b : Panthères nébuleuses mâles. (Photographies personnelles)



La panthère nébuleuse est considérée comme un félin de taille moyenne et a acquis la dénomination de « plus petit des grands félins » (Sunquist, 2002). Cette espèce est classée « Vulnérable » (Sanderson *et al.*, 2008) sur la liste rouge de l'IUCN (International Union for Conservation of Nature) et figure à l'appendice 1 de la CITES (Convention on International Trade of Endangered species). La perte de l'habitat et le trafic illégal sont les principales menaces pesant sur cet animal à l'heure actuelle (Nowell, 2007).

On trouve des panthères nébuleuses à l'état sauvage au Népal (région sud himalayenne), au nord-est de l'Inde (région de l'Assam), au Bhoutan, au sud de la Birmanie, au sud de la Chine, au Vietnam, au Laos, au Cambodge, en Thaïlande et dans la péninsule malaisienne (Sunquist, 2002; Wilting *et al.*, 2007).

En milieu captif, la panthère longibande est connue pour être l'une des espèces de félins les plus difficiles à faire reproduire en raison du caractère agressif des mâles envers les femelles et de la mortalité élevée des jeunes (Fletchall, 2000; Howard *et al.*, 2007; Nájera *et al.*, 2011).

Bien que l'élevage à la main ne soit pas une pratique très courante pour les félins sauvages, il peut être avantageux d'élever les petits que la mère a tendance à négliger ou ceux qui montrent des signes de maladie ou de faiblesse (Edwards et Hawes, 1997)(Rivas *et al.*, 2009) afin d'assurer la sauvegarde de l'espèce. Les panthères nébuleuses sont des animaux timides qui ont tendance à être particulièrement stressés en milieu captif en raison du confinement dans des enclos souvent petits et de la proximité humaine. Elever les petits à la main pourrait donc, selon certains auteurs, être recommandé pour diminuer le stress des animaux, pallier le manque d'adaptabilité des individus à de nouvelles situations (changement de lieu ou de soigneurs par exemple), ou encore pour les animaux faisant partie de programmes éducatifs occasionnant des contacts avec l'humain (McGovern, 1998; Rivas *et al.*, 2009).

Enfin, les panthères longibandes élevées à la main semblent être en mesure de former des couples plus facilement, puisque l'introduction avec un potentiel partenaire peut alors être effectuée à un plus jeune âge (Fazio, 2010).

Il est tout de même nécessaire de rappeler que des méthodes autres que l'élevage à la main (par exemple l'enrichissement du milieu et l'entraînement médical) permettent de diminuer le stress chez les félins en général, et les panthères nébuleuses en particulier.

Le poids des panthères nébuleuses adultes varie entre 11 et 23 kg. Les mâles sont plus lourds que les femelles à l'âge adulte (Wilson et Mittermeier, 2009). La gestation dure en moyenne 88 à 95 jours et les portées comprennent en général deux ou 3 petits (valeurs extrêmes : 1 à 5 petits). Lorsqu'ils sont élevés par la mère, les petits commencent à manger de la nourriture solide entre 7 et 10 semaines d'âge mais peuvent continuer à téter jusqu'à 14 semaines d'âge. Ils atteignent leur maturité sexuelle entre 20 et 30 mois d'âge (Wilson et Mittermeier, 2009).

Les informations concernant la dynamique de croissance des panthères nébuleuses dans la littérature sont peu abondantes (Fellner, 1968 ; McGovern, 1998). Des gains moyens de poids ont été publiés, mais les analyses ne prenaient pas en compte l'influence de paramètres comme le sexe de l'individu ou la taille de la portée (Fletchall, 2000). Connaître l'allure de la croissance dans cette espèce pourrait être utilisé comme un outil diagnostique pour évaluer l'état de santé des petits en croissance et améliorer les taux de survie pendant la phase de sevrage (Binczik et Reindl, 1987; Wack *et al.*, 1991).

2. Jeunes panthères nébuleuses utilisées dans l'étude

Les jeunes panthères utilisées dans cette étude ont été élevées dans 3 centres de sauvegarde différents : 18 portées sont nées au « Khao Kheow Open Zoo », dans la province de Chonburi en Thaïlande, 3 portées sont issues du « Center for Species Survival » de l'institut de biologie de la conservation du Smithsonian à Front Royal dans l'Etat de Virginie aux Etats-Unis, et 3 portées sont nées à « Point Defiance Zoo & Aquarium » dans l'Etat de Washington aux Etats-Unis. La Figure 3 présente la photographie d'une jeune panthère élevée à la main à Khao Kheow Open Zoo en Thaïlande.

Tous les individus de cette étude ont été retirés à leur mère durant leurs 24 premières heures de vie afin d'être élevés à la main suivant un protocole d'alimentation établi au préalable par une équipe du Smithsonian. [1]

Les panthères recevaient quotidiennement 28 % de leur poids en nourriture, proposée plusieurs fois par jour. Le processus d'élevage à la main était divisé en 3 périodes. Pendant les 28 premiers jours de vie, les petits ne recevaient que du lait. Durant la deuxième phase, entre 29 et 42 jours de vie, ils recevaient du lait et des protéines animales (nourriture pour jeunes à base de poulet). Enfin à partir de 43 jours de vie de la viande (poulet ou bœuf) était introduite en quantité de plus en plus importante.

[1] Le Smithsonian (Washington, D.C.) est considéré comme une des plus grandes institutions de recherche au monde. Il abrite 9 centres de recherche, 19 musées et galeries ainsi qu'un parc zoologique.

Figure 3 : Jeune femelle panthère nébuleuse de 4 mois et demi. (Photographie personnelle)

B. Collecte des données

La collecte des données doit suivre un protocole de préférence établi à l'avance, afin de standardiser au maximum les prises de mesures. Pour recueillir des données de poids, il est préférable d'utiliser la même balance, ou tout du moins de vérifier que les balances utilisées sont calibrées de la même façon. Dans ce cas, l'opérateur n'a que peu d'importance car il s'agit d'effectuer une simple lecture. Quand on associe cette pesée à la mesure de la taille de l'animal par exemple (longueur totale, taille au garrot, tour de thorax, …), des variations plus importantes peuvent intervenir si plusieurs opérateurs effectuent les mesures, dans le cas où ils n'utilisent pas exactement les mêmes repères anatomiques sur l'animal par exemple.

Dans le cadre de l'étude de la croissance pondérale des jeunes panthères nébuleuses, 49 individus (29 mâles et 20 femelles) ont été pesés au cours de leurs 90 premiers jours de vie. La pesée était toujours effectuée le matin, avant le premier nourrissage de la journée.

Il est possible de manipuler facilement les panthères pour la pesée, même quand elles sont plus âgées (Figure 4).

Les mesures ainsi que plusieurs caractéristiques individuelles ont ensuite été saisies dans une base de données.

- On retiendra donc qu'il est préférable d'établir un protocole détaillé à l'avance, afin de limiter les biais de mesure. Ceci est d'ailleurs valable quel que soit le type d'étude menée.

Figure 4 : Pesée d'une panthère nébuleuse femelle de 4 mois et demi. (Photographie personnelle)

C. Élaboration d'une base de données

La bonne élaboration de la base de données conditionne le déroulement des analyses et la confiance accordée aux résultats. Dans le cas de l'analyse de la croissance pondérale, il ne s'agit que d'un tableau unique (éventuellement issu de la fusion de plusieurs tableaux), que l'on peut concevoir dans Excel par exemple. La saisie des données doit être vérifiée (ce qui n'est pas prévu automatiquement par Excel) : il ne doit y avoir ni doublon, ni donnée aberrante (par exemple un nombre bien trop grand dû à une erreur de frappe de virgule). De plus, une donnée manquante ne doit pas être saisie par un zéro (si la mesure de poids a été impossible au jour j, cela ne veut certainement pas dire qu'à ce jour-là, l'animal pesait « zéro gramme » !). Zéro est une valeur chiffrée qui ne sera pas du tout prise en compte par les logiciels de la même manière qu'une case laissée vide dans Excel par exemple. Il faudra donc vérifier dans la base de données que les cases pour lesquelles il n'y a pas eu de mesure apparaissent avec un point (« . ») ou une autre codification indiquant qu'elles sont bien prises en compte comme des données non renseignées.

D'autres logiciels, comme EpiData ou Access par exemple, permettent d'élaborer un masque de saisie de données qui donne un « cadre » lors de la saisie de données et permet de diminuer le nombre d'erreurs de saisie. Par exemple, il est possible de préciser que l'on souhaite coder le sexe de l'animal en binaire (« 0 » pour les femelles et « 1 » pour les mâles par exemple). Ainsi, si le chiffre saisi est différent de ces deux valeurs, un message d'erreur apparait et interdit à l'opérateur de poursuivre la saisie.

La Figure 5 ci-dessous présente un extrait de la base de données utilisée pour l'analyse de croissance pondérale de jeunes panthères nébuleuses. Une ligne correspond à une mesure pour un individu donné à un âge donné. Les données sont ici ordonnées par nom (« name »), c'est-à-dire que les 90 premières lignes du tableau concernent un même individu (une ligne par jour de mesure), puis les 90 suivantes un autre animal et ainsi de suite.

En colonne, on trouve évidemment le nom de l'individu (« name »), son âge au moment de la mesure (« age »), le poids correspondant (« weight »), son sexe (« sex ») codé 1 pour mâle et 0 pour femelle, et la taille de la portée dont il fait partie (« litter_size »). La variable binaire « GI_problems » indique si l'animal a rencontré des problèmes gastro-intestinaux (épisodes de diarrhée le plus souvent) au cours de ses 3 premiers mois de vie, codée 1 pour oui et 0 pour non. Les deux dernières colonnes sont renseignées par des valeurs non nulles si l'individu a effectivement été malade durant ses 3 premiers mois de vie. Elles permettent de répertorier le premier jour de maladie (« first_day_sick ») et la durée des symptômes (« length_sick »).

- On retiendra donc que les données se présentent avec les individus en lignes (avec une ligne par jour de mesure par animal) et les variables en colonnes, et que le soin apporté à l'élaboration de la base de données et à la vérification de leur saisie conditionne la fiabilité des résultats issus de l'analyse.

Figure 5 : Capture d'écran d'un extrait de la base de données sous le logiciel SAS.

	Name	Age	Weight	Sex	GI_problems	Litter_size	first_day_sick	length_sick
2139	BISSAH	68	2080	1	0	3	.	0
2140	BISSAH	69	2140	1	0	3	.	0
2141	BISSAH	70	2140	1	0	3	.	0
2142	BISSAH	71	2140	1	0	3	.	0
2143	BISSAH	72	2080	1	0	3	.	0
2144	BISSAH	73	2060	1	0	3	.	0
2145	BISSAH	74	2120	1	0	3	.	0
2146	BISSAH	75	2140	1	0	3	.	0
2147	BISSAH	76	2180	1	0	3	.	0
2148	BISSAH	77	2220	1	0	3	.	0
2149	BISSAH	78	2200	1	0	3	.	0
2150	BISSAH	79	2220	1	0	3	.	0
2151	BISSAH	80	2300	1	0	3	.	0
2152	BISSAH	81	2260	1	0	3	.	0
2153	BISSAH	82	2320	1	0	3	.	0
2154	BISSAH	83	2340	1	0	3	.	0
2155	BISSAH	84	2380	1	0	3	.	0
2156	BISSAH	85	.	1	0	3	.	0
2157	BISSAH	86	.	1	0	3	.	0
2158	BISSAH	87	.	1	0	3	.	0
2159	BISSAH	88	.	1	0	3	.	0
2160	BISSAH	89	2580	1	0	3	.	0
2161	KELLY	0	280.1	1	1	3	69	9
2162	KELLY	1	294.3	1	1	3	69	9
2163	KELLY	2	307.5	1	1	3	69	9
2164	KELLY	3	331.7	1	1	3	69	9
2165	KELLY	4	349.7	1	1	3	69	9
2166	KELLY	5	368.3	1	1	3	69	9
2167	KELLY	6	379.6	1	1	3	69	9
2168	KELLY	7	410.6	1	1	3	69	9
2169	KELLY	8	436.7	1	1	3	69	9
2170	KELLY	9	460	1	1	3	69	9
2171	KELLY	10	480	1	1	3	69	9
2172	KELLY	11	520	1	1	3	69	9
2173	KELLY	12	560	1	1	3	69	9

Sur cette ligne, la 2139ème du fichier de données, on peut lire que le mâle panthère prénommé Bissah pesait 2080 grammes à l'âge de 68 jours. Par ailleurs il fait partie d'une portée de 3 petits et n'a présenté aucun problème digestif au cours de ses 3 premiers mois de vie.

Ces cases contiennent un « . » car Bissah n'a jamais été malade. Par conséquent le « premier jour de maladie » ne peut pas être renseigné. Si la valeur « 0 » était présente, cela signifierait qu'il a été malade dès son premier jour de vie, ce qui est complètement différent !

Cases pour lesquelles les données de poids sont manquantes.

Cette ligne est la première concernant Kelly, un mâle panthère issu également d'une portée de 3 petits. A la naissance, on peut lire qu'il pesait 280,1 grammes. La colonne « GI_problems » nous indique que Kelly a été malade au cours de ses 3 premiers mois de vie. On lit dans les deux colonnes les plus à droite que cet animal a été malade durant 9 jours à partir de son 69ème jour de vie.

C'est un exemple de saisie à vérifier dans le tableau de données avant de se lancer dans les analyses : chaque fois que le chiffre « 1 » apparait dans la colonne « GI_problems », il faudra s'assurer que les deux colonnes de droite contiennent des valeurs différentes de zéro ou «.».

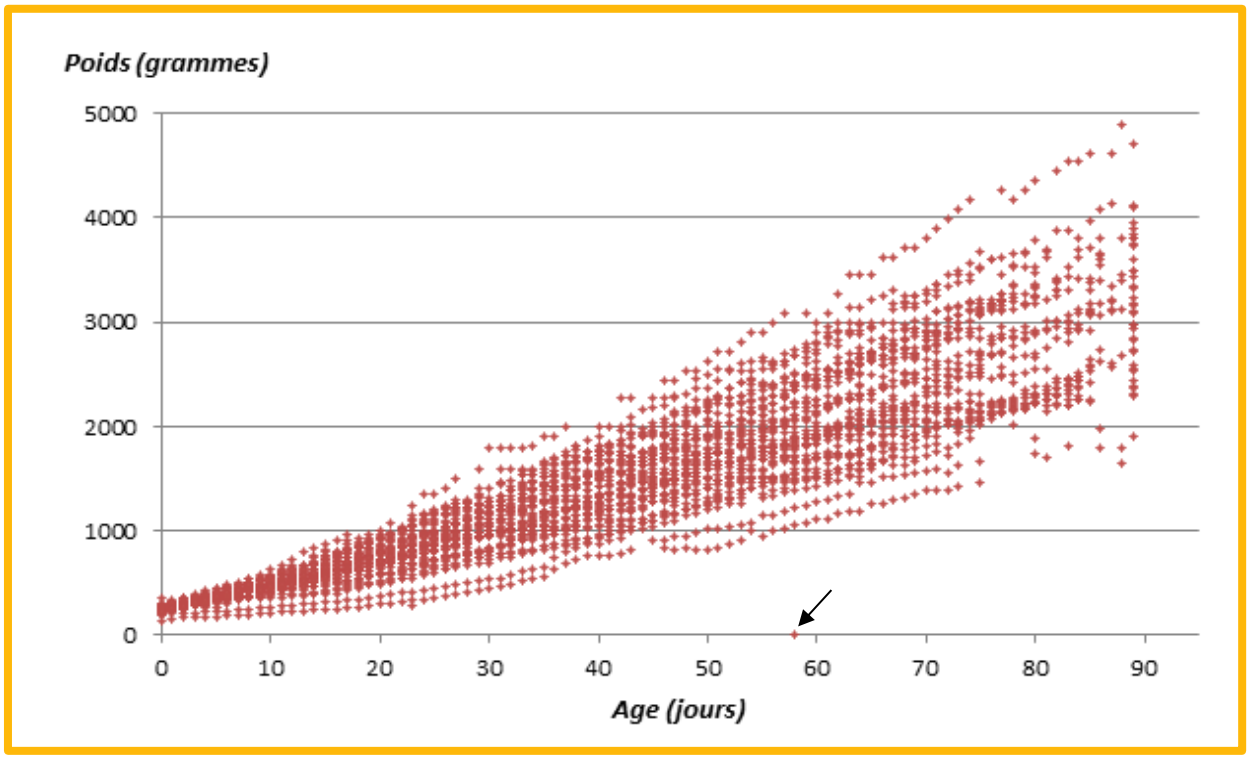
II. Représentations graphiques des données

Différentes représentations graphiques sont disponibles pour permettre de visualiser la dispersion des données et ainsi potentiellement orienter le choix du modèle par la suite.

A. Nuages de points

Pour avoir une idée de l'allure générale des données, il est possible de tracer un nuage de points (Figure 6), reprenant dans notre cas toutes les mesures de poids à tous les temps (âges des individus : de 0 à 90 jours). On remarque que pour les 49 animaux de la base de données, certains points se superposent. Si le nombre d'animaux de l'étude est trop important, on pourra choisir de ne sélectionner qu'un petit nombre d'entre eux afin que le graphique en nuage de points soit lisible.

Figure 6 : Graphique en nuage de points des poids des 49 jeunes panthères dans leurs 90 premiers jours de vie.

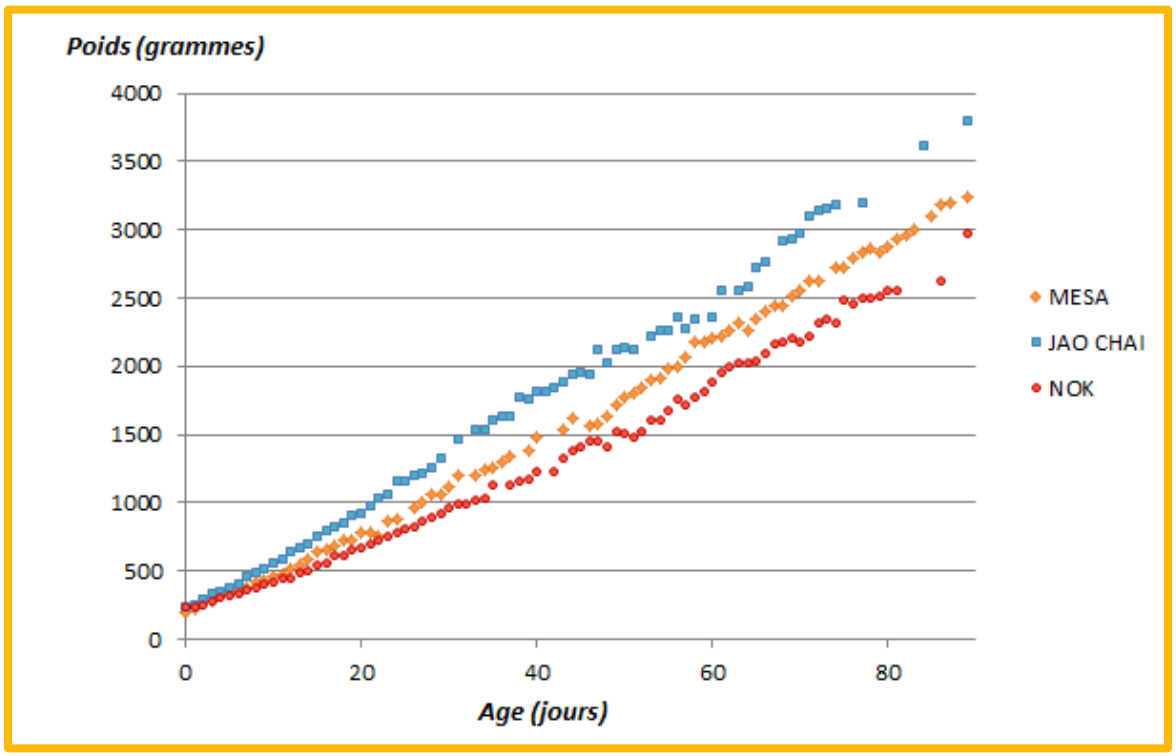


Le graphique en nuage de points peut également permettre de vérifier l'absence de valeurs aberrantes issues d'une mauvaise saisie de données. Sur la figure ci-dessus, on peut constater la présence d'une erreur de saisie à 58 jours (cf. flèche sur la Figure 6), très certainement une donnée de poids manquante marquée par un zéro. Il faudra chercher cette erreur dans le tableau de données et la corriger avant de débuter l'analyse.

B. Courbes individuelles

Pour avoir une idée des trajectoires de courbes et se familiariser avec les données, on peut également tracer des courbes individuelles au hasard comme présenté dans la Figure 7 ci-dessous.

Figure 7 : Graphique représentant les courbes individuelles de croissance de trois panthères nébuleuses durant leurs 90 premiers jours de vie.



C. Moyennes ± écarts types ou intervalles de confiance à 95 %

Il est également possible de calculer, séparément pour certaines catégories (ici les mâles et les femelles), les moyennes pour chaque jour, en leur associant des barres d'erreur de hauteurs égales aux écarts types (notés « SD » dans l'échantillon, pour « standard deviation ») ou aux intervalles de confiance à 95 %.

L'écart type (SD) permet de visualiser la variabilité du caractère mesuré : si la distribution du caractère quantitatif suit une loi normale, alors deux tiers des mesures des individus à un temps donné sont contenues dans l'intervalle de ± SD. L'écart type quantifie donc la dispersion des données de l'échantillon.

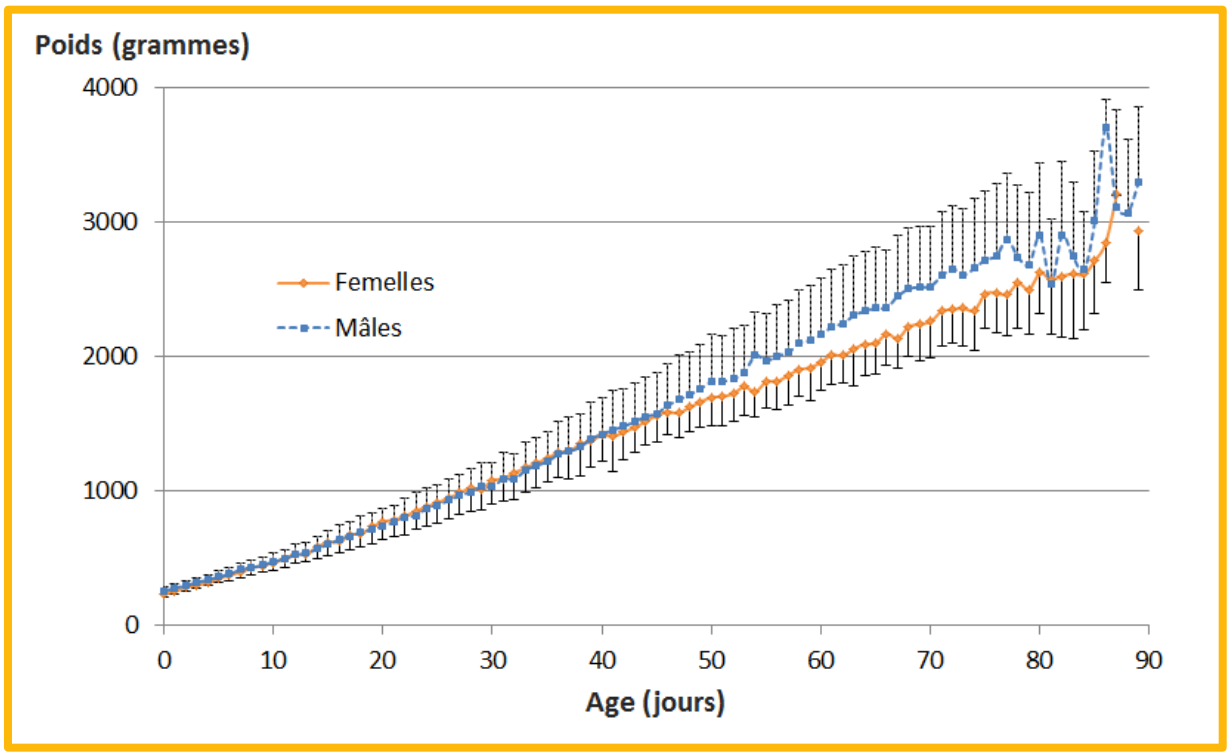
L'intervalle de confiance à 95 % de chaque moyenne, quant à lui, est utilisé pour donner une indication de la vraie valeur de la moyenne dans la réalité (dans la population cible qu'il faut avoir définie au préalable, en vérifiant l'absence de biais d'échantillonnage et de mesure). Pour rappel, l'intervalle de confiance s'exprime comme suit s'il y a plus de 30 mesures à un temps donné :

IC 95 % = moyenne ± 1,96 *SE (avec SE « standard error » = SD / $\sqrt{n}$ avec n le nombre de mesures au temps t donné)

Il y a alors 95 % de chances pour que la vraie valeur de la moyenne dans la population cible soit comprise dans cet intervalle de confiance.[2]

La Figure 8 ci-dessous présente deux courbes de moyennes avec les écarts types correspondant à chaque temps de mesure. Pour une meilleure lisibilité, les barres d'erreurs ont été mises d'un seul côté du tracé puisque les courbes sont relativement rapprochées dans le cas présenté.

Figure 8 : Graphique représentant les moyennes de poids et leurs écarts types pour tous les temps chez les mâles et les femelles panthères nébuleuses au cours des 90 premiers jours de vie.



↳ Ces graphiques permettent de visualiser les données brutes et/ou leur dispersion par rapport à la moyenne. Il est possible de tracer un graphique par « catégorie d'individus », par exemple un graphique pour les mâles et un pour les femelles, et ainsi d'avoir une première idée des tendances qu'il faudra mettre en évidence lors de l'analyse par modélisation. Dans le cas présent, les mâles semblent grandir plus vite que les femelles, au moins à partir d'une cinquantaine de jours de vie, mais seule une analyse statistique permet de savoir si ces différences sont significatives.

[2] Dans le cas où il y a moins de 30 mesures, il ne faut plus utiliser la valeur de « 1,96 », mais prendre la valeur d'un $t_{2,5\%}$ de Student dans les tables de Student (valeur dépendant du nombre de mesures). L'interprétation de cet intervalle de confiance reste bien entendu identique à celle énoncée ci-dessus.

A partir des allures des courbes tracées lors de l'étape préliminaire que nous venons de décrire, il est également possible d'avoir une idée du modèle mathématique à utiliser lors de l'étape suivante : la modélisation.

III. Choix du modèle

Une phrase de Box et Draper (Box et Draper, 1987), reprise par Cheng dans son article visant à expliquer comment construire un modèle linéaire « suffisamment bon » (Cheng *et al.,* 2010), résume assez bien la situation : « fondamentalement, tous les modèles sont faux, mais certains sont utiles (…) ». Voilà qui est rassurant ! Il faut en effet garder à l'esprit qu'en aucun cas le modèle ne sera capable de refléter exactement la réalité. Le but de la modélisation est de trouver une expression du modèle qui se rapproche le plus de la réalité (c'est-à-dire avec la meilleure adéquation possible aux données), tout en restant interprétable de façon simple, ce dernier point étant fondamental. Les paramètres du modèle devront faire référence à une caractéristique visuelle de la courbe correspondant au modèle pour que leur interprétation soit facile. Ainsi, un modèle « suffisamment bon » peut très bien convenir. Un exemple donné par Cheng est le suivant : on peut imaginer que le « modèle exact » dans la population puisse être non linéaire, alors que la modélisation basée sur les données disponibles d'un échantillon de cette population ne pourrait être effectuée que de façon linéaire. Un tel modèle linéaire pourrait tout de même être « suffisamment bon » pour approcher le modèle exact. Le modèle exact pour la population entière peut être plus complexe que nécessaire pour décrire la sous population échantillonnée. Parmi les modèles théoriquement utilisables sur ces données de l'échantillon, il en existe forcément un qui est le meilleur (Cheng *et al.,* 2010). Cependant, il y a sans doute plusieurs modèles qui en sont assez proches pour être considérés comme « suffisamment bons » pour répondre aux objectifs de l'analyse.

- Il ne faut donc jamais perdre de vue que le modèle est un outil pour répondre aux questions que l'on se pose.

Trouver un modèle « suffisamment bon » nécessite une bonne analyse de données, en ne perdant aucune information disponible dans la base de données. Il ne faut *a priori* pas extrapoler les résultats du modèle sur des plages de valeurs qui ne figurent pas dans l'échantillon. C'est pourquoi il faudra systématiquement, lorsque l'on teste un modèle, afficher le graphique correspondant à l'équation avec les paramètres estimés et le comparer aux courbes individuelles obtenues avec les données brutes.

Dans notre exemple d'étude sur les panthères nébuleuses, les représentations graphiques permettent de se rendre compte que les données pourraient être assez bien modélisées par des droites ou des segments de droites (il semble en effet possible de tracer, sur un intervalle de temps défini – par exemple 20 jours de vie –, une droite passant par un grand nombre de points). Un modèle utilisant des équations de droites est dit « modèle linéaire », et on dirait alors que le poids est associé linéairement au temps.

Cette partie III a pour vocation de décrire les modèles linéaires (et leur utilisation dans la modélisation de la croissance pondérale des jeunes panthères), puis présenter des exemples de modèles non-linéaires et enfin proposer une méthode destinée à tester l'adéquation des modèles au données.

A. Modèles utilisant des équations de droites : les modèles linéaires

1. Théorie mathématique sur les modèles linéaires (Diggle, 1994)

En général, on utilise des lettres capitales pour représenter les variables ou les matrices aléatoires, et des lettres minuscules pour les observations spécifiques. Les scalaires et les matrices sont en style normal, les vecteurs sont écrits en gras.

Y_{ij} représente une variable réponse et $\mathbf{x}_{ij}$ un vecteur de longueur p (p-vecteur) des p variables explicatives observées aux temps t_{ij} pour une observation j = 1, …, n_i sur le sujet i = 1, …, m. La moyenne et la variance de Y_{ij} sont représentées respectivement par $E(Y_{ij}) = \mu_{ij}$ et $Var(Y_{ij}) = v_{ij}$. Les mesures répétées pour un individu i sont collectées dans un n_i –vecteur, noté $\boldsymbol{Y_i} = (Y_{i1}, \ldots, Y_{ini})$, avec pour moyenne $E(\boldsymbol{Y_i}) = \boldsymbol{\mu_i}$ et pour matrice de covariance de dimension (n_i x n_i) $Var(\boldsymbol{Y_i}) = V_i$, où l'élément jk de V_i est la covariance entre Y_{ij} et Y_{ik}, définie par $Cov(Y_{ij}, Y_{ik}) = v_{ijk}$. On appelle R_i la matrice de dimension (n_i x n_i) de $\boldsymbol{Y_i}$.

$\boldsymbol{Y} = (\boldsymbol{Y_1}, \ldots, \boldsymbol{Y_m})$ fait référence aux réponses observées de toutes les unités d'étude (les individus).

La plupart des analyses longitudinales sont basées sur un modèle linéaire d'équation :

$$Y_{ij} = \beta_1 x_{ij1} + \beta_2 x_{ij2} + \ldots + \beta_p x_{ijp} + \varepsilon_{ij}$$

Où $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)$ est un p-vecteur de coefficients de régression quantifiant l'association entre chacune des p-expositions et la variable réponse. Ces coefficients sont estimés par le logiciel ; ils quantifient l'association entre les variables explicatives et la variable réponse Y.

$\boldsymbol{\varepsilon_{ij}}$ une variable aléatoire de moyenne nulle (appelée « résidu ») représentant l'écart entre la réponse observée et celle prédite par le modèle. $\boldsymbol{\varepsilon_{ij}}$ est donc distribuée selon une loi Normale de paramètres $(0, \sigma^2)$, où σ quantifie la variabilité des résidus autour de la valeur prédite.

x_{ijl} (l variant de 1 à p) est la valeur de l'exposition l pour le sujet i à sa mesure j. Très souvent par convention $x_{ij1} = 1$ pour tous les i et tous les j, et β_1 est alors le terme d'ordonnée à l'origine du modèle linéaire.

En notation matricielle, l'équation de régression pour le i-ème sujet prend donc la forme :

$$\boldsymbol{Y_i} = X_i.\boldsymbol{\beta} + \boldsymbol{\varepsilon_i}$$

Où X_i est une matrice de dimension (n_i x p) et $\boldsymbol{\varepsilon_i} = (\boldsymbol{\varepsilon_{i1}}, \ldots, \boldsymbol{\varepsilon_{ini}})$.

Application

Afin de rendre tout ceci plus concret, prenons l'exemple de deux bébés panthères : *i* varie donc dans ce cas de 1 à m, avec m=2.

Ces deux petits sont de sexes différents. La femelle sera considérée comme « individu 1 » et le mâle comme « individu 2 ».

On suppose que chacun des deux petits (la femelle = individu 1 et le mâle = individu 2) a été pesé 3 fois : à la naissance (0 jours), à 45 jours de vie et à 90 jours de vie. *j* varie donc de 1 à 3 pour l'individu 1 et pour l'individu 2. Donc *j* varie de 1 à n_1=3 pour la femelle et de 1 à n_2=3 pour le mâle.

Les variables explicatives de ce modèle sont le temps et le sexe du petit. Tous les termes des premières colonnes des matrices X sont égaux à 1 par convention comme décrit précédemment : c'est-à-dire que tous les termes du vecteur $\mathbf{x_1}$ sont égaux à 1.

Notons $\mathbf{x_2}$ le vecteur correspondant au temps (prenant ici 3 valeurs : 0, 45 ou 90 jours) et $\mathbf{x_3}$ le vecteur correspondant au sexe (prenant la valeur 0 si le petit est une femelle, 1 si le petit est un mâle).

NB : On admettra ici (bien que ce ne soit sans doute pas vrai en pratique) qu'il n'y a aucun terme d'interaction entre les expositions « temps » et « sexe », afin de ne pas compliquer la démonstration. La notion de terme d'interaction sera explicitée dans la partie 2.

Les matrices X_1 et X_2 (regroupant les variables explicatives pour la femelle et le mâle, respectivement) s'écrivent donc comme suit :

$$X_1 = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 45 & 0 \\ 1 & 90 & 0 \end{bmatrix} \quad \text{et} \quad X_2 = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 45 & 1 \\ 1 & 90 & 1 \end{bmatrix}$$

$\boldsymbol{Y} = (\boldsymbol{Y_1}, \boldsymbol{Y_2})$ fait référence aux mesures observées des poids des deux panthères.

Si l'on suppose que les poids de la 1ère panthère à 0, 45, et 90 jours de vie sont respectivement de 225, 1250, et 2810 grammes, et si l'on suppose que les poids de la 2ème panthère à 0, 45, et 90 jours de vie sont respectivement de 260, 1530, et 3120 grammes, alors les deux vecteurs $\boldsymbol{Y_1}$ et $\boldsymbol{Y_2}$ s'écrivent comme suit :

$$\boldsymbol{Y_1} = \begin{pmatrix} 225 \\ 1250 \\ 2810 \end{pmatrix} \quad \text{et} \quad \boldsymbol{Y_2} = \begin{pmatrix} 260 \\ 1530 \\ 3120 \end{pmatrix}$$

On peut alors écrire les équations correspondantes pour les deux individus :

- Pour la femelle :

$$\boldsymbol{Y_1} = X_{1.}\boldsymbol{\beta} + \boldsymbol{\varepsilon_1}$$

C'est-à-dire : $$\begin{pmatrix} 225 \\ 1250 \\ 2810 \end{pmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 45 & 0 \\ 1 & 90 & 0 \end{bmatrix} . \begin{pmatrix} a \\ b \\ c \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \end{pmatrix}$$

« a » correspond au premier terme du vecteur $\boldsymbol{\beta}$, et aurait aussi pu être noté β_1.

« b » correspond au second terme du vecteur $\boldsymbol{\beta}$, et aurait aussi pu être noté β_2.

Enfin, « c » correspond au troisième terme du vecteur $\boldsymbol{\beta}$, et aurait aussi pu être noté β_3.

Après multiplication des deux matrices (que nous avons tous faite un jour dans notre vie !), on obtient alors l'équation suivante :

$$\begin{pmatrix} 225 \\ 1250 \\ 2810 \end{pmatrix} = \begin{pmatrix} a \\ a + 45b \\ a + 90b \end{pmatrix} + \begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \end{pmatrix} \qquad (E_1)$$

- <u>Pour le mâle</u>:

$$\boldsymbol{Y_2} = X_2 . \boldsymbol{\beta} + \boldsymbol{\varepsilon_2}$$

C'est-à-dire : $$\begin{pmatrix} 260 \\ 1530 \\ 3120 \end{pmatrix} = \begin{bmatrix} 1 & 0 & 1 \\ 1 & 45 & 1 \\ 1 & 90 & 1 \end{bmatrix} . \begin{pmatrix} a \\ b \\ c \end{pmatrix} + \begin{pmatrix} \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \end{pmatrix}$$

« a » correspond au premier terme du vecteur $\boldsymbol{\beta}$, et aurait aussi pu être noté β_1.

« b » correspond au second terme du vecteur $\boldsymbol{\beta}$, et aurait aussi pu être noté β_2.

Enfin, « c » correspond au troisième terme du vecteur $\boldsymbol{\beta}$, et aurait aussi pu être noté β_3.

Notons que les valeurs « a », « b », et « c » ci-dessous sont égales aux valeurs de « a », « b », et « c » de la panthère femelles : ce sont les paramètres du modèle qui sont tels que le modèle « fonctionne » bien pour tous les individus ! Les termes du vecteur $\boldsymbol{\varepsilon_i}$ sont justement là pour tolérer un écart entre les données observées, et celles prédites par le modèle justement avec ces trois valeurs « a », « b », et « c ».

On obtient alors l'équation suivante :

$$\begin{pmatrix} 260 \\ 1530 \\ 3120 \end{pmatrix} = \begin{pmatrix} a + c \\ a + 45b + c \\ a + 90b + c \end{pmatrix} + \begin{pmatrix} \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \end{pmatrix} \qquad (E_2)$$

La résolution des équations (E_1) et (E_2) conduira le logiciel à fournir les résultats pour les paramètres a, b et c qui forment le vecteur $\boldsymbol{\beta}$ ainsi que les résultats pour la matrice des résidus, comprenant $\boldsymbol{\varepsilon_1}$ et $\boldsymbol{\varepsilon_2}$.

- Cet exemple était destiné à expliciter la partie très théorique sur les modèles linéaires. Nous verrons dans la partie suivante qu'en pratique, et c'est tant mieux, il n'y a pas besoin d'écrire les matrices et les vecteurs. Toutefois, la compréhension du fonctionnement global est essentielle, même si les applications que nous allons voir permettent de n'écrire qu'une équation simple, en laissant le logiciel l'appliquer *de facto* à tous les sujets pour tous les temps et toutes les expositions.

2. Modélisation linéaire de la croissance pondérale chez les panthères nébuleuses

Dans le cas de l'étude de la croissance des panthères nébuleuses, les représentations graphiques des données permettent d'orienter le choix vers un modèle linéaire. Afin que l'adéquation d'un modèle linéaire aux données soit bonne, il est possible de lui autoriser une ou plusieurs ruptures de pentes. Cette partie présente les équations, représentations graphiques et explications associées à 2 types de modèles linéaires : un modèle sans rupture de pente et un modèle à une rupture de pente.

Par souci de simplicité, nous considèrerons dans les deux sous-parties suivantes seulement deux variables explicatives : le temps et le sexe du petit.

L'interaction entre les deux variables explicatives utilisées sera cette fois prise en compte dans le modèle. On dit qu'il y a interaction entre deux variables explicatives si l'association entre une variable explicative et la variable dépendante (aussi appelée variable réponse, représentée ici par le poids) n'est pas la même selon les valeurs prises par l'*autre* variable explicative. En épidémiologie, ceci correspond à la notion de modification d'effet d'une variable explicative sur la variable réponse par une autre variable explicative. Afin de prendre en compte une interaction potentielle entre les variables explicatives (ici « temps » et « sexe »), il est nécessaire d'introduire un terme d'interaction dans l'équation, défini comme le produit des deux variables (noté ici « t.sexe » et de valeur t * sexe). Si les résultats de l'analyse statistique donnent une valeur de coefficient non nulle devant ce terme d'interaction, alors il existe bel et bien une interaction numérique dans l'échantillon entre les deux variables explicatives (si ce terme est de plus significativement différent de zéro, alors on pourra écrire qu'il y a de grandes chances pour cette interaction soit réelle dans la population). Dans notre cas, cela reviendrait à dire que l'effet du temps sur le poids (c'est-à-dire la dynamique de croissance) ne serait pas le même selon le sexe de l'animal.

La démarche présentée ci-après est celle pour le sexe du petit, qui est une variable de type binaire. Cette démarche est identique avec un autre type de variable explicative (comme par exemple une variable en 3 classes ou plus, ou une variable quantitative).

a. Modèle sans rupture de pente

L'équation du modèle linéaire sans rupture de pente dans le cas présent s'écrit comme suit :

$$\textbf{Poids (t, sexe) = a + b.t + c.sexe + d.t.sexe + } \varepsilon$$

Il est ainsi possible de réécrire cette expression pour chacun des deux sexes :

Pour les femelles : Poids (t, sexe=0) = a + b.t **(f_1)**

« a » représente donc le poids à la naissance (quand t=0) pour les femelles.

« b » représente la pente de la courbe de croissance des femelles, c'est-à-dire leur gain moyen quotidien de poids (GMQ).

Pour les mâles : Poids (t, sexe=1) = a + c + (b+d).t **(m_1)**

« a+c » représente donc le poids à la naissance (quand t=0) pour les mâles.

« b+d » représente la pente de la courbe de croissance des mâles, c'est-à-dire leur gain moyen quotidien de poids (GMQ).

En effectuant la soustraction (m_1) – (f_1), on obtient :

Poids (t, sexe=1) - Poids (t, sexe=0) = c + d.t **(m_1) – (f_1)**

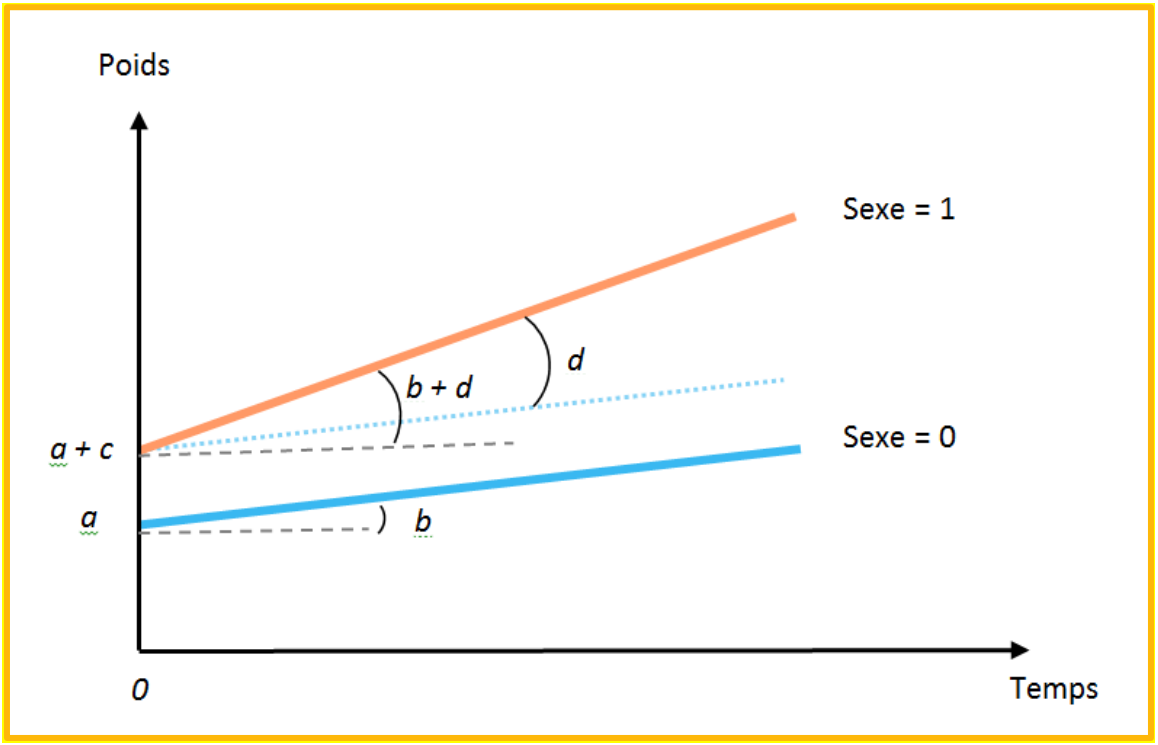
Il apparait alors que « c » est la différence constante de poids entre les mâles et les femelles à tous les temps t. En effet, quel que soit l'instant t considéré, le paramètre « c » possède toujours la même valeur.

« d » représente la différence de GMQ entre les mâles et les femelles. Si d>0, le GMQ des mâles est plus important que celui des femelles. Si d<0, les femelles grandissent plus vite que les mâles.

Ainsi, en l'absence de différence de GMQ entre les mâles et les femelles sur l'une ou l'autre des portions de la courbe (d=0), les deux droites seraient parallèles, séparées par une distance de valeur c. Il se lit très facilement sur le graphique pour t=0 et correspond alors à la différence de poids de naissance entre les mâles et les femelles.

Les paramètres du modèle « a », « b », « c » et « d » peuvent être représentés sur des courbes de croissance schématiques pour les mâles et les femelles. Sur le graphique présenté en Figure 9 ci-après, les mâles (sexe=1) grandissent plus vite que les femelles (sexe=0), et « d » a alors une valeur positive.

Figure 9 : Graphique des courbes de croissance schématiques pour un modèle linéaire sans rupture de pente pour chacun des deux sexes, présentant les coefficients du modèle.



Ce modèle est simple, mais peut être trop simple pour représenter la réalité, si par exemple le GMQ augmente ou diminue sur une période plus petite que la période totale de recueil de données. On peut en effet imaginer que le poids n'évoluerait pas linéairement durant les 90 jours considérés, nécessitant alors d'être modélisé par plusieurs équations linéaires, correspondant à des portions de droites de pentes différentes. Si tel est le cas, il faut alors écrire un modèle permettant au moins une rupture de pente.

b. Modèle à une rupture de pente

Construire un modèle à une rupture de pente revient à autoriser un changement de pente de la droite à un temps t_1.

Ce temps t_1 peut être fixé au préalable pour une raison biologique, ou exigée par l'étude dans le cas où le but de l'étude serait de tester si le GMQ est différent entre avant et après cette valeur de t_1 précisément (t_1 pouvant correspondre à un changement alimentaire, un changement d'environnement (Sirot, 2012), une valeur trouvée dans la littérature …).

Le temps de rupture t_1 peut également être recherché par le logiciel pour fournir la meilleure adéquation possible aux données, à l'aide de boucles itératives.

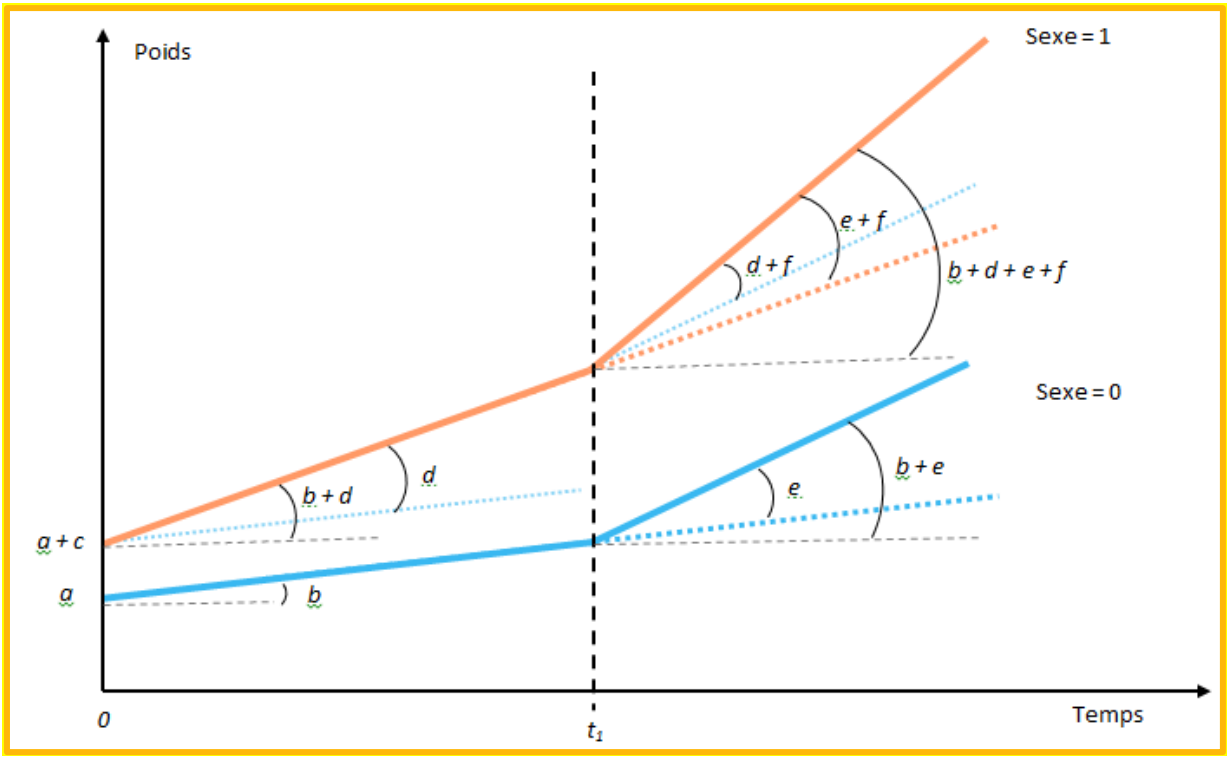
Par rapport au modèle sans rupture de pente présenté dans la partie précédente, deux termes supplémentaires sont ajoutés dans l'équation, précédés des paramètres « e » et « f » du modèle :

$$\textbf{Poids (t, sexe)} = a + b.t + c.sexe + d.t.sexe + e.(t-t_1)_+ + f.(t-t_1)_+ .sexe + \varepsilon$$

Avec $(t-t_1)_+ = 0$ si $t < t_1$ et $(t-t_1)_+ = t-t_1$ si $t > t_1$

Les courbes pour les mâles et les femelles associées à cette équation sont présentées en Figure 10.

Figure 10 : Graphique des courbes de croissance schématiques pour un modèle linéaire à une rupture de pente pour chacun des deux sexes, présentant les coefficients du modèle.



Avant t_1, le raisonnement est similaire à celui effectué pour un modèle sans rupture de pente.

Pour les femelles avant t_1: Poids (t, sexe=0) = a + b.t **(f_1)**

« a » représente donc le poids à la naissance (quand t=0) pour les femelles.

« b » représente le gain de poids quotidien moyen (GMQ) avant t_1 chez les femelles.

Pour les mâles avant t_1: Poids (t, sexe=1) = a + c + (b+d).t **(m_1)**

« a+c » représente donc le poids à la naissance (quand t=0) pour les mâles.

« b+d » représente le gain de poids quotidien moyen (GMQ) avant t_1 chez les mâles.

En effectuant la soustraction (m_1) – (f_1), on obtient :

Poids (t, sexe=1) - Poids (t, sexe=0) = c + d.t **(m_1) – (f_1)**

« c » est la différence de poids entre mâles et femelles à la naissance. Pour être exact, « c » est une différence de poids constante entre les mâles et les femelles avant t_1.

« d » est la différence de GMQ entre mâles et femelles avant t_1.

Après t_1, il faut réécrire les équations pour les deux sexes :

Pour les femelles après t_1: Poids (t, sexe=0) = a + b.t + e.(t-t_1) **(f_2)**

« b + e » représente le GMQ après t_1 chez les femelles.

Donc « e » est la différence de GMQ des femelles entre les deux périodes : avant et après t_1.

Pour les mâles après t_1: Poids (t, sexe=1) = a + c + (b+d).t + (e+f).(t-t_1) **(m_2)**

« b + d + e + f » représente le GMQ après t_1 chez les mâles.

Donc « e + f » est la différence de GMQ des mâles entre les deux périodes : avant et après t_1.

En effectuant la soustraction (m_2) – (f_2), on obtient :

Poids (t, sexe=1) - Poids (t, sexe=0) = c + d.t + f.(t-t_1) **(m_2) – (f_2)**

On constate que « d + f » est la différence de GMQ entre mâles et femelles après t_1.

Remarque : « c – f.t_1 » est une différence constante de poids entre mâles et femelles après t_1.

La démarche présentée ci-avant peut être appliquée pour créer des modèles à plus d'une rupture de pente. Toutefois, si l'allure des données nécessite de concevoir un modèle à plus d'une rupture de pente, il est possible d'envisager l'utilisation d'un modèle plus complexe, également susceptible de fournir une bonne adéquation aux données.

B. Modèles plus complexes : les modèles non-linéaires

Cette partie présente de façon théorique quelques modèles complexes disponibles dans la littérature. La liste des types de modèles est non exhaustive, de même que celle de leurs équations, qui peuvent prendre des formes variées (Campana et Jones, 1992).

1. Modèle autorisant des points d'inflexion : le modèle polynomial

Les équations polynomiales peuvent efficacement résumer des données de croissance pondérale.

Elles sont de la forme :

$$\textbf{Poids (t)} = a + b_1.t + b_2.t^2 + b_3.t^3 + \ldots + b_n.t^n$$

Avec « a » et « b_1, b_2, …., b_n » les paramètres du modèle à estimer. Le nombre n choisi doit être supérieur de 1 au nombre de points d'inflexion de la courbe (qui ne dépasse pas 4 dans la plupart des courbes de croissance).

2. Modèles de croissance curviligne à un point d'inflexion ou moins

Le principal avantage de ce type de modèles est leur souplesse : ils sont capables de s'adapter à de nombreuses formes de courbes, dès lors qu'elles présentent une partie incurvée et un point d'inflexion potentiel.

a. Modèle exponentiel

Les courbes exponentielles sont les analogues curvilignes des courbes tracées en modèles linéaires simples. Leur équation est la suivante :

Poids (t) = A.exp(G.t)

A est donc le poids estimé de l'animal à la naissance et G représente la vitesse de croissance instantanée.

b. Modèle de Gompertz

Le modèle de Gompertz, aussi appelé modèle de Laird-Gompertz, convient bien pour décrire les croissances sigmoïdes. Les équations de ce modèle possèdent la particularité de contenir deux fonctions exponentielles composées. En voici un exemple :

Poids (t) = P_{max} . exp[-k.exp (-G.t)]

P_{max} est le poids asymptotique, G la vitesse de croissance instantanée, et k un paramètre adimensionel.

c. Modèle logistique

Le modèle logistique donne un ajustement de la courbe très semblable à celui du modèle de Gompertz. Ils sont cependant différents car le modèle logistique impose une symétrie entre la partie en dessous du point d'inflexion et celle au-dessus, contrairement au modèle de Gompertz. Un exemple d'équation pour le modèle logistique est le suivant :

Poids (t) = P_{max} .[l + c.exp (-G.t)] $^{-1}$

P_{max} est le poids asymptotique, G la vitesse de croissance instantanée, c un paramètre à estimer.

Les modèles non linéaires sont très utilisés pour modéliser la croissance des poissons (Campana et Jones, 1992), notamment selon les stades biologiques (larves ou juvéniles) que l'on souhaite étudier dans ces cas particuliers. Ils ont été également utilisés à de nombreuses reprises pour modéliser la croissance des mammifères (Kaufmann, 1981; Zullinger *et al.*, 1984; Begall, 1996; Strathe *et al.*, 2010), malgré le fait que certains de leurs paramètres soient très difficilement interprétables.

↳ Nous venons de voir qu'il existe deux grandes familles de modèles : les modèles linéaires et les modèles non linéaires, d'équations plus complexes. Il est possible d'utiliser des équations de type linéaire mais tolérant un changement de pente en introduisant des seuils temporels auxquels l'inclinaison des portions de droites est susceptible de changer.

C. Adéquation du modèle aux données

1. Une méthode largement utilisée : le critère d'Akaike

La modélisation de la croissance passe souvent par la comparaison de plusieurs modèles d'équations différentes sur un même jeu de données, comme par exemple dans deux études visant à modéliser la croissance chez des animaux de rente tels que des agneaux (Lambe *et al.* 2006) ou des porcelets (Strathe *et al.* 2010). Dans chacune de ces deux études, 4 modèles d'équations différentes étaient testés et comparés sur les mêmes données de poids. Cette comparaison permet le choix du modèle qui correspond le mieux aux données (modèle qui possède la meilleure « goodness of fit »). Pour comparer ces modèles, on utilise en général le critère d'Akaike (Burnham, Anderson 1998).

Le critère d'information d'Akaike (en anglais « Akaike information criterion » ou « AIC ») est une mesure de la qualité d'un modèle statistique proposée par Hirotugu Akaike en 1973. Lorsque l'on estime un modèle statistique, il est possible d'augmenter la vraisemblance du modèle en ajoutant un ou des paramètres. Le critère d'information d'Akaike, tout comme le critère d'information bayésien, permet de pénaliser les modèles en fonction du nombre de paramètres afin de satisfaire le critère de parcimonie. On choisit alors le modèle avec le critère d'information d'Akaike le plus faible, et gardant ainsi seulement les paramètres d'intérêt principal. Il s'écrit comme suit : $AIC = 2k - 2*\ln(L)$, avec k le nombre de paramètres du modèle et L la valeur maximum de la fonction de vraisemblance du modèle (Akaike 1974).

- Il faut donc retenir que plus le critère d'Akaike est petit, plus l'adéquation du modèle aux données est bonne.

Il est également possible de comparer les résidus des différents modèles et de choisir celui pour lequel les valeurs présentes dans la matrice des résidus sont les moins variables.

2. Cas concret de l'étude sur les panthères nébuleuses : combien de ruptures de pentes choisir ?

Afin de savoir quel modèle linéaire utiliser dans cette étude (c'est-à-dire combien de ruptures de pente utiliser pour avoir la meilleure adéquation du modèle aux données) deux possibilités s'offraient à nous. Il était possible d'effectuer une recherche du meilleur modèle en autorisant des points de rupture à tous les temps à l'aide d'une boucle dans le programme ou de tester les modèles auxquels des ruptures avaient été attribuées aux moments correspondant aux changements alimentaires. C'est cette deuxième solution qui a été choisie dans cet exemple (avec des changements alimentaires ayant eu lieu à 28 et 42 jours). La première méthode sera détaillée dans la seconde partie de cette thèse sur la croissance pondérale des jeunes microcèbes.

Trois modèles d'équations différentes ont donc été testés dans le cas présent :

Modèle sans rupture de pente : $\textbf{Poids}(t) = a + b.t$

Modèle à une rupture de pente : $\textbf{Poids}(t) = a' + b'.t + c'.(t-28)_+$

Modèle à deux ruptures de pente : $\mathbf{Poids\ (t) = a'' + b''.t + c''.(t-28)_+ + d''.(t-42)_+}$

Avec : $(t-28)_+ = 0$ si $t < 28$ jours et $(t-28)_+ = t-28$ si $t > 28$ jours

$(t-42)_+ = 0$ si $t < 42$ jours et $(t-42)_+ = t-42$ si $t > 42$ jours

L'analyse statistique du modèle à deux ruptures de pentes a permis de montrer que le coefficient d'' n'était pas significativement différent de zéro. Ces résultats orientent le choix vers un modèle autorisant une seule rupture de pente, à 28 jours.

De plus, l'indice d'Akaike était meilleur pour les deux modèles autorisant des ruptures de pentes (indices d'Akaike égaux à 24269 et 24315) que pour le modèle sans rupture de pente (indice d'Akaike égal à 24372), ce qui nous conforte dans notre choix de modèle à une rupture de pente.

Après que le choix du modèle a été effectué, il s'agit d'utiliser la modélisation pour répondre aux questions que l'on se pose, qui sont souvent de la forme : « Est ce que telle variable explicative a un effet sur la croissance ? ». La relation de causalité (traduite dans la phrase précédente par le mot « effet ») est souvent difficile à prouver, et il faut dans tous les cas se montrer très prudent lorsque l'on interprète causalement les résultats des analyses (Desquilbet, 2013). Ces analyses, présentées dans la partie suivante doivent permettre de savoir s'il existe une association (brute ou ajustée) entre la variable explicative et la variable à expliquer, cette association pouvant être causale ou non (voilà pourquoi il est délicat d'employer le mot « effet » *a priori* !). La partie IV ci-après présente donc la méthode pour introduire des variables dans le modèle et interpréter les résultats de l'analyse statistique fournis par le logiciel.

IV. Introduction de variables explicatives à effets fixes dans le modèle et analyse des résultats

Les deux premières parties ci-après présentent la marche à suivre lors d'analyses d'associations brutes entre la croissance pondérale et deux variables explicatives différentes. Le principe de l'analyse ajustée est décrit dans la troisième partie.

A. Analyse de l'association brute entre le sexe de la panthère et sa croissance

1. Description de l'échantillon et du modèle

La composition de l'échantillon est en général présentée dans un tableau où figurent les pourcentages ainsi que les effectifs entre parenthèses. Si la variable explicative est binaire, seulement une des deux classes est décrite, les effectifs de l'autre s'obtenant par soustraction comme par exemple dans le Tableau 1 ci-après.

Tableau 1 : Composition de l'échantillon de panthères nébuleuses pour l'analyse de l'association brute entre le sexe et la croissance.

Catégorie	Pourcentage (n) (N=49)
Panthères non malades	**63 % (31)**
Panthères femelles parmi les non malades	32 % (10)

Dans cet exemple, l'analyse brute de l'association entre le sexe de la panthère et sa croissance a été réalisée seulement sur les données des animaux n'ayant pas été malades au cours des 90 premiers jours de vie. Il y avait 31 petits n'ayant pas présenté de problèmes gastro-intestinaux durant tout le sevrage, dont 10 femelles et 21 mâles.

L'équation du modèle à une rupture de pente utilisée était la suivante :

$$\textbf{Poids (age, sexe) = a + b.age + c.sexe + d.age.sexe + e.age_etoile + f.age_etoile.sexe}$$

Avec age_etoile = max (0, age – age_rupture) = max (0, age – 28). Avant le 28ème jour de vie, age_etoile prend donc systématiquement la valeur 0. C'est donc l'équivalent de $(t-t_1)_+$ que nous avions utilisé dans la partie A.2.b.

Rappelons que « sexe » est ici défini comme une variable binaire de valeur 0 si l'animal est une femelle, 1 si l'animal est un mâle.

Les résultats à l'issue de l'analyse statistique effectuée à l'aide de la procédure « proc mixed » du logiciel SAS (Annexe 1) sont présentés dans la partie suivante.

2. Lecture des résultats

En général, on fixe le risque d'erreur de première espèce α égal à 5 %. C'est-à-dire que si la valeur de p est inférieure à 0,05 la valeur du paramètre considéré (ici « a », « b », « c », « d », « e » ou « f ») est significativement différente de 0.

Il est possible de présenter les résultats sous forme d'un tableau (Tableau 2) reprenant les valeurs estimées des paramètres, les erreurs standards (SE) et les valeurs de p associées. Une capture d'écran de la fenêtre de résultats fournie par le logiciel SAS est disponible en Annexe 2.

Tableau 2 : Résultats de l'analyse statistique de l'association brute entre le sexe et la croissance chez les panthères nébuleuses.

Paramètre	Valeur estimée	SE*	p
a	218,77	33,80	<0,01
b	28,02	2,23	<0,01
c	16,93	41,06	0,68
d	- 2,16	2,71	0,42
e	2,70	2,50	0,29
f	9,71	3,05	<0,01

* SE : Standard Error

Le poids moyen (± SE) des petits à la naissance était estimé à 218,8 ± 33,8 grammes (paramètre « a »). D'après le tableau de résultats, la valeur de « c » n'était pas significativement différente de 0, ce qui permet de conclure que le sexe et le poids de naissance n'étaient pas associés. Ainsi, les petits mâles ne naissaient pas plus lourds que les petits femelles. Toutefois, la modélisation n'est bien évidemment pas le meilleur moyen d'effectuer une comparaison ponctuelle (à un temps précis, comme par exemple à la naissance). Il faut toujours préférer les analyses ponctuelles permettant de comparer les moyennes des données réelles.

La modélisation reste quant à elle le meilleur outil pour comparer les dynamiques de croissance, c'est-à-dire, dans le cas présent, les pentes représentant les Gains Moyens Quotidiens (GMQ). La comparaison des pentes se fait ici par l'analyse des coefficients « b », « d », « e » et « f » et/ou de leurs sommes le cas échéant.

Le coefficient « d » n'était pas significativement différent de zéro. La pente de la droite avant 28 jours n'était donc pas significativement différente entre les mâles et les femelles.

Le coefficient « e » n'était pas significativement différent de zéro. Cela signifie qu'il n'y avait pas de différence significative de GMQ entre avant et après 28 jours pour les femelles.

Pour savoir si le GMQ des mâles avant 28 jours était différent de celui après 28 jours, il faut déterminer si la somme des paramètres « e+f » (cf. Figure 10) était significativement différente de zéro.

La significativité d'une somme de deux paramètres demande de calculer une covariance entre ces deux paramètres (ce que peut fournir le logiciel), mais les calculs sont ensuite un peu laborieux, et nous les présenterons seulement en annexe (Annexe 3). Dans notre cas, la méthode présentée en Annexe 3 a permis de déterminer que la somme « e+f » était positive et significativement différente de zéro. Il est donc possible de conclure qu'il y avait une différence significative de GMQ chez les mâles entre avant et après 28 jours de vie : le GMQ des mâles était plus important après 28 jours de vie.

Pour savoir s'il existait une association brute entre le sexe et la croissance pondérale des jeunes panthères nébuleuses après 28 jours de vie, il faut déterminer si la somme des paramètres « d+f » (cf. Figure 10) était significativement différente de zéro. En employant la méthode décrite en Annexe 3, il apparait que la somme « d+f » était positive et significativement différente de zéro (d+f [± IC 95 %] = 7,6 [1,9 ; 13,2]). Le GMQ des mâles était donc significativement plus élevé (+ 7,6 grammes en moyenne) que celui des femelles après 28 jours de vie.

A partir des résultats du modèle on peut également écrire que les mâles avaient un GMQ plus élevé que les femelles dans notre échantillon de données, avec une différence de GMQ égale à 7,5 grammes (d+f = -2,2 + 9,7).

- Il est toujours possible de se prononcer sur les résultats dans l'échantillon, en donnant les valeurs chiffrées des paramètres et/ou de leurs sommes et en les comparant entre les classes (ici femelles *versus* mâles par exemple). Lorsque les différences sont significatives, il est alors autorisé de faire de l'inférence statistique, c'est-à-dire de penser que les différences observées à l'échelle de l'échantillon existent aussi à une échelle plus générale, dans la population cible[3] (dans notre cas la population cible peut être définie comme « les panthères nébuleuses maintenues en captivité», par exemple).

B. Analyse de l'association brute entre la durée des problèmes digestifs et la croissance des panthères

1. Description de l'échantillon

Afin de simplifier la démonstration et l'analyse des résultats, une variable binaire nommée « LS » a été créée et définie telle que LS = 0 si length_sick ≤ 7 jours et LS = 1 si length_sick > 8 jours. On crée donc deux classes en fonction du nombre de jours où des problèmes gastro-intestinaux ont été rapportés : si aucun problème n'était rapporté ou qu'il était rapporté pour une durée inférieure à une semaine, les petits appartenaient à la classe « LS = 0 ». Dans le cas contraire ils faisaient partie de la classe « LS = 1 ».

- De manière générale, l'interprétation est souvent facilitée lorsque l'on crée des variables en classes, et plus particulièrement des variables binaires. C'est donc une technique qu'il peut être judicieux d'adopter aussi souvent que possible.

Le Tableau 3 ci-après présente les effectifs utilisés lors de cette analyse.

[3] Sous réserve de représentativité de l'échantillon, mais nous n'en discuterons pas ici.

Tableau 3 : Composition de l'échantillon de panthères nébuleuses pour l'analyse de l'association brute entre la durée de la maladie et la croissance.

Catégorie	Pourcentages (n) (N=49)
Panthères non malades	**63 % (31)**
Panthères malades durant moins d'une semaine	**14 % (7)**
Panthères malades durant au moins une semaine	**23 % (11)**

L'équation du modèle à une rupture de pente utilisée était la suivante :

$$\text{Poids (age, LS)} = a + b.age + c.LS + d.age.LS + e.age_etoile + f.age_etoile.LS$$

Avec age_etoile = max (0, age – age_rupture) = max (0, age – 28).

« LS » est défini comme une variable binaire de valeur 0 si l'animal a été malade moins de 8 jours consécutifs ou n'a jamais été malade. Elle prend la valeur 1 si la jeune panthère a été malade plus de 7 jours consécutifs.

Les résultats à l'issue de l'analyse statistique effectuée à l'aide de la procédure « proc mixed » du logiciel SAS (Annexe 4) sont présentés dans la partie suivante.

2. Lecture des résultats

Les résultats obtenus sont présentés dans le Tableau 4.

Tableau 4 : Résultats de l'analyse statistique de l'association brute entre la durée de la maladie et la croissance chez les panthères nébuleuses.

Paramètre	Valeur estimée	SE*	p
a	216,06	8,85	<0,01
b	28,44	1,32	<0,01
c	1,91	18,64	0,9183
d	- 5,39	2,78	0,0529
e	8,21	1,37	<0,01
f	- 6,27	2,90	0,0307

***** SE : Standard Error

Le coefficient « d » n'était pas significativement différent de zéro. La pente de la droite avant 28 jours était donc la même pour les petits quelle que soit la durée de leur maladie.

Le coefficient « e » était significativement différent de zéro. Cela signifie qu'il y avait une différence significative de GMQ entre avant et après 28 jours pour les petits non malades ou malades durant moins d'une semaine (« LS=0 »).

La valeur de « f » était également significativement différente de zéro. Cependant, comme nous l'avons vu précédemment, la comparaison des GMQ entre avant et après 28 jours cette fois-ci pour les petits malades durant plus d'une semaine (« LS=1 ») nécessite de calculer la covariance entre les paramètres « e » et « f » afin de déterminer si leur somme était significativement différente de zéro. Les résultats obtenus suivant la méthode décrite en Annexe 3 permettent de dire que la somme « e+f » n'était pas significativement différente de zéro. Pour les petits ayant été malades pendant plus d'une semaine, il n'y avait donc pas de différence significative de GMQ entre avant et après 28 jours de vie.

De même, la comparaison des GMQ après 28 jours de vie entre les petits ayant été malades plus d'une semaine et les autres requiert un calcul de covariance entre les deux paramètres « d » et « f » afin de savoir si leur somme était significativement différente de zéro. Dans le cas présent, la méthode décrite en Annexe 3 a permis de conclure que la somme « d+f » était négative et significativement inférieure à zéro (d+f [± IC 95 %] = -11,6 [-19 ; -14,3]). Par conséquent, on peut dire que les jeunes panthères nébuleuses ayant été malades durant plus d'une semaine grandissaient moins vite que celles ayant été malades moins d'une semaine ou n'ayant pas été malades.

Dans notre échantillon, les panthères ayant été malades pendant plus d'une semaine prenaient en moyenne 11,7 grammes de moins (d+f = -5,4 + (-6,3) = -11,7) que ceux ayant été malades moins longtemps ou n'ayant pas été malades.

C. Notion de modèle ajusté, prise en compte de facteurs de confusion

Après avoir réalisé les analyses brutes présentées dans les paragraphes précédents, il est nécessaire de travailler avec un modèle ajusté sur les variables d'intérêt principal afin d'étudier les associations des différentes variables avec la croissance pondérale, indépendamment des associations éventuelles avec les autres variables explicatives (pouvant alors jouer le rôle de facteurs de confusion). Afin de faciliter l'interprétation des résultats, ce modèle ajusté fait intervenir seulement les deux variables binaires définies précédemment : « sexe » et « LS ».

Remarque : un tel modèle ne doit pas présenter de « cases vides ». C'est-à-dire que toutes les valeurs ou classes de toutes les variables introduites dans le modèle doivent se recouper sans qu'il y ait de situations de croisement sans individus ou avec trop peu d'individus (par exemple, il ne serait pas possible d'introduire dans un même modèle les deux variables « sexe » et « nombre de gestations de l'animal » car il n'y a aucun individu dans la case mâle = 1 et gestation ≠ 0 !). Il s'agit donc en pratique de choisir comment faire figurer les variables dans le modèle et quelles variables conserver par rapport à l'analyse brute. Dans le cas présent, afin que le problème de « case vide » ne se pose pas, l'analyse a été menée sur le fichier de données comprenant les 49 microcèbes après vérification qu'il y avait bien des mâles pour lesquels LS=0 et LS=1, et des femelles remplissant la même condition.

L'équation du modèle ajusté était la suivante :

$$\text{Poids (age,sexe,LS)} = a + b.age + c_1.sexe + c_2.LS + d_1.age.sexe + d_2.age.LS + e.age_etoile + f_1.age_etoile.sexe + f_2.age_etoile.LS$$

Le Tableau 5 présente les résultats de cette analyse ajustée. Une « proc mixed » a été utilisée dans le logiciel SAS (Annexe 5).

Tableau 5 : Résultats de l'analyse statistique ajustée sur le sexe et la durée de la maladie.

Paramètre	Valeur estimée	SE*	p
a	192,12	13,35	<0,01
b	29,32	2,02	<0,01
c_1	37,92	15,92	0,1701
c_2	8,60	18,74	0,6465
d_1	- 1,39	2,41	0,05629
d_2	- 5,63	2,84	0,0474
e	4,12	2,02	0,0471
f_1	6,49	2,41	<0,01
f_2	- 5,14	2,84	0,0712

***** SE : Standard Error

L'analyse ajustée nous apprend que « d_1 » n'était pas significativement différent de 0. On peut donc conclure qu'indépendamment de la durée de la maladie de l'animal, les mâles ne grandissaient pas plus vite ou moins vite que les femelles durant les 28 premiers jours de vie.

En revanche, « d_2 » était négatif et significativement différent de zéro. On peut alors dire qu'indépendamment de leur sexe, les petits ayant été malades pendant plus d'une semaine grandissaient moins rapidement que les autres avant 28 jours.

Il est possible de dire que les panthères femelles incluses dans l'étude avaient un GMQ moyen de 29,3 grammes avant 28 jours et 33,4 grammes (b+e = 29,3 + 4,1 = 33,4) après 28 jours. Les mâles avaient quant à eux un GMQ moyen de 27,9 grammes ($b+d_1$ = 29,3-1,4 = 27,9) avant 28 jours et 38,5 ($b+d_1+e+f_1$ = 29,3-1,4+4,1+6,5 = 38,5) grammes après 28 jours.

Afin de savoir s'il est possible d'inférer les différences observées entre ces résultats (à l'échelle de l'échantillon de panthères nébuleuses de notre étude) à la population cible, les analyses de significativité de sommes sont nécessaires (telles que décrites en Annexe 3).

« f_1 » était positif et significativement différent de 0. La somme des paramètres « e » et « f_1 » était positive et significativement différente de zéro : indépendamment de la durée de la maladie de l'animal, les mâles grandissaient plus vite après 28 jours de vie qu'avant 28 jours de vie.

La somme « d_1+f_1 » était non significativement différente de zéro : indépendamment de la durée de la maladie, les mâles ne grandissaient pas plus vite que les femelles après 28 jours de vie.

Dans l'échantillon, les panthères non malades ou malades durant moins d'une semaine avaient un GMQ moyen de 29,3 grammes avant 28 jours et 33,4 grammes (b+e = 29,3 + 4,1 = 33,4) après 28 jours. Les jeunes ayant été malades plus d'une semaine avaient quant à eux un GMQ moyen de 23,7 grammes (b+d_2 = 29,3-5,6 = 23,7) avant 28 jours et 22,7 grammes (b+d_2+e+f_2 = 29,3-5,6+4,1-5,1 = 22,7) après 28 jours.

La somme des paramètres « e + f_2 » était non significativement différente de zéro : indépendamment de leur sexe, les petits ont des GMQ non différents avant et après 28 jours quelle que soit la durée de leur maladie.

La somme « d_2+f_2 » était négative et significativement différente de zéro : indépendamment du sexe de l'animal, les jeunes panthères malades pendant plus d'une semaine grandissaient moins vite que les autres après 28 jours de vie. A la lumière de ce qui a été écrit ci-avant, on peut dire que les jeunes panthères malades durant plus d'une semaine grandissaient moins rapidement que les autres sur toute la durée de l'étude (de 0 à 90 jours).

On constate que les conclusions énoncées ci-dessus pour les jeunes panthères avant 28 jours rejoignent celles de l'analyse brute étudiant l'association entre la croissance et le sexe de l'animal. En revanche, elles sont différentes de celles de l'analyse brute faisant intervenir la durée de la maladie. Ceci permet de penser que le sexe jouait le rôle de facteur de confusion dans la relation entre la variable « LS » et la croissance de la jeune panthère étudiée lors de l'analyse brute, durant la première période de vie. En ce qui concerne les conclusions après 28 jours de vie, ce sont celles énoncées pour l'effet du sexe qui diffèrent entre l'analyse brute et l'analyse ajustée : cette dernière a montré que les mâles ne grandissaient pas plus rapidement que les femelles après 28 jours de vie, une fois la durée de la maladie prise en compte, alors que l'analyse brute uniquement chez les petits sains avait permis de conclure que les mâles grandissaient plus vite que les femelles. Dans le cas présent, on peut penser que la différence de résultats entre l'analyse brute chez les petits sains et l'analyse ajustée sur la durée de la maladie est due à un phénomène d'interaction entre le sexe et la durée de la maladie. Il est en effet possible que les mâles soient moins résistants que les femelles à la maladie et que leur courbe de croissance en pâtisse davantage. Ainsi, les GMQ moyens seraient très différents après 28 jours entre les mâles sains et les mâles malades alors les GMQ moyens après 28 jours entre les femelles saines et les femelles malades ne seraient que peu éloignés. En prenant en compte les animaux malades et sains, la différence entre mâles et femelles serait donc beaucoup moins prononcée qu'en ne prenant en compte que les sains.

- Les résultats de l'analyse ajustée peuvent être différents de ceux énoncés après les analyses brutes (c'est le cas ici pour la variable « LS » avant 28 jours de vie, et pour la variable « sexe » après 28 jours de vie). La prise en compte des autres variables explicatives peut donc mettre en évidence d'importants biais de confusion ! Dans une étude à visée analytique (recherche de la relation causale), destinée à être publiée, l'analyse ajustée doit être présentée en plus de l'analyse brute, puisque c'est celle qui permet d'ouvrir vers la discussion à visée causale. Toutefois, il est intéressant d'effectuer les analyses brutes car elles permettent de savoir, pas comparaison avec l'analyse ajustée, si les autres variables explicatives jouent ou non un rôle de facteur de confusion.

V. Contraintes imposées par la non-indépendance des données

A. Notion de non-indépendance

Plusieurs situations conditionnent la non-indépendance des données collectées. Les données sont non-indépendantes si :

- On effectue plusieurs mesures par sujet, plus ou moins espacées dans le temps. Si elles sont espacées dans le temps dans le but de mettre en évidence une évolution temporelle (exemple : mesures de poids quotidiennes dans les 90 premiers jours de vie) on parle de données longitudinales.
- On effectue des mesures sur plusieurs animaux d'un même groupe, « cluster » en anglais (exemple : tous les petits d'une même portée).

Lorsque l'on choisit de modéliser la croissance pondérale au cours du temps, un même animal est pesé plusieurs fois. La non-indépendance des données est donc systématique, puisqu'il y a plusieurs mesures consécutives pour un même animal. Ceci constitue un premier niveau de non indépendance, dont on ne peut pas s'affranchir.

Il peut également y avoir plusieurs petits par portée, donc le poids d'un petit peut potentiellement être corrélé au poids de son/ses frère(s) et sœur(s), ceci constituant alors un deuxième niveau de non-indépendance. Enfin, une même femelle est susceptible d'avoir plusieurs portées au cours de sa vie, ce qui crée un troisième niveau de non-indépendance : celui des portées entre-elles. Ces deux autres niveaux de non indépendance ne sont pas rencontrés systématiquement : ils n'ont, par exemple, pas lieu d'être si l'on travaille sur une espèce qui ne met au monde que des petits uniques et que l'on choisit de n'étudier qu'une seule mise bas par mère.

De même, il n'est pas nécessaire de prendre en compte les deux derniers niveaux de non indépendance évoqués ci-dessus dans le cas où les petits sont élevés à la main. En effet, dans ce cas le poids d'un animal à un âge donné sera *a priori* indépendant de celui de ses frères et sœurs de portée et également indépendant du poids des animaux des autres portées de la même mère. L'intervention humaine, par le fait de donner une quantité de nourriture déterminée selon un protocole préalable, annihile, en grande partie *a priori*, ces deux niveaux de non-indépendance des données.

Il y a donc 3 niveaux potentiels de non-indépendance des données. Le premier niveau est celui de l'animal, et doit être obligatoirement pris en compte dès lors que l'on possède plusieurs mesures sur le même animal. Les deux autres niveaux (portée et mère) doivent être pris en compte lorsque nécessaire, en fonction du contexte, des caractéristiques de l'espèce étudiée, et des données disponibles.

Dans les études, la prise en compte de cette non-indépendance des mesures est rendue possible en introduisant dans le modèle ce que l'on appelle des effets aléatoires portant sur les 3 niveaux (l'animal lui-même et éventuellement la mère et/ou la portée).

Dans le cas de notre exemple des panthères nébuleuses élevées à la main, il n'y a qu'un seul niveau de non indépendance : le niveau individuel. Il existe une ligne dans le programme permettant de spécifier que l'on souhaite le prendre en compte sous forme d'effets aléatoires.

B. Prise en compte de la non indépendance des données : introduction d'effets aléatoires dans le modèle

Les modèles comprenant simplement des effets aléatoires ne sont pas utilisés en pratique : ils ne possèderaient que peu d'intérêt puisque la modélisation est justement utilisée pour observer les associations entre une ou des variable(s) explicative(s) et la variable réponse. Les effets aléatoires figurent donc dans le modèle *en plus* des effets fixes. Les modèles prenant en compte à la fois des variables à effets fixes et des effets aléatoires sont dits « modèles mixtes ».

A des fins pédagogiques, l'équation utilisée en exemple dans cette partie est l'équation du modèle linéaire à une rupture de pente à 28 jours, pour une panthère i, au temps t.

$$\textbf{Poids}_{i,t} = a + b.t + c.(t-28)_+$$

Avec : $(t-28)_+ = 0$ si $t < 28$ jours et $(t-28)_+ = t-28$ si $t > 28$ jours

Un seul niveau d'effets aléatoires sera pris en compte : le niveau individuel. Ainsi, on obtient l'équation suivante :

$$\textbf{Poids}_{i,t} = (a + a_i) + (b + b_i).t + (c + c_i).(t-28)_+$$

« a_i », « b_i » et « c_i» sont donc les effets aléatoires liés à l'individu, pris en compte sur chacun des trois paramètres de ce modèle linéaire à une rupture de pente (« a », « b » et « c »).

« a_i », « b_i » et « c_i» suivent chacun une loi Normale de paramètres $(0, \sigma^2)$, c'est-à-dire une loi Normale de moyenne nulle et de variance σ^2 à estimer (l'estimation sera effectuée par le logiciel). C'est-à-dire que la moyenne des « a_i » pour tous les animaux sera égale à 0 (de même pour « b_i » et « c_i»).

Les « a_i » permettent de prendre en compte la variabilité individuelle à la naissance, c'est-à-dire le fait que toutes les panthères n'ont sans doute pas le même poids à t=0.

Les « b_i » et « c_i» permettent de prendre en compte la variabilité de GMQ entre les différents animaux sur les deux périodes considérées. En effet, le bagage génétique étant différent d'un animal à l'autre, il est possible que certaines jeunes panthères grandissent plus vite que d'autres.

Dans le cas présenté, les effets aléatoires vont donc permettre au modèle de prendre en compte, sur chacun des paramètres du modèle, la variabilité liée aux individus eux-mêmes.

Dans certains cas, il n'est pas nécessaire de faire figurer les effets aléatoires sur *tous* les paramètres du modèle comme cela a été présenté ici pour les jeunes panthères.

- L'introduction des effets aléatoires dans le modèle lui confère le statut de modèle à effets mixtes (« mixed-effect model » en anglais), puisqu'il contient à la fois des effets fixes et aléatoires. Les effets aléatoires se présentent théoriquement dans l'équation sous forme de coefficients que l'on additionne aux paramètres du modèle. Toutefois, ils sont pris en compte sous forme de lignes particulières dans les programmes, sans que l'opérateur ne soit obligé de les faire figurer dans l'équation, afin de ne pas alourdir cette dernière. L'introduction d'effets aléatoires sera détaillée en pratique pour les deux types de modèle dans la seconde partie de ce manuscrit.

C. Cas particulier des données répétées

Lorsque les mesures sont réalisées à une fréquence élevée (tous les jours par exemple), la mesure de poids à un âge donné dépend obligatoirement du poids de l'individu à la mesure précédente. Pour décrire ce phénomène, on parle de « données répétées » (« repeated measurements » en anglais).

Pour énoncer les choses de façon plus générale, le terme de « données répétées » fait référence à des réponses prises en séquence sur la même unité expérimentale : l'individu. Ces mesures sont prises au cours du temps dans le cas de modélisation de la croissance mais il est également envisageable de parler de données répétées (tout comme de non-indépendance des mesures au niveau individuel) dans le cas où la séquence n'est plus temporelle mais spatiale, comme par exemple des mesures portant sur différentes parties du corps d'un même animal (Littel, 1998).

Pour prendre en compte les données répétées, il est nécessaire d'écrire une ligne particulière dans le programme du logiciel lui signifiant que les données qu'il analyse sont des données répétées.

- La non-indépendance des données est à prendre en compte systématiquement dans une étude longitudinale. L'espèce étudiée ainsi que le contexte et le protocole de collecte des données conditionnent le nombre de niveaux de non-indépendance à prendre en compte. Plus ce nombre est important, plus la prise en compte dans le modèle est complexe, et informatiquement « lourde ». Par ailleurs, dès lors que plusieurs mesures sont effectuées sur le même animal, la prise en compte des données répétées est également nécessaire.

Une fois la non-indépendance des données prise en compte, les résultats s'interprètent comme vu précédemment dans la partie traitant des effets fixes. La prise en compte de la non-indépendance des données est essentielle : il est très rare qu'une fois que l'on a introduit les effets aléatoires une tendance observée lors de l'analyse avec les effets fixes seuls s'inverse. En revanche, la prise en compte de la non-indépendance des données peut jouer sur le degré de significativité, et ainsi changer complètement la conclusion issue des résultats. Pour des raisons pédagogiques, la notion de non-indépendance des données a été introduite après avoir expliqué la modélisation avec les effets fixes. Il est toutefois évident que dans une démarche d'analyse classique, l'introduction d'effets aléatoires dans le modèle doit être effectuée avant de faire tourner le programme pour les effets fixes afin que les résultats obtenus prennent en compte la non indépendance des données.

VI. Logiciels de modélisation et procédures

Plusieurs logiciels capables de prendre en compte les équations des modèles cités dans les parties précédentes sont disponibles.

Certaines procédures du logiciel SAS (de la société SAS Institute Incorporation), évoquées tout au long de cette première partie, permettent des analyses prenant en compte les effets aléatoires, notamment sur plusieurs niveaux.

Le tutoriel qui m'a permis de prendre ce logiciel en main est disponible en ligne (Desquilbet, 2012). L'aide en ligne proposée par SAS (http://support.sas.com) est également très bien conçue et fournit un guide pour s'approprier chacune des procédures du logiciel. Les procédures utilisées dans les analyses présentées dans cette thèse sont toutes disponibles en annexes. En s'appuyant sur le tutoriel cité plus haut, l'aide en ligne et le contenu de ces annexes, il est possible pour un débutant de réussir à « faire tourner » les analyses pour modéliser une croissance pondérale.

SAS fonctionne sur différents systèmes d'exploitation (Windows, Macintosh, Linux…). Son inconvénient majeur reste le coût de sa licence à l'année.

Un second logiciel disponible pour effectuer de telles analyses est le logiciel R, qui présente l'avantage d'être téléchargeable gratuitement, tout en fonctionnant également sur les différents systèmes d'exploitation.

- PARTIE B -

RETOUR D'EXPÉRIENCE SUR L'ANALYSE DE LA CROISSANCE PONDÉRALE DE JEUNES MICROCÈBES MURINS À L'AIDE DE DEUX MODÈLES

DONNÉES BIBLIOGRAPHIQUES SUR L'ESPÈCE *MICROCEBUS MURINUS*

Le microcèbe murin (*Microcebus murinus*) est un petit prosimien appartenant à la famille des Cheirogaleidae, une des cinq familles de primates lémuriformes (Figure 11). Son aire de répartition naturelle couvre les forêts sèches des parties sud et ouest de Madagascar. Ce petit lémurien nocturne et arboricole se nourrit principalement d'insectes et de fruits. En période défavorable, il peut également consommer des gommes végétales (Dammhahn, Kappeler 2008).

Figure 11 : Dessin d'un microcèbe murin (source : Handbook of the Mammals of the World, Volume 3 (Wilson, Mittermeier 2009))

En captivité, les microcèbes possèdent une espérance de vie moyenne de 5 à 6 ans (Perret 1997). Le microcèbe est considéré comme un modèle privilégié d'étude du vieillissement chez les primates, notamment au niveau cognitif. Son appartenance à l'ordre des Primates, qui lui confère une proximité phylogénétique avec l'Homme, sa durée de vie relativement courte, sa petite taille le rendant facilement manipulable et la possibilité d'accélérer son vieillissement en manipulant la durée des cycles photopériodiques en font un modèle de choix dans l'étude du vieillissement (Guignard 2009).

Le microcèbe est un mammifère monomorphique : à l'âge adulte, les mâles et les femelles se ressemblent morphologiquement (Wright 1999). La polyandrie est une caractéristique de cette espèce : les femelles peuvent s'accoupler avec plusieurs mâles durant l'œstrus (Eberle et Kappeler 2004; Huchard 2012). Ainsi, même en milieu captif il est souvent impossible de connaitre le père des petits. La mère met au monde des portées comprenant entre 1 et 4 petits, une à deux fois par an, après une période moyenne de gestation de deux mois.

Une étude sur l'investissement comportemental de la mère avec ses petits (Colas 1999) a montré que les contacts entre les mères et leurs petits étaient plus fréquents pour les mères multipares que pour les mère primipares. Les mères avaient davantage tendance à interagir avec leurs petits dans le cas de portées dans lesquelles au moins un mâle était présent. Cependant, au sein des portées mixtes, le nombre de contacts maternels avec les petits ne différait pas suivant le sexe de ces derniers. L'auteur émettait l'hypothèse que la présence de mâles dans une portée pouvait être interprétée comme un stimulateur de comportement maternel.

Les petits commencent à s'alimenter par eux-mêmes aux alentours de 40-45 jours de vie, dans la nature comme en captivité, et on considère qu'ils deviennent complètement autonomes au bout de deux mois de vie environ (Perret 1990). Ils atteindront leur maturité sexuelle un peu avant l'âge d'un an.

Les microcèbes, et les prosimiens en général, font partie des espèces de mammifères possédant une phase pré-sevrage courte avec taux de croissance pondérale important, néanmoins considéré comme modéré par rapport à d'autres groupes de mammifères tels que les lagomorphes, les artiodactyles ou les canidés. Les marsupiaux, qui ont une croissance pré-sevrage caractérisée par un gain de poids quotidien faible (Cockburn, Johnson 1988), et les grands primates sont connus pour avoir un taux de croissance faible (Case 1978). Il existe donc, au sein des mammifères des différences interspécifiques de dynamique de croissance, dont les aspects de modélisation sont évoqués ci-après.

PARTIE EXPÉRIMENTALE

I. Contexte et objectifs de l'étude

Certains microcèbes ont été ramenés de Madagascar en 1967 pour former une colonie de microcèbes captifs à Brunoy (91), au sein de l'UMR 7179 du Muséum. La Figure 12 présente une photographie d'un microcèbe adulte issu de cet élevage. La colonie de Brunoy est d'ailleurs aujourd'hui la plus importante colonie de microcèbes captifs au monde. Les données nécessaires à cette étude ont été collectées au cours des 40 dernières années au sein de cet élevage de microcèbes. Le travail présenté ici sur la dynamique de croissance du microcèbe s'inscrit dans un sujet de recherche plus vaste de l'UMR 7179. En effet, une première caractérisation du processus de croissance juvénile et ses déterminants dans cette espèce permettra par la suite de tester les coûts à long terme de cette croissance par le biais d'analyses de survie.

Figure 12 : Photographie d'un microcèbe adulte de l'élevage de Brunoy (91) (Crédit photographie : Eric Gueton, MNHN)

Le travail présenté dans cette partie expérimentale s'est déroulé en 3 phases distinctes après documentation sur l'espèce modèle.

- La première phase a été marquée par la mise en place d'une unique base de données de travail à partir des requêtes issues de la base de données de l'UMR 7179. Les données ont été triées puis ordonnées sous SAS pour permettre l'analyse dans de bonnes conditions.
- La seconde phase a consisté à rechercher le modèle de croissance adapté au mieux aux données de poids de cette espèce et dont les paramètres descriptifs seraient assez facilement interprétables. Il a également été nécessaire de réfléchir aux différents effets aléatoires à prendre en compte.
- La troisième phase a permis de tester les effets de plusieurs caractéristiques individuelles, maternelles et de la portée sur la croissance des petits microcèbes en étudiant leur association avec les paramètres descriptifs du modèle. L'analyse a été menée de façon brute puis ajustée.

II. Matériel et méthodes

A. Matériels animaux et recueil de données

Toutes les données sont issues de l'élevage de microcèbes captifs de l'UMR 7179 du Muséum National d'Histoire Naturelle basé à Brunoy. Les poids des petits sont recueillis au dixième de gramme près en moyenne 9 fois entre le premier et le 100ème jour de vie. Les jeunes microcèbes sont donc manipulés quotidiennement (Figure 13). Des feuilles sont remplies à la main par les personnes chargées des pesées puis les données sont ensuite saisies dans une base de données Access.

Figure 13 : Jeunes microcèbes âgés de moins d'une semaine manipulés avec précaution pour la pesée. (Crédit photographie : Eric Gueton, MNHN)

B. Base de données

Données directement disponibles

Les données nécessaires pour réaliser cette étude ont été recueillies à partir de la base de données disponible sous Access à Brunoy. Suite à plusieurs requêtes lancées dans ce logiciel, deux tableaux ont été mis à notre disposition. Le premier tableau Excel (tableau « microcèbes ») classait tous les microcèbes par identifiant et indiquait pour chacun le nom de la mère, le nombre de petits mâles ou femelles dans la portée, le poids dans les 100 premiers jours de vie accompagné de l'âge à la date de mesure et enfin une case « remarque ». Le second tableau présentait cette fois les identifiants des femelles ayant mis bas ou avorté au moins une fois (tableau « mères »). Pour chaque femelle, nous disposions de la date de mise-bas ou de l'avortement, du poids suite à la mise bas et de l'âge de la mère à la mise bas.

Données filtrées

Il a été nécessaire d'utiliser des filtres durant les requêtes effectuées sous Access puis par la suite de « filtrer » manuellement les données pour que les animaux inclus dans l'étude soient des microcèbes ayant grandi avec leur mère, dans des conditions standard. Les microcèbes élevés avec des rythmes de photopériodes accélérés, ceux dont les mères avaient subi des expérimentations comme une restriction calorique par exemple, ceux à qui il manquait un membre à la naissance (ceci pouvant influer sur leur poids), ainsi que ceux élevés à la main ont donc été exclus d'office des tableaux de données. Enfin, les animaux nés en période SD (pour « short day » en anglais, c'est-à-dire avec un rythme photopériodique raccourci par rapport à la normale), ont aussi été exclus de l'étude.

Par ailleurs, l'identification individuelle des petits ne se fait qu'après 60 jours de vie. C'est pourquoi, pour les portées comportant plusieurs animaux de même sexe, il est possible qu'il y ait eu des erreurs d'attribution de poids. En effet, s'il y a eu par exemple deux mâles dans une portée, lors de la première pesée le poids le plus élevé a été noté en face de l'identifiant « mâle n°1 », et le poids le plus faible a donc été attribué au « mâle n°2 ». A la pesée suivante, le poids le plus élevé sera reporté sur la ligne du « mâle n°1 » et le poids le plus faible sur la ligne du « mâle n°2 », suivant la même logique que pour la première pesée et indépendamment de l'identité réelle des animaux. Afin de s'affranchir de ce problème d'identités non certaines, nous avons choisi de ne conserver que les microcèbes pour lesquels il ne pouvait pas y avoir d'erreur d'identification. Ce sont donc les microcèbes seuls dans leurs portées, les microcèbes mâles n'ayant aucun frère et les microcèbes femelles n'ayant aucune sœur.

Le tableau de données « donnees_certaines » contient le résultat de cette phase de sélection.

Variables créées

La programmation sous SAS a permis de créer de nouvelles variables à partir des variables existantes :

- Le nombre de mises bas précédentes de la mère
- La variable binaire « multipare » et plusieurs variables en classes selon le nombre d'accouchements précédents
- Le nombre d'autres petits de la portée
- La variable binaire « petit_unique ».

Fusion des tables de données « microcèbes » et « mères »

Après de nombreux remaniements et à l'aide de la commande « MERGE » de SAS, les deux tables de données initiales ont été fusionnées grâce aux clefs de fusion « nom » (l'identifiant unique de l'animal), « mère » et « date de naissance » pour n'avoir qu'un seul tableau de données sur lequel travailler, et regroupant toutes les variables nécessaires à l'analyse. Ce nouveau tableau de données a été nommé « microcebe.final ».

Fusion des tables de données « final » et « donnees_certaines »

Toujours grâce à la commande « MERGE » de SAS, les deux tables de données « final » et « donnees_certaines » ont été fusionnées grâce aux clefs de fusion « nom », « mère » et « date de naissance » pour donner le tableau « fusion_certaines » à partir duquel nous serons amenés à effectuer les analyses.

Le programme correspondant est disponible en Annexe 6.

C. Méthodes statistiques

Cette partie a pour objectif de présenter, de façon peu conventionnelle mais dans un but pédagogique, la construction en parallèle de deux modèles différents en adéquation avec les données de poids des microcèbes de cette étude. Le but est de déterminer *in fine* lequel des deux modèles possède le plus d'avantages et semble le plus approprié pour poursuivre l'analyse. En pratique, dans de nombreux articles scientifiques, un simple tableau résume les modèles testés en comparant leurs indices d'Akaike et une phrase spécifie quel modèle a été choisi. Nous verrons que l'indice d'Akaike permet effectivement de comparer les deux modèles, mais que de nombreux autres éléments sont à considérer afin de choisir le modèle final.

1. Construction de deux modèles adaptés à la croissance des microcèbes

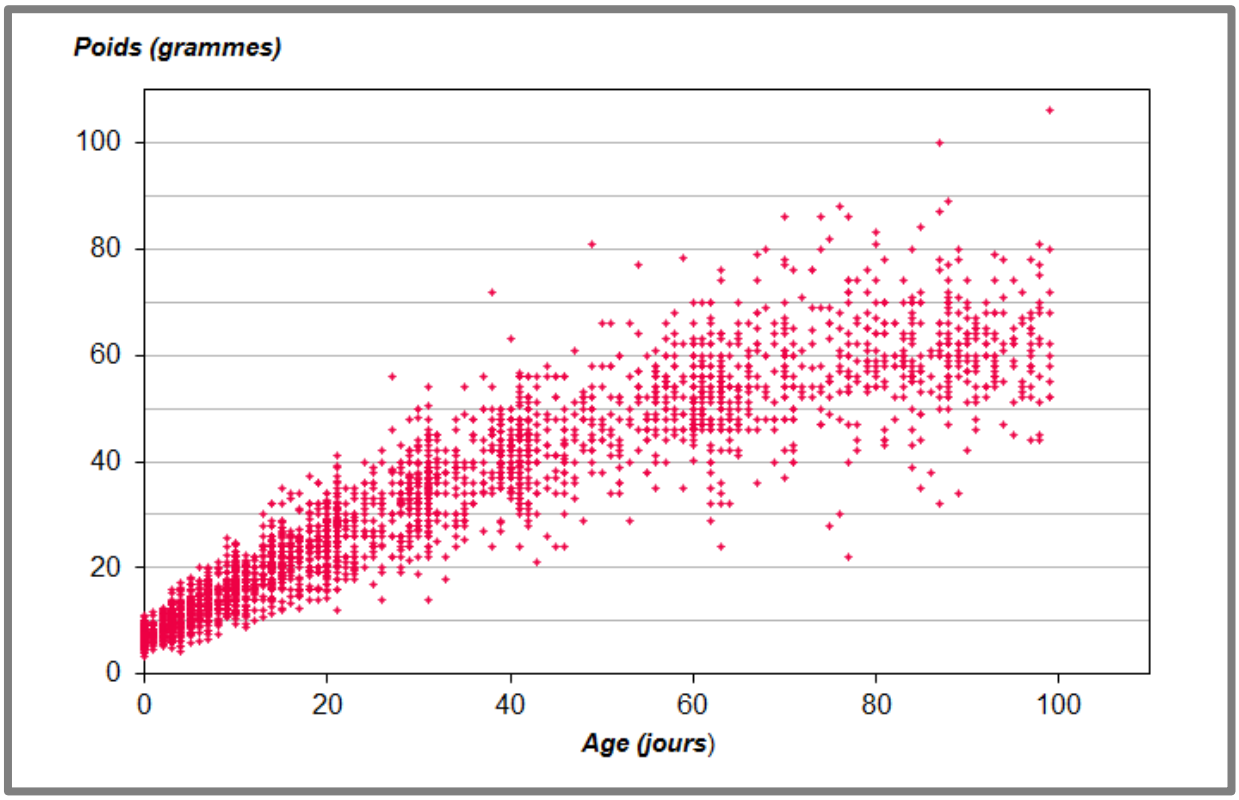
a. Visualisation des données brutes en nuage de points

La base de données contient 645 microcèbes avec en moyenne (± écart-type) 8,6 ± 1,9 mesures par animal au cours de ses 100 premiers jours de vie. Pour rendre le nuage de points lisible, une sélection aléatoire a été effectuée afin de ne garder que 150 individus. La « technique » consiste à faire générer pour chaque microcèbe un nombre aléatoire unique (typiquement un nombre compris entre 0 et 1 avec de nombreux chiffres après la virgule), puis à demander au programme de ne garder que les 150 plus petits nombres aléatoires dans un nouveau fichier de données à partir duquel sera effectué le tracé en nuages de points.

La Figure 14 ci-dessous présente le résultat de ce tracé en nuage de points pour 150 microcèbes au cours de leurs 100 premiers jours de vie.

Le programme utilisé pour sélectionner les données est disponible en Annexe 6.

Figure 14 : Graphique en nuage de points des poids de 150 microcèbes murins de l'étude au cours de leurs 100 premiers jours de vie.



La forme du nuage de points de la figure permet de s'orienter vers certains modèles pour la suite de l'étude. Les deux parties suivantes (b et c) présentent les deux modèles choisis, l'un linéaire et l'autre non, la justification du choix de ces modèles ainsi que la « construction » de leurs équations pour cette base de données en particulier.

b. Modèle linéaire

Nous avons vu dans la partie théorique de cette thèse qu'il est possible de complexifier un peu les modèles linéaires en leur autorisant des ruptures de pentes. Au vu du tracé en nuages de points, il nous a semblé indiqué de tester une rupture de pente pour ce modèle linéaire. Un programme particulier (appelé « macro ») a été écrit, permettant de chercher le meilleur point de rupture à l'aide d'une boucle : le logiciel fait tourner le modèle sur toutes les données, en faisant varier le point de rupture entre 0 et 100 jours et fournit, pour chaque âge de rupture pris en compte, les paramètres estimés obtenus ainsi que l'indice d'Akaike correspondant. Il suffit ensuite de regarder pour quel âge de rupture l'indice d'Akaike est le plus faible. Dans le cas présent, le plus faible indice a été obtenu pour une rupture de pente à l'âge de 46 jours. L'âge de rupture de pente a donc été fixé dans la suite de l'étude à 46 jours.

L'équation de ce modèle linéaire est la suivante :

$$\textbf{Poids} = \beta_1 + \beta_2 * \textbf{agepesee} + \beta_3 * \textbf{agepesee_etoile}$$

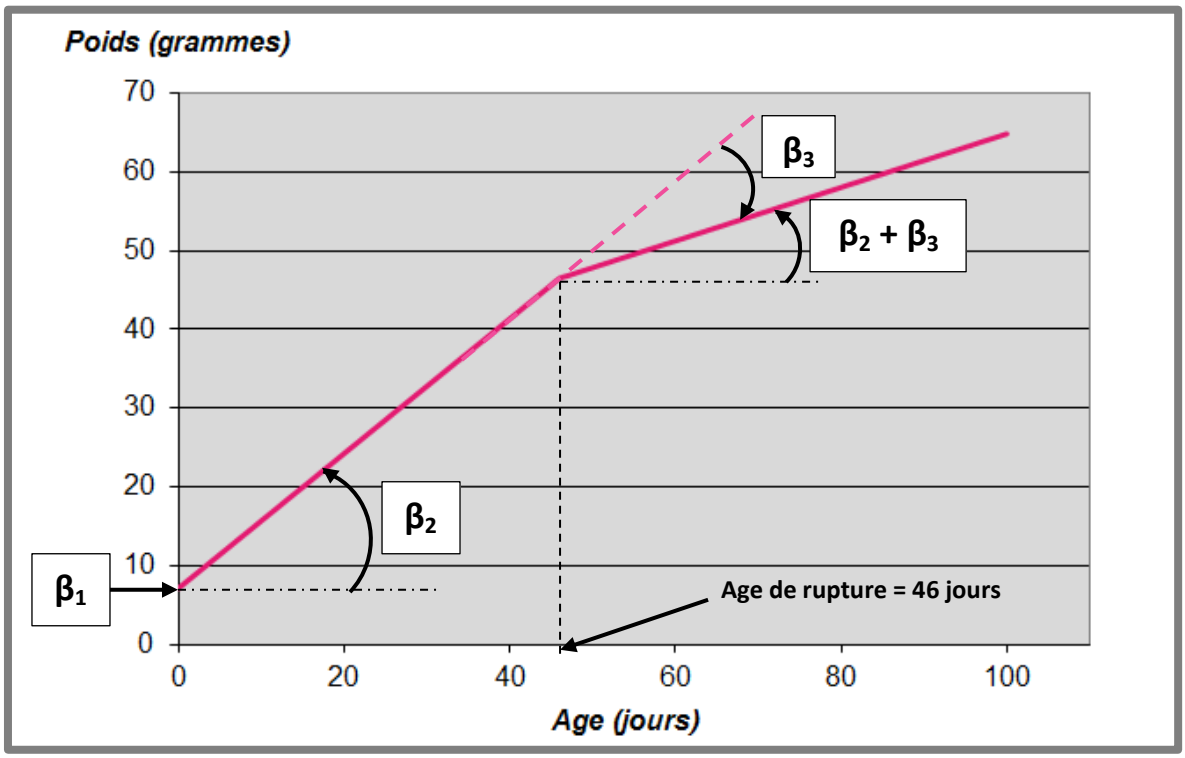
Avec agepesee_etoile = max (0, agepesee – age_rupture) = max (0, agepesee – 46). Avant le 47ème jour de vie, agepesee_etoile prend donc systématiquement la valeur 0.

De manière plus explicite, β_1 est l'ordonnée à l'origine, c'est-à-dire le poids estimé par le modèle à la naissance.

β_2 représente la pente de la courbe jusqu'à 46 jours de vie, c'est-à-dire le gain de poids quotidien moyen des microcèbes estimé par le modèle pour cette période. β_2 est un nombre positif.

Enfin, β_3 représente la différence de gain de poids moyen quotidien entre la première portion de la courbe (avant 46 jours) et la seconde (après 46 jours). La pente de la courbe à partir du 47ème jour de vie est donc égale à ($\beta_2 + \beta_3$). La Figure 15 ci-dessous permet de voir que, dans le cas présent, β_3 est négatif.

Figure 15 : Graphique schématisant la modélisation des données avec une équation linéaire autorisant une rupture de pente à 46 jours.



L'équation présentée ci-dessus est donc l'équation de base du modèle linéaire à une rupture de pente à 46 jours, potentiellement utilisable dans la suite de cette étude. Le programme permettant de rechercher l'âge de rupture de pente est disponible en Annexe 7.

c. Modèle non linéaire

Les recherches bibliographiques effectuées sur la modélisation non linéaire de la croissance ont conduit à la conclusion que le modèle le plus adapté à la modélisation de la croissance pour une espèce classique de mammifère était souvent un modèle de type Gompertz, contenant deux fonctions exponentielles composées. Un modèle de Gompertz simplifié avait d'ailleurs déjà été utilisé dans une étude précédente de l'UMR 7179 sur le microcèbe pour comparer l'évolution du poids et de la longueur du fémur des microcèbes au cours du temps, en fonction de la disponibilité de la nourriture dans le milieu (Canale *et al.* 2012).

Nous avons donc choisi d'utiliser ce type de modèle non linéaire avec notre jeu de données. Une première difficulté a été de trouver une expression du modèle de Gompertz possédant des paramètres assez facilement interprétables puisqu'il allait falloir discuter de l'impact de certaines caractéristiques du microcèbe, de sa mère ou de sa portée sur ces paramètres du modèle. La seconde difficulté a été de trouver un modèle avec lequel le programme de SAS fonctionnait au mieux.

Après exécution des programmes et comparaison des indices d'Akaike pour les différentes expressions de modèles de Gompertz trouvées dans la littérature, l'équation qui a été retenue pour cette étude est la suivante (Strathe *et al.* 2010) :

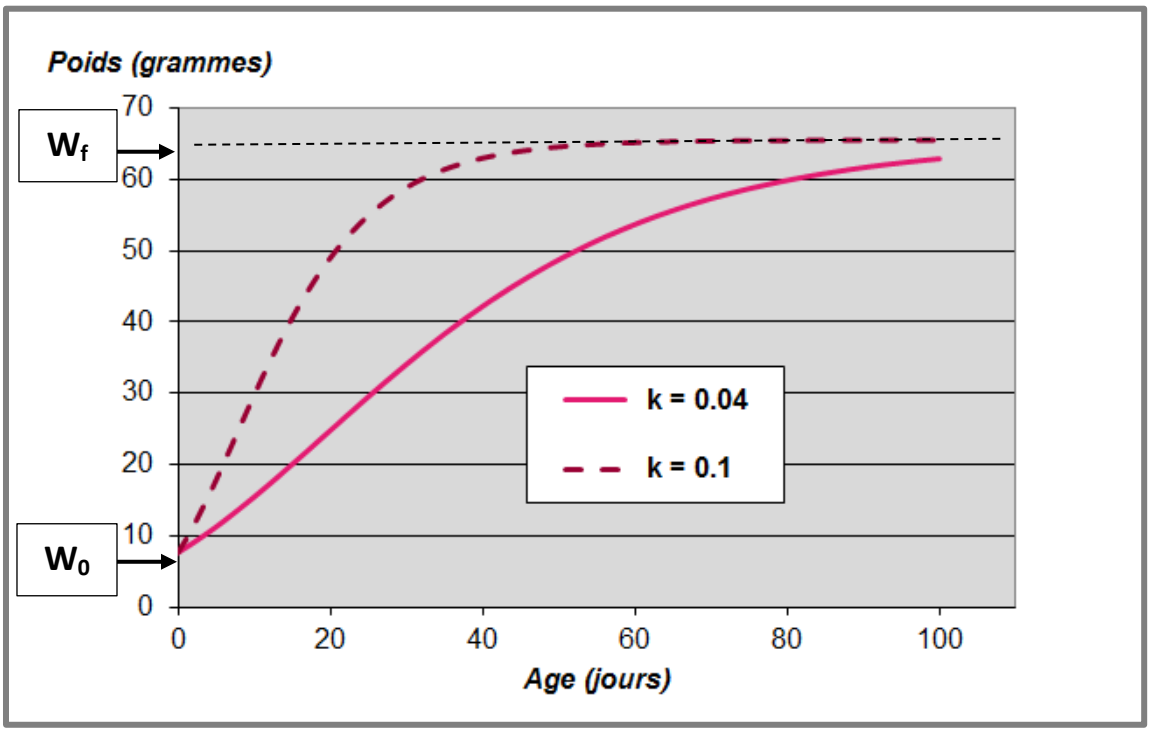
$$\textbf{Poids} = W_0 * \exp\{(1 - \exp(-k * Age)) * \ln(W_f / W_0)\}$$

Les 3 paramètres du modèle sont :

- W_0 : le poids à la naissance
- W_f : le poids théorique atteint en fin de croissance (mathématiquement, à t = + ∞)
- k : la vitesse à laquelle W_f est atteint.

La Figure 16 présente en trait plein la courbe tracée à partir de l'équation ci-dessus utilisée sur les données de poids des petits microcèbes. Dans le cas présent, les valeurs des paramètres du modèle étaient telles que W_0 = 7,6 ; W_f = 65,6 et k = 0,04. La seconde courbe tracée en pointillés est destinée à mieux se représenter l'impact d'une variation du paramètre k. Les valeurs de W_0 et W_f ont été gardées identiques mais la valeur de k a été choisie arbitrairement égale à 0,1. On constate que plus k est grand, plus la pente de la courbe à l'origine est importante.

Figure 16 : Graphique schématisant la modélisation non linéaire des données avec une équation de type Gompertz.



Le paramètre k sera le seul qui sera interprété pour évaluer la vitesse de croissance initiale des petits microcèbes. En effet, le modèle de Gompertz a tendance à surestimer les poids à la naissance et sous-estimer les poids en + l'infini (Zullinger *et al.* 1984).

Par ailleurs, si l'on souhaite estimer une différence de poids à un temps donné entre deux catégories d'individus, et si les données sont mesurées suffisamment fréquemment pour un même animal, il est alors préférable d'effectuer des tests statistiques testant deux moyennes de mesures comprises dans une fourchette de temps très restreinte (en choisissant la mesure la plus proche du centre de la fourchette si au moins deux mesures pour un animal sont incluses dans cette fourchette), plutôt que de comparer des valeurs théoriques issues d'un modèle, quel qu'il soit.

2. Prise en compte de la non-indépendance des données dans les deux types de modèles

Plusieurs mesures consécutives ont été réalisées pour un même microcèbe, ce qui signifie que la mesure de poids à un âge donné dépend obligatoirement du poids du microcèbe au temps de mesure précédent : ces mesures consécutives sont donc non indépendantes les unes des autres. Le fait qu'un animal soit pesé plusieurs fois au cours de l'étude introduit dans le cas présent, comme dans toutes les études longitudinales, une non-indépendance entre les différentes mesures pour cet animal. (Il faudra par ailleurs prendre en compte le fait que l'étude réalisée comporte des mesures répétées sur le même animal. Cette prise en compte prend l'apparence d'une ligne de code particulière dans les logiciels d'analyse). Il peut également y avoir plusieurs petits par portée, donc le poids d'un petit peut potentiellement être corrélé au poids de son/ses frère(s) et sœur(s). Enfin, une même femelle peut avoir plusieurs portées au cours de sa vie, ce qui crée une non-indépendance des portées entre-elles. Il y a donc 3 niveaux de non-indépendance des données. Dans cette étude, la prise en compte de cette non-indépendance est envisagée en introduisant dans le modèle des effets aléatoires portant sur les 3 niveaux (l'individu lui-même, la mère et la portée).

Pour chacun des deux modèles présentés, on peut envisager des effets aléatoires de l'individu, de la mère ou de la portée sur chacun des trois paramètres : (W_0, W_f et k) pour le modèle de Gompertz, (β_1, β_2 et β_3) pour le modèle linéaire. Il a donc été nécessaire de s'interroger dans un premier temps sur les paramètres (W_0, W_f et k ou β_1, β_2 et β_3) devant « recevoir » un effet aléatoire dans le modèle pour des raisons biologiques.

Pour le niveau individuel, tout laisse à penser que le temps n'aura pas le même impact sur le poids selon les animaux, qui possèdent des caractéristiques intrinsèques différentes les uns des autres. Certains peuvent avoir un métabolisme plus efficace que d'autres, et grossiraient ainsi plus vite, et il serait alors nécessaire de prendre en compte un effet aléatoire sur le paramètre k de notre modèle de Gompertz ainsi que sur les paramètres β_2 et β_3 du modèle linéaire. Pour les mêmes raisons, prendre en compte un effet aléatoire sur le paramètre W_f semble tout à fait justifié. En revanche, il est plus délicat de vouloir prendre en compte un potentiel effet aléatoire sur W_0 ou β_1. En effet, ces microcèbes sont tous issus des mêmes animaux capturés en milieu sauvage et ont donc une proximité génétique importante qui tendrait à douter de l'existence d'un effet aléatoire individuel sur le poids à la naissance (caractérisé par β_1 ou W_0 dans chacun de nos deux modèles).

Pour les effets aléatoires liés à la mère et à la portée, on peut envisager qu'ils existent sur le poids à la naissance et sur la vitesse de croissance initiale du microcèbe. En revanche, il ne semble pas pertinent de les prendre en compte sur le poids en fin de croissance (W_f) ou sur la vitesse de croissance tardive (donc sur β_3) dans la mesure où la croissance se termine après que les petits sont sevrés, donc une fois qu'ils ont accès à une nourriture *ad libidum*, ce qui annihile certainement les effets aléatoires liés au fait qu'une mère ait plusieurs portées au cours de sa vie, et/ou qu'il y ait plusieurs petits par portée.

Le Tableau 6 résume donc les effets aléatoires à prendre en compte pour les deux modèles.

Tableau 6 : Comparaison de la prise en compte des effets aléatoires de l'individu, la portée et la mère sur les paramètres des deux modèles.

	Modèle linéaire			Modèle de Gompertz		
	β_1	β_2	β_3	W_0	k	W_f
EA* individuel		✓	✓		✓	✓
EA* de la portée	✓	✓		✓	✓	
EA* de la mère	✓	✓		✓	✓	

* EA : Effet Aléatoire

Plusieurs niveaux d'effets aléatoires sont gérés par les procédures du logiciel SAS avec lequel a été réalisée cette étude permettant la modélisation avec des modèles linéaires (proc mixed ou proc glimmix par exemple.

$$\textbf{Poids} = (\beta_1 + EA_\beta_1_mere + EA_\beta_1_portee) + (\beta_2 + EA_\beta_2_ind + EA_\beta_2_mere + EA_\beta_2_portee) * agepesee + (\beta_3 + EA_\beta_3_ind) * agepesee_etoile$$

Avec EA_β_i_mere l'effet aléatoire de la mère sur β_i.

Avec EA_β_i_portee l'effet aléatoire de la portée sur β_i.

Avec EA_β_i_ind l'effet aléatoire de l'individu sur β_i.

Notons qu'en pratique cette équation n'a pas besoin d'être écrite puisque la prise en compte des effets aléatoires fait l'objet de lignes spécifiques dans le programme qui évitent la rédaction de telles expressions

Plusieurs niveaux d'effets aléatoires peuvent donc être gérés par les procédures de SAS dans le cas de modèles linéaires.

En revanche, les procédures pré-programmées du logiciel SAS ne permettent pas la prise en compte d'effets aléatoires sur plus de un niveau de non indépendance pour des modèles non linéaires. Il a donc été nécessaire, dans le cas du modèle de Gompertz, de télécharger une macro SAS, nommée *%nlinmix*, afin de permettre la prise en compte de plusieurs niveaux de non indépendance.

Une macro SAS est un programme que l'utilisateur écrit lui-même et qui nécessite un langage de programmation spécifique à SAS. Elle permet ainsi d'exécuter des calculs ou de faire tourner des modèles lorsque les programmes fournis par le logiciel ne le permettent pas. En l'occurrence la macro *%nlinmix* (téléchargeable grâce à ce lien : http://support.sas.com/kb/25/032.html) est programmée pour être capable d'effectuer par défaut 60 itérations mais elle peut converger avant. Elle fournit entre autres un indice d'Akaike, les valeurs des paramètres estimés du modèle (W_0, k et W_f) et les valeurs des différents coefficients associés aux effets fixes ainsi que leurs degrés de signification *p*. Les paramètres du modèle (en l'occurrence W_0, W_f et k) sont estimés à partir de valeurs fixées *a priori* par l'utilisateur à l'instar de la théorie bayésienne. La macro fournit par la suite une estimation *a posteriori* de ces paramètres. Dans cette étude, les valeurs *a priori* des paramètres W_0 et W_f ont été déterminées en calculant les moyennes des poids des microcèbes à 0 et 100 jours. La valeur *a priori* de k a été déterminée lors de la construction de la courbe (représentée en Figure 16) issue du fonctionnement du programme avec le modèle de Gompertz sans prise en compte d'effets aléatoires. Les valeurs *a priori* ont donc été fixées pour toutes les analyses telles que $W_0 = 7$, $W_f = 66$ et k = 0,04.

Cependant, il s'est avéré que, dans le cas présent, la macro ne supportait pas la prise en compte de plus d'un niveau de non indépendance, très certainement à cause du grand nombre de données (plus de 5000 mesures de poids de 645 microcèbes). Pour le modèle de Gompertz, il a donc fallu faire le choix de ne prendre en compte que niveau d'effet aléatoire le plus important : celui lié à l'individu lui-même.

Le modèle retenu est alors celui qui prend en compte deux effets aléatoires individuels, l'un sur k (noté « effet_aleat_k »), l'autre sur W_f (noté « effet_aleat_W_f ») :

$$\textbf{Poids} = W_0 * \exp\{(1 - \exp(-(k + \text{effet_aleat_k}) * \text{Age})) * \ln((W_f + \text{effet_aleat_}W_f) / W_0)\}$$

Même si cette partie expérimentale reprend le schéma général (« plan ») à adopter lors de la rédaction d'une publication scientifique, il est clair que, dans un tel document, un seul des deux modèles (celui choisi *in fine*) aurait été présenté. Deux types de modèles ont été présentés précédemment pour donner un exemple de réflexion dans chacun des deux cas. La partie suivante vise à comparer les avantages et les inconvénients de chacun des deux modèles dans le cas présent (même s'il est possible que le lecteur se soit déjà forgé une opinion… N'est-ce pas?!).

3. Modèle linéaire ou non linéaire : lequel choisir ?

Dans le cas présent, et avec le logiciel SAS en particulier, on constate qu'il est plus facile de prendre en compte plusieurs niveaux de non indépendance dans le cas où l'on choisit de travailler avec un modèle linéaire puisque les procédures à utiliser sont déjà pré-programmées ! Il n'y a donc pas besoin de chercher une macro adéquate à télécharger.

Il faut de plus mentionner que le temps pour faire tourner la macro *%nlinmix* (dans le cas du modèle non linéaire, donc) était très long (environ 15 minutes par analyse, au moins), indépendamment de la vétusté de l'ordinateur (les temps de calcul étaient similaires entre l'ordinateur sur lequel j'ai travaillé, âgé de 5 ans, et un ordinateur beaucoup plus récent, âgé de quelques mois). La durée importante des calculs était peut être due au grand nombre de données (> 5000 observations à traiter) et à la complexité du modèle mathématique utilisé (modélisation non linéaire affublée d'une modélisation à effets mixtes). Dans le cas de la procédure utilisée pour le modèle linéaire, le temps d'analyse n'excédait pas 3 minutes.

En pratique, lors de l'analyse avec le modèle non linéaire, il a été nécessaire d'essayer plusieurs équations pour trouver celle qui correspondait le mieux à nos données (meilleur indice d'Akaike) mais en ayant auparavant fait le tri parmi toutes les équations disponibles pour en trouver une dont les paramètres soient interprétables (c'est-à-dire « parlants » une fois que le logiciel donne des estimations chiffrées de ces paramètres). L'expression du modèle de Gompertz choisie est une expression non linéaire relativement compliquée. Elle possède l'avantage, selon plusieurs références bibliographiques, d'être adaptée en général pour modéliser la croissance juvénile des mammifères (Zullinger *et al.* 1984; Helmink *et al.* 2000; Lewis *et al.* 2002). Elle présente cependant l'inconvénient de faire tourner le programme pour que celui-ci estime les valeurs de 3 paramètres, dont deux (W_0 et W_f) ne sont finalement pas interprétables. Il faut en effet préférer les analyses ponctuelles à des âges précis aux valeurs pondérales estimées par le modèle aux valeurs de W_0 et W_f estimées (Zullinger *et al.* 1984). En choisissant d'emblée d'utiliser un modèle linéaire simple, on s'affranchit des étapes longues et fastidieuses de tri des différentes expressions du modèle. Dans le cas d'un modèle linéaire à rupture(s) de pente(s), la recherche du (ou des) point(s) de rupture constitue certes une étape supplémentaire par rapport à un modèle sans rupture de pente, mais qu'il est facile d'automatiser informatiquement à l'aide d'une boucle dans le programme. Seul le paramètre β_1 (c'est-à-dire, correspondant à l'ordonnée à l'origine moyenne estimée par le logiciel) peut être considéré comme « non utile » car il est facile (et plus juste) de faire une moyenne ponctuelle pour obtenir la véritable valeur moyenne du poids à la naissance.

Les éléments mis en lumière dans la partie 2 de ces matériel et méthodes ont montré qu'il aurait été nécessaire de prendre en compte les effets aléatoires dus à la non-indépendance des données de poids des animaux d'une même portée ou d'une même mère sur les paramètres k et W_f du modèle. Cependant, alors que la macro avait été choisie pour son avantage théorique de pouvoir supporter plusieurs niveaux de non-indépendance pour des modèles non linéaires mixtes, les effets aléatoires de la mère et de la portée n'ont pas pu être pris en compte en raison des limites informatiques dans ce contexte précis. Ces effets aléatoires étaient en revanche tout à fait gérés par la procédure utilisée lors de la modélisation linéaire.

Enfin, il est possible de comparer objectivement les deux modélisations en utilisant les critères d'Akaike. Avec le modèle linéaire autorisant une rupture de pente à 46 jours de vie, l'indice d'Akaike avait pour valeur 33369. Lors de l'analyse avec le modèle non linéaire, l'indice d'Akaike valait 33956.

⚠ Pour comparer les critères d'Akaike de deux modèles, il faut avoir fait fonctionner les modèles sur un même jeu de données (c'est-à-dire sur le même nombre de lignes sans données manquantes sur les variables figurant dans ces deux modèles) !

Toutes ces comparaisons sont synthétisées dans le Tableau 7 ci-après.

Tableau 7 : Présentation des avantages et inconvénients des deux types de modèles (modèle linéaire *versus* modèle de Gompertz).

	Modèle linéaire	**Modèle non linéaire (Gompertz)**
Facilité d'analyse avec le logiciel SAS	Procédure pré-programmée dans le logiciel SAS	Macro à télécharger, non disponible dans le package de base
Temps d'analyse	Assez court, n'excédant pas 3 minutes par analyse	Long, environ 15 minutes par analyse
Choix préalable du modèle	Un seul type d'équation linéaire avec éventuelle détermination automatisée de point(s) de rupture	Nombreuses expressions du modèle de Gompertz : choix fastidieux d'une équation selon le contexte
Prise en compte des 3 niveaux d'effets aléatoires	Possible théoriquement et bien gérée par la procédure glimmix en pratique sur nos données	Possible théoriquement avec la macro *%nlinmix* mais impossible en pratique sur nos données
Critères d'Akaike	33369	33956

Tous ces arguments ont permis de choisir quel modèle était le plus adapté pour répondre à la problématique posée dans le cas présent : modéliser au mieux la croissance pondérale des jeunes microcèbes afin d'analyser les associations entre cette croissance et le sexe du microcèbe, la taille de la portée et la parité de la mère. Ainsi, le modèle retenu dans la suite de l'étude est le modèle linéaire présentant une rupture de pente au 46ème jour de vie. Un exemple de modélisation avec l'équation de Gompertz sera tout de même présenté à titre pédagogique dans la toute première partie d'analyse brute qui concerne l'association entre la croissance du microcèbe et le nombre de petits dans la portée.

On fixera durant toute la durée de l'étude le risque d'erreur de première espèce α égal à 5 %.

III. Résultats

Les quatre premières parties qui vont suivre présentent l'échantillon et les résultats issus de l'analyse brute de l'impact de 3 différentes caractéristiques (le sexe du microcèbe, la parité de la mère et le nombre d'autres petits dans la portée) sur les paramètres du modèle.

La cinquième et dernière partie présente un modèle complet ajusté sur toutes les variables d'intérêt principal.

A. Description de l'échantillon

Les données de poids de 645 microcèbes ont été utilisées dans cette étude. On dénombre 349 femelles et 296 mâles.

La table de données comprend des petits microcèbes issus de 366 mères différentes (279 primipares et 87 multipares).

Toutes les portées conservées (481 portées en tout) sont des « portées élevées », c'est-à-dire que tous les animaux nés ont atteint le 100ème jour de vie. Au total, 128 microcèbes proviennent de portées où ils sont petits uniques, 357 microcèbes ont été élevés avec un autre petit (portée de 2 petits), 156 microcèbes ont été élevés avec 2 autres petits (portée de 3 petits), et enfin 4 microcèbes ont été élevés avec 3 autres petits (portée de 4 petits).

Les effectifs énoncés précédemment sont repris dans le Tableau 8.

Tableau 8 : Effectifs utilisés lors de l'étude de la dynamique de croissance des microcèbes.

Catégorie	Pourcentages (n) (N=645)
Microcèbes femelles	54 % (349)
Microcèbes issus de mères primipares	43 % (279)
Microcèbes élevés dans des portées de plus d'un petit	80 % (517)

Le programme permettant le calcul de ces pourcentages avec le logiciel SAS à partir de la table de données complète est disponible en Annexe 8.

B. Effet brut du nombre de petits dans la portée sur la croissance du microcèbe

Dans un but pédagogique, nous présenterons ici une partie de l'analyse avec le modèle linéaire autorisant une rupture de pente au 46ème jour de vie, mais aussi avec le modèle de Gompertz, afin de montrer la marche à suivre pour effectuer ce type d'analyse en modélisation non linéaire.

Le Tableau 9 récapitule à nouveau les proportions de microcèbes de la base de données ayant grandi seuls, avec un seul, deux ou trois autre(s) petit(s).

Tableau 9 : Pourcentages et effectifs de l'échantillon de 645 microcèbes suivant le nombre de petits dans la portée.

	Pas d'autre petit	1 autre petit	2 autres petits	3 autres petits	TOTAL
Nombre de microcèbes	128	357	156	4	645
Pourcentages	20 %	55 %	25 %		100 %

Au vu du faible nombre de microcèbes (4 animaux) appartenant à des portées composées de 4 petits (1 petit plus 3 autres petits dans la portée), les catégories « 2 autres petits » et « 3 autres petits » ont été fusionnées lors de l'analyse statistique pour former une seule classe « deux ou 3 autres petits», contenant 160 microcèbes.

1. Analyse après création d'une variable binaire (utilisation des deux modèles)

a. Présentation de la variable et des équations des modèles

La première partie de cette analyse brute, pour laquelle nous allons détailler les deux modèles, a été effectuée avec une variable binaire appelée «petit_unique» (codée 1 si le petit est seul dans la portée, et 0 sinon) afin de comparer la croissance des petits uniques à celle des petits non uniques dans leur portée (c'est-à-dire ayant au moins un frère ou une sœur).

Les équations utilisées pour modéliser le poids des petits microcèbes en fonction de leur âge dans cette analyse sont les suivantes :

$$\textbf{Poids} = \beta_1 + \beta_2 * \text{agepesee} + \beta_3 * \text{agepesee_etoile} + \alpha_1 * \text{petit_unique} + \alpha_2 * \text{agepesee} * \text{petit_unique} + \alpha_3 * \text{agepesee_etoile} * \text{petit_unique}$$

NB : Les 3 niveaux d'effets aléatoires définis dans les matériels et méthodes ont été pris en compte.

$$\textbf{Poids} = (W_0 * (1 + \alpha * \text{petit_unique})) * \exp\{(1 - \exp(-(k * (1 + \beta * \text{petit_unique})) * \text{Age})) * \ln((W_f * (1 + \delta * \text{petit_unique})) / (W_0 * (1 + \alpha * \text{petit_unique})))\}$$

NB : Les effets aléatoires sur k et W_f liés à l'individu ne figurent pas dans cette équation pour des raisons de lisibilité mais ont bien évidemment été pris en compte.

Les programmes correspondant à cette analyse sous SAS figurent en Annexe 9. Les résultats et leurs interprétations pour chacun des deux modèles sont présentés ci-après.

b. Résultats de l'analyse avec le modèle linéaire

Les estimations des paramètres du modèle linéaire données par le logiciel sont présentées dans le Tableau 10 ci-dessous.

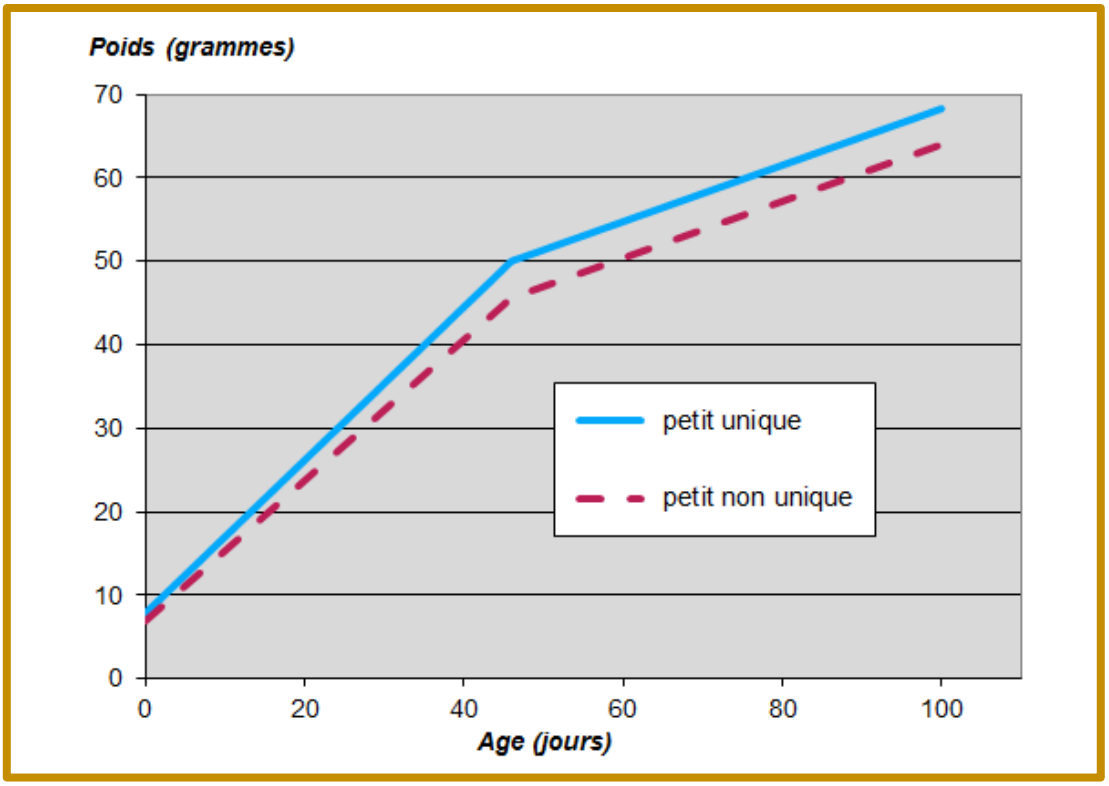
Tableau 10 : Estimations des paramètres du modèle linéaire données par le logiciel lors de l'analyse avec la variable binaire « petit_unique ».

Paramètre	Valeur estimée	SE*	p
β_1	6,8877	0,1241	<0,01
β_2	0,8421	0,0080	<0,01
β_3	- 0,5021	0,0139	<0,01
α_1	0,9271	0,2372	<0,01
α_2	0,0752	0,0163	<0,01
α_3	- 0,0769	0,0310	0,0131

* SE : Standard Error

La courbe correspondante est disponible en Figure 17.

Figure 17: Allure de croissance des petits uniques ou non uniques en analyse brute avec le modèle linéaire.



Cette analyse brute a montré que :

Parmi les microcèbes uniques dans leur portée ou parmi les microcèbes non uniques dans leur portée, le gain de poids moyen quotidien (GMQ) était plus faible dans la deuxième période (après 46 jours de vie) que dans la première (significativité de β_3 et de $\beta_3+\alpha_3$, respectivement).

Les petits uniques étaient significativement plus lourds à la naissance (0,9 grammes (±SE = 0,2)) que les petits non uniques (significativité de α_1).

Les valeurs des GMQ pour les deux catégories aux deux périodes sont retranscrites dans le Tableau 11.

Tableau 11 : Valeurs des GMQ pour les petits non uniques et les petits uniques lors de l'analyse brute.

	Petits non uniques	Petits uniques
GMQ avant le 46ème jour de vie	β_2 = **0,8421**	$\beta_2 + \alpha_2$ = **0,9173**
GMQ après le 46ème jour de vie	$\beta_2 + \beta_3$ = **0,3400**	$\beta_2 + \beta_3 + \alpha_2 + \alpha_3$ = **0,3383**

D'après les valeurs de ce tableau, les petits uniques grandissaient un peu plus rapidement que les petits non uniques avant 46 jours de vie (significativité de α_2). Il existait donc une association brute entre la vitesse de croissance des microcèbes et le nombre de petits de la portée durant les 46 premiers jours de vie. En revanche, les vitesses de croissance des microcèbes des deux catégories n'étaient pas significativement différentes après 46 jours de vie ($\alpha_2+\alpha_3$ non significativement différente de zéro).

c. Résultats de l'analyse avec le modèle non linéaire

Les estimations des paramètres du modèle non linéaire données par le logiciel sont présentées dans le Tableau 12 ci-dessous.

Tableau 12 : Estimations des paramètres du modèle non linéaire données par le logiciel lors de l'analyse avec la variable binaire « petit_unique ».

Paramètre	Valeur estimée	SE*	p
W_0	8,6352	0,1156	<0,01
k	0,0366	0,0008	<0,01
W_f	63,4494	0,9626	<0,01
α	- 0,1084	0,0094	<0,01
β	0,0820	0,0208	<0,01
δ	0,0381	0,0135	<0,01

* SE : Standard Error

Si l'on reprend l'équation de Gompertz citée plus haut, les valeurs présentées ici pour W_0, W_f et k correspondent à la classe possédant la valeur « 0 » en binaire, c'est-à-dire aux microcèbes issus de portées de plusieurs petits.

Les valeurs des estimations pour les coefficients correspondant à un petit unique s'obtiennent de la façon suivante (d'après l'équation mentionnée plus haut, exemple donné pour k) :

$k_{petit\ unique} = k * (1+ \beta * 1) = k * (1+ \beta) = k * (1+ 0.0820) = 0{,}0396$

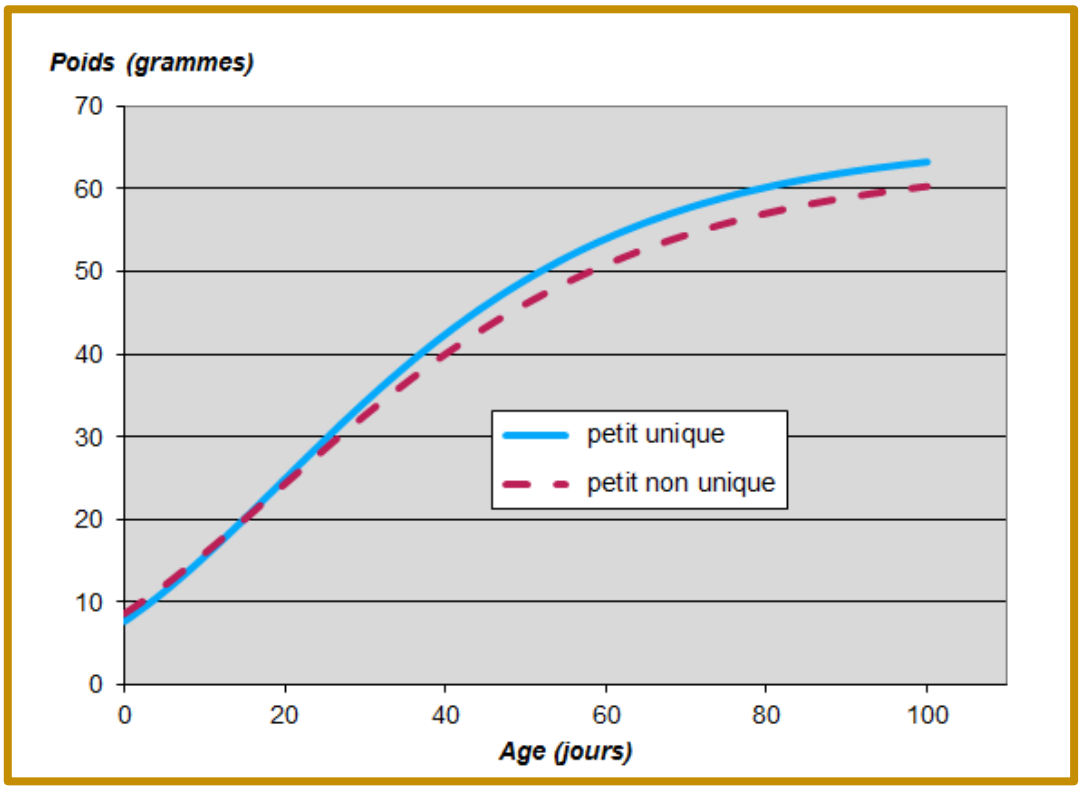
On présente donc en général les résultats dans un tableau plus synthétique tel que celui présenté ci-dessous (Tableau 13). Un graphique (Figure 18) permet de mieux visualiser les résultats.

Tableau 13 : Valeurs des estimations des paramètres du modèle non linéaire pour les deux groupes (petit unique versus petit non unique).

	Petits uniques	Petits non uniques
W_0 *	7,6991	8,6352
k *	0,0396	0,0366
W_f *	65,8662	63,4494

* = différence significative entre les deux classes.

Figure 18 : Graphique représentant l'allure de croissance des petits uniques ou non uniques en analyse brute avec le modèle de Gompertz.



Les microcèbes seuls dans leur portée avaient une croissance plus rapide que ceux qui n'étaient pas petits uniques. Les k obtenus pour les deux groupes étaient significativement différents l'un de l'autre (Tableau 13). Il existait donc une association brute entre la vitesse de croissance des microcèbes dans les premiers jours de vie et la composition de la portée (petit unique *versus* petit non-unique).

2. Analyse à l'aide de la variable quantitative correspondant au nombre d'autres petits dans la portée dans le modèle linéaire

a. Equation du modèle utilisée

L'équation utilisée lors de cette analyse est la suivante :

$$\textbf{Poids} = \beta_1 + \beta_2 * \textbf{agepesee} + \beta_3 * \textbf{agepesee_etoile} + \alpha_1 * \textbf{nb_autres_petits} + \alpha_2 * \textbf{agepesee} * \textbf{nb_autres_petits} + \alpha_3 * \textbf{agepesee_etoile} * \textbf{nb_autres_petits}$$

NB : Les 3 niveaux d'effets aléatoires définis dans les matériels et méthodes ont été pris en compte.

b. Vérification de la linéarité

La vérification de la linéarité est une étape préalable nécessaire à toute analyse faisant intervenir une variable quantitative (quel que soit le type de modèle, linéaire ou non).

La variable quantitative à étudier doit être transformée en variable en classes (au minimum 3 classes). Il s'agit ensuite de remplacer la variable quantitative par les différentes classes dans le modèle. Trois classes ont donc été définies à partir de la variable quantitative « nb_autres_petits » : la « classe 0 » correspondait à un nombre d'autres petits égal à 0, la « classe 1 » à un nombre d'autres petits égal à 1 et la « classe 2 » à un nombre d'autres petits supérieur ou égal à 2.

NB : pour introduire une variable en classes dans le modèle, il faut définir une classe, appelée « classe de référence » que l'on sortira du modèle. La classe 0 a été fixée arbitrairement comme classe de référence dans notre exemple.

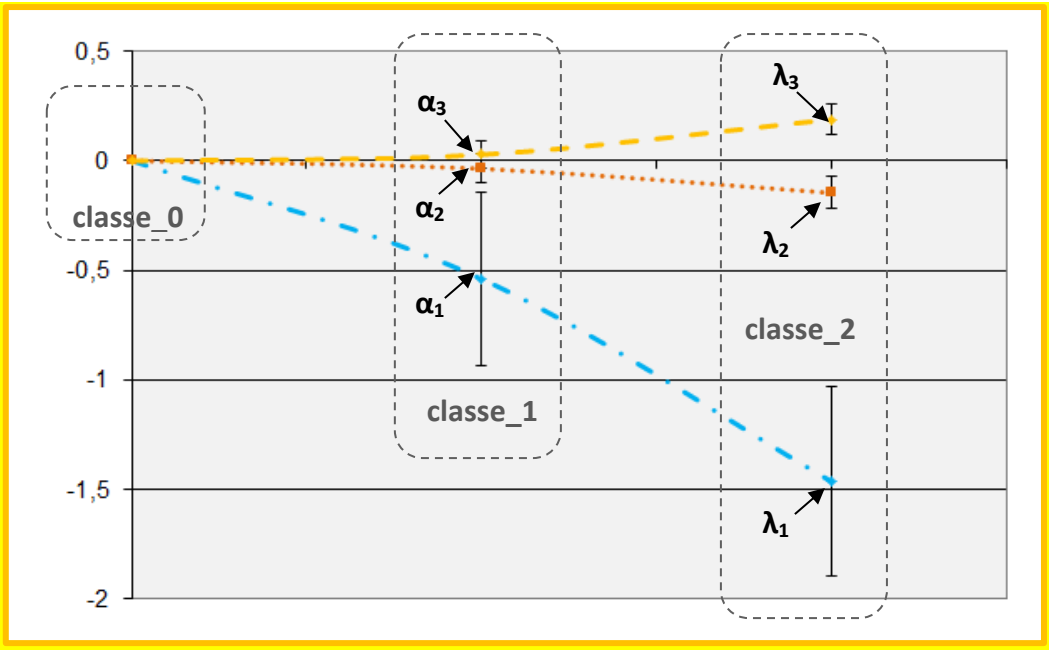
Le modèle utilisé pour la vérification de la linéarité était donc le suivant :

$$\textbf{Poids} = \beta_1 + \beta_2 * \textbf{agepesee} + \beta_3 * \textbf{agepesee_etoile} + \alpha_1 * \textbf{classe_1} + \lambda_1 * \textbf{classe_2} + \alpha_2 * \textbf{agepesee} * \textbf{classe_1} + \lambda_2 * \textbf{agepesee} * \textbf{classe_2} + \alpha_3 * \textbf{agepesee_etoile} * \textbf{classe_1} + \lambda_3 * \textbf{agepesee_etoile} * \textbf{classe_2}$$

Vérifier la linéarité consiste à utiliser les valeurs estimées des coefficients α_i et λ_i (ainsi que les erreurs standards associées) fournies par le logiciel une fois que le programme a tourné. La valeur des coefficients associés à la « classe 0 » est nulle par convention. Ces valeurs doivent être placées sur un graphique comme présenté sur la Figure 19.

Le programme sous SAS est disponible en fin d'Annexe 9.

Figure 19 : Graphique représentant en ordonnées les valeurs des coefficients α_i et λ_i (avec leur intervalle de confiance à 95 %) en fonction des 3 classes de la variable quantitative "nb_autres_petits" (0 autre petit, 1 autre petit, ≥ 2 autres petits) en abscisse.



S'il est possible de tracer une droite passant par les 3 points (dont les valeurs sont associées à chacune des 3 classes) de chaque courbe du graphique, alors la linéarité est vérifiée.

Dans le cas présent, il a été nécessaire de tracer 3 courbes puisque l'on retrouve chacune des classes mentionnée 3 fois dans l'équation du modèle. Le graphique a montré qu'il était possible de tracer des droites passant par les points d'ordonnées α_i et λ_i. Par conséquent la linéarité de la variable « nb_autres_petits » a été vérifiée et l'analyse peut être effectuée en utilisant cette variable quantitative.

c. Résultats de l'analyse

Les résultats de l'analyse (avec l'équation évoquée en partie 2.a) sont présentés dans le Tableau 14.

Tableau 14 : Estimations des paramètres du modèle linéaire données par le logiciel lors de l'analyse avec la variable quantitative correspondant au nombre d'autres petits dans la portée.

Paramètre	Valeur estimée	SE*	p
β_1	7,9028	0,1781	<0,01
β_2	0,9375	0,0119	<0,01
β_3	- 0,6211	0,0224	<0,01
α_1	- 0,7444	0,1358	<0,01
α_2	- 0,0745	0,0092	<0,01
α_3	0,0989	0,0180	<0,01

* SE : Standard Error

D'une part, on remarque que plus le nombre d'autres petits était grand, plus le poids estimé à la naissance était petit ($\alpha_1 < 0$).

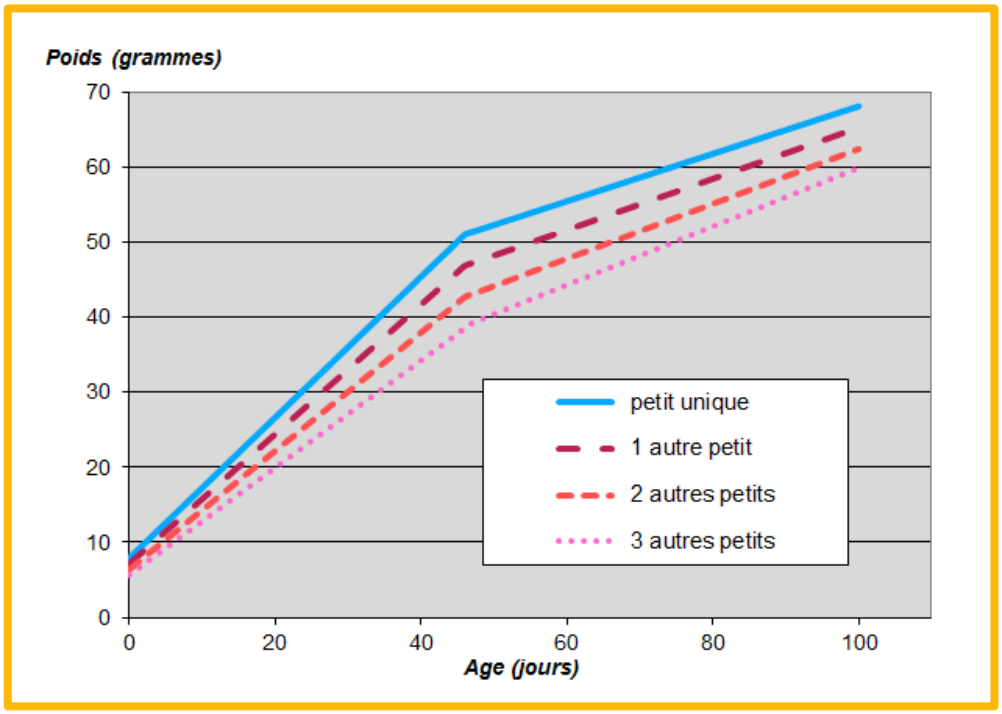
D'autre part, les conclusions que l'on peut tirer de la lecture de ces résultats rejoignent celles présentées pour l'analyse avec la variable binaire « petit_unique » à l'échelle de notre échantillon de microcèbes. Le Tableau 15 donne les valeurs des GMQ avant et après 46 jours pour des tailles de portée différentes.

Tableau 15 : Valeurs des GMQ selon la composition de la portée, lors de l'analyse brute.

	Petits uniques	**1 autre petit**	**2 autres petits**	**3 autres petits**
GMQ avant le 46ème jour de vie	β_2 = **0,9375**	$\beta_2 + \alpha_2$ = **0,8630**	$\beta_2 + 2\alpha_2$ = **0,7886**	$\beta_2 + 3\alpha_2$ = **0,7141**
GMQ après le 46ème jour de vie	$\beta_2 + \beta_3$ = **0,3164**	$\beta_2 + \beta_3 + \alpha_2 + \alpha_3$ = **0,3408**	$\beta_2 + \beta_3 + 2\alpha_2 + 2\alpha_3$ = **0,3652**	$\beta_2 + \beta_3 + 3\alpha_2 + 3\alpha_3$ = **0,3897**

D'après le Tableau 15, plus le nombre de petits était grand plus le GMQ était faible dans la première période (avant 46 jours). En revanche, après 46 jours la tendance s'inversait. Le graphique ci-dessous (Figure 20) permet de visualiser cette modélisation de croissance dans notre échantillon.

Figure 20 : Graphique représentant l'allure de croissance des petits microcèbes en fonction pour différentes tailles de portée en analyse brute avec le modèle linéaire.



C. Effet brut du sexe sur la croissance du microcèbe

L'équation utilisée lors de cette analyse est la suivante :

$$\textbf{Poids} = \beta_1 + \beta_2 * \text{agepesee} + \beta_3 * \text{agepesee_etoile} + \delta_1 * \text{male} + \delta_2 * \text{agepesee} * \text{male} + \delta_3 * \text{agepesee_etoile} * \text{male}$$

Les résultats de cette analyse fournis par le logiciel sont présentés dans le Tableau 16.
Le programme correspondant dans SAS est disponible en Annexe 10.

Tableau 16 : Estimations des paramètres du modèle linéaire données par le logiciel lors de l'analyse avec la variable binaire « male».

Paramètre	Valeur estimée	SE*	p
β_1	7,1256	0,128	<0,01
β_2	0,8802	0,0094	<0,01
β_3	- 0,4964	0,0171	<0,01
δ_1	0,1318	0,1891	0,4858
δ_2	- 0,0396	0,0139	0,0044
[4]δ_3	- 0,0464	0,0253	0,0665

* SE : Standard Error

Le Tableau 17 donne les valeurs du GMQ avant et après le 46ème jour de vie.

Tableau 17 : Valeurs des GMQ obtenus lors de l'analyse brute chez les mâles et les femelles.

	Femelles	Mâles
GMQ avant le 46ème jour de vie	β_2 = **0,8802**	$\beta_2 + \delta_2$ = **0,8406**
GMQ après le 46ème jour de vie	$\beta_2 + \beta_3$= **0,3838**	$\beta_2 + \delta_2 + \beta_3 + \delta_3$ = **0,2978**

Cette analyse brute nous permet donc seulement de dire qu'avant 46 jours de vie, les mâles grandissaient un peu moins vite que les femelles (significativité de δ_2). La différence de GMQ entre les mâles et les femelles était la même avant 46 jours qu'après 46 jours (non significativité de $\delta_2+\delta_3$). On peut donc conclure que les mâles grandissaient un peu moins vite que les femelles durant les 100 premiers jours de vie.

[4] Remarque : la valeur de p associée à δ_3 est proche de 0,05. Par conséquent, si δ_3 dans la population est effectivement différent de zéro, une taille d'échantillon plus importante aurait pu permettre d'obtenir un p<0,05.

D. Effet brut de la parité de la mère sur la croissance du microcèbe

L'équation utilisée lors de cette analyse est la suivante :

$$\textbf{Poids} = \beta_1 + \beta_2 * \textbf{agepesee} + \beta_3 * \textbf{agepesee_etoile} + \lambda_1 * \textbf{multipare} + \lambda_2 * \textbf{agepesee} * \textbf{multipare} + \lambda_3 * \textbf{agepesee_etoile} * \textbf{multipare}$$

Les résultats de cette analyse fournis par le logiciel sont présentés dans le Tableau 18.
Le programme correspondant à cette analyse est disponible en Annexe 11.

Tableau 18 : Estimations des paramètres du modèle linéaire données par le logiciel lors de l'analyse avec la variable binaire « multipare».

Paramètre	Valeur estimée	SE*	p
β_1	7,0833	0,1342	<0,01
β_2	0,8593	0,0089	<0,01
β_3	- 0,5258	0,0160	<0,01
λ_1	0,1001	0,2066	0,6280
λ_2	0,0010	0,0138	0,9439
λ_3	0,0214	0,0255	0,4008

* SE : Standard Error

Les valeurs des paramètres λ_1, λ_2 et λ_3 étaient non significativement différentes de 0. Il y a donc des chances pour que, réellement, il n'existe aucune différence de croissance des microcèbes selon la parité de leur mère.

Cette analyse a permis de mettre en évidence qu'il n'existait pas d'association brute entre la parité de la mère et la croissance du petit microcèbe dans notre échantillon.

E. Ajustement : prise en compte de facteurs de confusion potentiels

Après avoir réalisé les analyses brutes présentées dans les paragraphes précédents, il a été nécessaire de travailler avec un modèle ajusté sur les variables d'intérêt principal afin d'étudier les associations avec la croissance pondérale indépendamment d'autres éventuelles associations. Afin de faciliter l'interprétation des résultats, ce modèle ajusté fait intervenir seulement des variables binaires, déjà définies précédemment : « male », « petit_unique » et « multipare ».

Remarque : Dans le cas présent, le problème de « case vide » ne se pose pas et l'analyse ajustée peut être menée sur le même fichier de données que les analyses brutes.

L'équation du modèle ajusté est la suivante :

$$\textbf{Poids} = \beta_1 + \beta_2 * \text{agepesee} + \beta_3 * \text{agepesee_etoile} + \alpha_1 * \text{petit_unique} + \alpha_2 * \text{agepesee} * \text{petit_unique} + \alpha_3 * \text{agepesee_etoile} * \text{petit_unique} + \delta_1 * \text{male} + \delta_2 * \text{agepesee} * \text{male} + \delta_3 * \text{agepesee_etoile} * \text{male} + \lambda_1 * \text{multipare} + \lambda_2 * \text{agepesee} * \text{multipare} + \lambda_3 * \text{agepesee_etoile} * \text{multipare}$$

Les résultats obtenus sous SAS sont présentés dans le Tableau 19 et explicités dans le Tableau 20 présentant les GMQ pour les différentes catégories.

Le programme pour cette analyse ajustée figure en Annexe 12.

Tableau 19 : Estimations des paramètres du modèle linéaire données par le logiciel lors de l'analyse ajustée sur les variables d'intérêt principal.

Paramètre	Valeur estimée	SE*	p
β_1	6,7144	0,1621	<0,01
β_2	0,8544	0,0111	<0,01
β_3	- 0,4889	0,0208	<0,01
α_1	0,9478	0,2389	<0,01
α_2	0,0770	0,0163	<0,01
α_3	- 0,0743	0,0310	0,0166
δ_1	0,2123	0,1477	0,1507
δ_2	- 0,0325	0,0116	0,0051
δ_3	- 0,0438	0,0248	0,0769
λ_1	0,1915	0,2049	0,3500
λ_2	0,0061	0,0136	0,6544
λ_3	0,0159	0,0254	0,5318

* SE : Standard Error

Tableau 20 : Valeurs des GMQ obtenus pour les différentes catégories lors de l'analyse ajustée.

	Petits non uniques femelles	Petits non uniques mâles	Petits uniques femelles	Petits uniques mâles
GMQ avant le 46ème jour de vie	$\beta_2 =$ **0,8544**	$\beta_2 + \delta_2 =$ **0,8219**	$\beta_2 + \alpha_2 =$ **0,9314**	$\beta_2 + \alpha_2 + \delta_2 =$ **0,8989**
GMQ après le 46ème jour de vie	$\beta_2 + \beta_3 =$ **0,3655**	$\beta_2 + \beta_3 + \delta_2 + \delta_3 =$ **0,2862**	$\beta_2 + \beta_3 + \alpha_2 + \alpha_3 =$ **0,3673**	$\beta_2 + \beta_3 + \alpha_2 + \alpha_3 + \delta_2 + \delta_3 =$ **0,2919**

Grâce à ces résultats, on peut notamment conclure que :

- Indépendamment du sexe du petit et de la parité de la mère, les petits uniques dans leur portée grandissaient plus vite que les petits ayant au moins un frère ou une sœur ($\alpha_2 > 0$ et significativement différent de 0, de même que la somme $\alpha_2 + \delta_2$) durant les 46 premiers jours de vie. Les vitesses de croissance des microcèbes des deux catégories de sexe étaient par ailleurs légèrement moins élevées pour les petits non uniques après 46 jours de vie.

- Indépendamment du statut de petit unique dans la portée et indépendamment de la parité de la mère, les petits mâles grandissaient légèrement moins vite que les femelles dans les premiers jours de vie ($\delta_2 < 0$ et significativement différent de 0). La différence de GMQ entre les mâles et les femelles persistait par la suite après 46 jours de vie.

- Indépendamment du sexe du petit et du statut de petit unique dans la portée, la parité de la mère n'était pas associée à la dynamique de croissance du petit microcèbe (λ_2 et $\lambda_2 + \lambda_3$ non significativement différents de 0, respectivement).

IV. Discussion

Il faut tout d'abord noter que les différences significatives sont à interpréter d'un point de vue épidémiologique et biologique. En effet, dans cette étude, de nombreux paramètres sont significativement différents de zéro et conduisent à l'interprétation qu'il existe une différence réelle entre telle et telle catégories de microcèbes. Même une très petite différence entre deux paramètres, ne signifiant quasiment rien d'un point de vue biologique (par exemple non visible sur un graphique représentant les deux courbes car elle représente quelques millièmes de grammes), peut apparaitre significative statistiquement. Il faut donc rester prudent et garder un esprit critique en interprétant les résultats dans leur contexte épidémiologique ou biologique. En cas de doute, il est toujours possible de tracer les courbes (comme cela a été fait lors de l'analyse brute), afin de regarder si la différence se voit « à l'œil nu » sur un graphique. Rappelons que ce sont les observations biologiques et/ou cliniques qui amènent en premier lieu à se poser des questions auxquelles la (bio)statistique peut *ensuite* fournir des réponses. Autrement dit, l'analyse statistique dans le domaine médical est un outil de réponse à des questions que l'on a des raisons biologiques ou cliniques de se poser.

L'analyse ajustée a montré qu'indépendamment du statut de petit unique dans la portée et indépendamment de la parité de la mère, les petits mâles grandissaient légèrement moins vite que les femelles dans les premiers jours de vie. On retrouve en général plutôt la tendance inverse chez de nombreuses espèces de mammifères telles que les ruminants sauvages ou les chiens (Birgersson et Ekvall, 1997; Helmink *et al,.* 2000). Toutefois, dans la mesure où l'espèce étudiée est monomorphique à l'âge adulte (Wright, 1999), les résultats obtenus ne semblent pas aberrants de prime abord. Ils vont cependant à l'encontre de ce que l'on aurait pu attendre suite à la lecture de l'étude de Colas (Colas, 1999), qui rapporte que la mère microcèbe aurait tendance à allouer davantage de ressources à ses petits mâles. Cette tendance à l'augmentation des contacts mère-petit en présence de petits de sexe mâle décrite dans cette étude sur les microcèbes n'est pas rapportée chez d'autres espèces de primates, peut être en raison du fait que les études éthologiques portent principalement sur les singes (en particulier les grands singes), qui n'ont en général qu'un seul petit à la fois.

L'analyse ajustée a également mis en évidence qu'indépendamment du sexe du petit et du nombre de petits dans la portée, la parité de la mère n'était pas associée à la dynamique de croissance du petit microcèbe. Ces résultats n'étaient pas spécialement attendus : deux facteurs décrits dans la littérature, l'un physiologique, l'autre comportemental, auraient pu contribuer à faire apparaitre une différence de croissance du jeune microcèbe associée à la parité de sa mère. On aurait en effet pu penser qu'à partir de sa deuxième lactation, la mère microcèbe aurait donné davantage de lait, autorisant une meilleure croissance chez les petits de la portée. L'accroissement de la production laitière au fur et à mesure des lactations successives avait déjà été montré dans de très nombreuses études chez les vaches laitières (Doreau *et al.*, 1991; Pérochon *et al.,* 1996; Phillips et Rind, 2001). De plus, il avait été mis en évidence dans une étude comportementale sur les microcèbes (Colas, 1999) qu'une mère multipare interagissait davantage avec ses petits si elle avait déjà eu une expérience précédente de maternité, leur autorisant un meilleur accès alimentaire.

Ceci avait été également prouvé dans une étude sur les campagnols (*Microtus ochrogaster*), qui a mis en évidence que les mères multipares quittaient moins le nid que les primipares, et avaient donc des petits dont la croissance était plus rapide et le taux de survie meilleur que les petits issus de mères primipares (Wang et Novak, 1994).

Dans l'étude sur la croissance des faons de daim (*Dama dama*) (Birgersson et Ekvall, 1997), les vitesses de croissance n'étaient pas différentes pour les petits de sexe femelle dont les mères étaient primipares ou multipares. En revanche, les petits mâles issus de femelles multipares grandissaient plus vite que ceux issus de femelles primipares.

Enfin, l'analyse ajustée a mis en évidence qu'indépendamment du sexe du petit et de la primiparité de la mère, les petits uniques dans leur portée grandissaient plus vite que les petits ayant au moins un frère ou une sœur au cours de leurs 46 premiers jours de vie. On retrouve ces conclusions dans une étude concernant l'effet de la taille de portée sur la croissance postnatale d'un rougeur, le Néotoma du désert (*Neotoma lepida*)(Cameron, 1973). On peut expliquer cela par le fait que, le petit étant tout seul dans sa portée, toutes les ressources maternelles lui étaient destinées *in utero*, pouvant justifier un poids de naissance supérieur, puis par la suite, lui permettant ainsi d'avoir une vitesse de croissance initiale supérieure à celle de petits qui doivent partager les ressources maternelles avec un ou plusieurs frères et sœurs. La même analyse a en revanche mis en évidence que les petits non uniques grandissaient plus rapidement que les petits uniques après 46 jours de vie. Or, d'après la bibliographie, les petits commencent à s'alimenter par eux-mêmes entre 40 et 45 jours de vie (Perret, 1990). Ce passage à une alimentation solide disponible à volonté pourrait tout à fait correspondre à l'inversion de tendance observée après 46 jours de vie : les microcèbes qui n'étaient pas petits uniques pourraient ainsi « compenser » dans la seconde période et rattraper un éventuel « retard » de croissance en première période, lié au partage des ressources maternelles entre plusieurs petits.

Les variables d'intérêt principal et donc les principaux facteurs de confusion potentiels ont été pris en compte lors de cette analyse ajustée. La bonne standardisation des protocoles d'alimentation et de pesée des petits permettent d'analyser les résultats avec confiance par rapport à d'éventuels biais liés aux conditions d'élevage des petits.

Il aurait été intéressant d'évaluer l'impact de la sex ratio des autres petits de la portée sur la croissance du jeune microcèbe selon son sexe. En effet, selon deux précédentes études de l'UMR 7179 du Muséum National d'Histoire Naturelle, la présence de mâles dans la portée encouragerait les comportements maternels dans la première quinzaine de vie (Perret et Colas, 1997; Colas, 1999) . La première de ces deux études tire notamment la conclusion que les femelles issues de portées composées uniquement de femelles ont une croissance pondérale plus lente que les autres microcèbes (les femelles issues de portées mixtes ou les mâles quelle que soit la composition de la portée). Toutefois, une telle analyse s'est avérée impossible dans notre contexte du fait de la construction même de la base de données. En effet, pour être certains de n'avoir aucun problème d'erreur d'identification des petits microcèbes, nous avions sélectionné des animaux uniques dans leur portée, les mâles n'ayant que des sœurs et les femelles n'ayant que des frères.

Cette étude, et plus particulièrement la partie concernant le choix du modèle, a montré qu'il est difficile de calibrer un modèle de croissance défini. Malgré un grand nombre de références bibliographiques prônant la bonne adéquation du modèle de Gompertz à des données reflétant la croissance pondérale de plusieurs mammifères, il semble que dans le cas du microcèbe, ce modèle ne soit pas le plus adapté. Le modèle à deux pentes capable de prendre en compte plusieurs niveaux d'effets aléatoires a offert une meilleure adéquation avec les données disponibles. Les résultats issus de ces analyses seront utilisés pour des analyses de survie réalisées dans le cadre d'une thèse de sciences visant à évaluer les coûts de la rapidité de la croissance sur les paramètres de fertilité future et la survie des microcèbes.

CONCLUSION

La modélisation de la croissance pondérale demande beaucoup de méthode et de rigueur afin de ne pas se perdre dans des analyses inutiles ou risquer d'arriver à des conclusions erronées. Les étapes décrites dans ce manuscrit, et en particulier dans la première partie, sont toutes nécessaires à une analyse rigoureuse des données.

Tout d'abord, il est obligatoire de standardiser au mieux les pesées des animaux afin de ne pas introduire de biais (de mesure, ou de classement) en tout début d'étude. Si les mesures de différentes expositions ont été faites avant l'écriture du protocole de l'étude, les biais (de confusion, et dans une moindre mesure, de sélection) doivent obligatoirement être discutés, et l'on doit essayer de s'en affranchir au maximum, quitte à ne pas inclure certains individus dans les analyses.

Visualiser les données est souvent très utile afin de se faire une idée de leur dispersion et ainsi choisir au mieux le modèle.

Il faut retenir qu'il n'est pas forcément nécessaire d'utiliser des modèles non-linéaires, souvent plus difficiles à comprendre et à manipuler, si l'on sait travailler avec des ruptures de pentes dans un modèle linéaire.

La prise en compte de la non indépendance des données est primordiale. Si elle n'est pas (correctement) effectuée, les résultats peuvent être faussés. Des effets aléatoires doivent donc être spécifiés dans le programme. De même, les données répétées doivent être prises en compte.

L'association entre une variable explicative (sexe, taille de la portée, parité de la mère, durée d'une maladie, …) et la croissance pondérale des jeunes mammifères doit systématiquement être ajustée sur les autres variables explicatives, qui représentent des facteurs de confusion potentiels. C'est la clef permettant une interprétation causale rigoureuse et assurant une confiance dans les résultats énoncés.

Modéliser la croissance avec un modèle adéquat peut permettre d'obtenir une bonne représentation de la dynamique de croissance dans une espèce donnée afin de la comparer à d'autres espèces proches, ou afin de comparer certaines catégories entre elles au sein de cette espèce (mâles *versus* femelles par exemple). Les comparaisons temporelles ou spatiales sont aussi possibles afin de savoir si le gain moyen quotidien pour une espèce donnée en milieu sauvage est le même qu'il y a dix ans, ou s'il est le même à deux endroits du globe entre deux populations différentes.

La modélisation de la croissance pondérale peut également s'avérer particulièrement utile dans le cas de jeunes mammifères sauvages en captivité afin de savoir si l'alimentation est adaptée (ressources de la mère suffisantes si les petits sont élevés par celle-ci, ou alimentation fournie par l'humain dans le cas de petits élevés à la main), en comparant les résultats avec des données bibliographiques déjà publiées.

Enfin, l'étude de la croissance pondérale des jeunes microcèbes est un bon exemple d'utilisation de la modélisation de la croissance dans une étude épidémiologique et biologique beaucoup plus vaste visant à tester les coûts à long terme sur la longévité et les performances de reproduction qui permettrait de savoir si l'hypothèse « early life hypothesis » (selon laquelle plus on grandirait vite, moins on vivrait longtemps) est vérifiée dans le cas d'un petit primate proche de l'humain (Péron, 2010).

BIBLIOGRAPHIE

AKAIKE H. (1974). A new look at the statistical model identification. *IEEE. T. Automat. Contr.,* **19**, n° 6, 716–723.

BEGALL S. (1996) The application of the Gompertz model to describe body growth. *Growth. Develop. Aging.*, **61**, n° 2, 61–67.

BINCZIK G, REINDL NJ. (1987). A neonatal growth model for captive amur tigers. *In :* Tilson R, Seal U, (editors). *Tigers of the world.* Noyes Publications Park Ridge, New Jersey. 167-170

BIRGERSSON B, EKVALL K. (1997). Early growth in male and female fallow deer fawns. *Behav. Ecol.*, **8**, n° 5, 493–499.

BOX GEP, DRAPER NR. (1987). *Empirical Model-Building and Response Surfaces.* Whiley. New York. 424p.

BURNHAM KP, ANDERSON DR. (1998). *Model Selection and Multimodel Inference - A Practical Information-Theoretic Approach.* Springer. 488p.

CAMERON GN. (1973). Effect of Litter Size on Postnatal Growth and Survival in the Desert Woodrat. *J. Mammal.*, **54**, n° 2, 489–493.

CAMPANA S, JONES C. (1992). Analysis of otolith microstructure data. *Can. Spec. Publ. Fish. Aquat. Sci.*, n° 117, 73–100.

CANALE CI, HUCHARD E, PERRET M *et al.* (2012). Reproductive resilience to food shortage in a small heterothermic primate. *PLoS One*, **7**.

CASE TJ. (1978). On the evolution and adaptive significance of postnatal growth rates in the terrestrial vertebrates. *Q. Rev. Biol.,* **53**, n° 3, 243–82.

CHENG J, EDWARDS LJ, MALDONADO-MOLINA MM *et al.* (2010). Real longitudinal data analysis for real people: Building a good enough mixed model. *Stat. Med. ,***29**, n° 4, 504–520.

COCKBURN A, JOHNSON CN. (1988). Patterns of Growth. *In* : Tyndale-Biscoe C.H., Janssens PA (editors). *The Developing Marsupial*, Springer Berlin Heidelberg, 28–40

COLAS S. (1999). Evidence for Sex-Biased Behavioral Maternal Investment in the Gray Mouse Lemur (Microcebus murinus). *Int. J. Primatol.,* **20**, n° 6, 911–926.

DAMMHAHN M, KAPPELER PM. (2008). Comparative Feeding Ecology of Sympatric Microcebus berthae and M. murinus. *Int. J. Primatol.*, **29**, n° 6, 1567–1589.

DESQUILBET L. (2012). Initiation à SAS® PC. *In : Outils de biostatistique et d'épidémiologie analytique, Ecole Nationale Vétérinaire d'Alfort.*
[http://eve.vet-alfort.fr/pluginfile.php/12577/mod_resource/content/0/Tuto_pdf/Loic_D/ Tuto_SAS _V1_1_ dpp_.pdf] (consulté le 03/01/2014)

DESQUILBET L. (2013). Comment établir une relation de cause à effet en médecine - Application à la médecine vétérinaire. *In : Outils de biostatistique et d'épidémiologie analytique, Ecole Nationale Vétérinaire d'Alfort.* [file:///C:/Users/Laurie/Downloads/Relation_cause_effet_v2.9_CA%20(7).pdf] (consulté le 15/10/2014)

DIGGLE P. (1994). *Analysis of longitudinal data.* 2nd ed., Clarendon Press. Oxford University Press, Oxford : New York, 15-16

DOREAU M, BOULOT S, MARTIN-ROSSET W. (1991). Effect of parity and physiological state on intake, milk production and blood parameters in lactating mares differing in body size. *Cambridge Journals Online.* **53**, n° 01, 111–118.

EBERLE M, KAPPELER PM. (2004). Selected polyandry: female choice and inter-sexual conflict in a small nocturnal solitary primate (Microcebus murinus). *Behav. Ecol. Sociobiol.*, **57**, n° 1, 91–100.

EDWARDS MS, HAWES J. (1997). An overview of small felid hand-rearing techniques and a case study for Mexican margay Leopardus wiedii glaucula at the Zoological Society of San Diego. *In : International Zoo Yearbook*, **35**, n° 1, 90–94.

FAZIO JM. (2010). Behavioral assessment of the clouded leopard (*Neofelis nebulosa*): a comparative analysis of reproductive success. *In : dissertation.* Fairfax,VA: George Mason University.

FLETCHALL N. (2000). Clouded leopard (Neofelis nebulosa) Husbandry Guidelines. Grand Rapids, Michigan: John Ball Zoological Garden. 32-40

GUIGNARD A. (2009). Effets de la restriction calorique chronique, ou d'un mimétique, sur les performances cognitives chez un primate non humain, microcebus murinus. Thèse Méd. Vét. Alfort.

HELMINK SK, SHANKS RD, LEIGHTON EA. (2000). Breed and sex differences in growth curves for two breeds of dog guides. *J Anim. Sci.*, **78**, n° 1, 27–32.

HOWARD JG, PELICAN K, SCHWARTZ R *et al.* (2007). Developing an integrated in situ/ex situ clouded leopard conservation program in Thailand. *In: 2007 Proceedings AAZV, AAWV, AZA/NAG Joint Conference.*

HUCHARD E, CANALE CI, LE GROS C *et al*. (2012). Convenience polyandry or convenience polygyny? Costly sex under female control in a promiscuous primate. *Pro. Biol. Sci.*, **279**, n° 1732.

KAUFMANN KW. (1981).Fitting and using growth curves. *Oecologia*, **49**, n° 3, 293–299.

LAMBE NR, NAVAJAS EA, SIMM G *et al.* (2006). A genetic investigation of various growth models to describe growth of lambs of two contrasting breeds. *J. Anim. Sci.*, **84**, n° 10, 2642–2654.

MCGOVERN IM. (1998). Bringing out the reclusive leopard: hand-rearing clouded leopards (*Neofelis nebulosa)*. Animal Keepers Forum: 25 (7): 275-279.

NÁJERA F, TONGTHAINAN D, ASAITHAMMAKUL W *et al.* (2011). An overview of breeding clouded leopards (Neofelis nebulosa) in the Thailand Clouded Leopard Consortium (Khao Kheow Open Zoo). *In: 2011 Proceedings of the 19th SEAZA Annual Conference. Thailand*.

NÁJERA F, BROWN J, WILDT D *et al.* (in press). Body mass dynamics in hand-reared clouded leopards (*Neofelis nebulosa*) cubs from birth to weaning. *Zoo Biol.*

NOWELL K. (2007). Asian big cat conservation and trade control in selected range States: evaluating implementation and effectiveness of CITES recommendations. *Cambridge: TRAFFIC International*.

PÉROCHON L, COULON JB, LESCOURRET F. (1996). Modelling lactation curves of dairy cows with emphasis on individual variability. *Anim. Sci.*, **63**, n° 02, 189–200.

PERON G, GIMENEZ O, CHARMANTIER A, *et al.* (2010). Age at the onset of senescence in birds and mammals is predicted by early-life performance. *Proc. R. Soc.*, **277**, 2849-2856.

PERRET M. (1990). Influence of social factors on sex ratio at birth, maternal investment and young survival in a prosimian primate. *Behav. Ecol. Sociobiol.*, **27**, n° 6, 447–454.

PERRET M. (1997). Change in Photoperiodic Cycle Affects Life Span in a Prosimian Primate (*Microcebus murinus*). *J. Biol. Rhythms*, **12**, n° 2, 136–145.

PERRET M, COLAS S. (1997). Manipulation of sex ratio at birth and maternal investment in female mouse lemurs (Microcebus murinus, Primates). *Appl. Anim. Behav. Sci.*, **51**, n° 3–4, 275–283.

PHILLIPS CJC, RIND MI. (2001). The Effects on Production and Behavior of Mixing Uniparous and Multiparous Cows. *J. Dairy Sci.*, **84**, n° 11, 2424–2429.

RIVAS A, MARTINEZ F, SANCHEZ I *et al.* (2009). Hand-rearing of Iberian lynx cubs. *In:* Vargas A, Breitenmoser C, Breitenmoser U (editors). *Iberian Lynx ex situ conservation: An interdisciplinary approach.* Madrid: Fundación Biodiversidad. 108-126.

SANDERSON ., KHAN J, GRASSMAN L *et al.* (2008). *The IUCN Red List of Threatened Species.* [en ligne] (mise à jour 2014). [http://www.iucnredlist.org/details/14519/0]. (consulté le 10/11/2014)

SIROT C. (2012). Analyse comparative d'otholites de poissons amphidromes de milieu tropical insulaire: recherche d'indicateurs de qualité des habitats traversés. Thèse Méd. Vét. Alfort.

STRATHE AB, DANFÆR A, SØRENSEN H *et al.* (2010). A multilevel nonlinear mixed-effects approach to model growth in pigs. *J. Anim. Sci.,* **88**, n° 2, 638–649.

SUNQUIST ME. (2002). *Wild cats of the world*. University of Chicago Press, Chicago. 452 p.

WACK RF, KRAMER LW, CUPPS W *et al.* (1991). Growth rate of 21 captive-born,mother-raised cheetah cubs. *Zoo Biol.,* **10**, 273-276.

WANG Z, NOVAK MA. (1994). Parental Care and Litter Development in Primiparous and Multiparous Prairie Voles (*Microtus ochrogaster*). *J. Mammal.,* **75**, n° 1, 18.

WILSON DE, MITTERMEIER RA. (2009). *Handbook of the mammals of the world*. Lynx Editions, Conservation International, IUCN, Barcelona.

WILTING A, BUCKLEY-BEASON VA, FELDHAAR H *et al.* (2007). Clouded leopard phylogeny revisited: support for species recognition and population division between Borneo and Sumatra. *Front. Zool.,* **4**, n° 1, 15.

WRIGHT PC. (1999). Lemur traits and Madagascar ecology: coping with an island environment. *Am. J. Phys. Anthropol.,* **29**, 31–72.

ZULLINGER EM, RICKLEFS RE, REDFORD KH *et al.* (1984). Fitting Sigmoidal Equations to Mammalian Growth Curves. *J. Mammal.,* **65**, n° 4, 607.

ANNEXES

ANNEXE 1: ANALYSE DE L'EFFET BRUT DU SEXE DES JEUNES PANTHÈRES SUR LEUR CROISSANCE PONDÉRALE

CRÉATION D'UN NOUVEAU TABLEAU DE DONNÉES SUR LEQUEL EFFECTUER LES ANALYSES

```
data weights_all_healthy;
```
/* après « data », on définit le nom du nouveau tableau de données que l'on va créer */

```
set cl.cl_data;
```
/* après « set », on nomme le tableau de données à partir duquel on souhaite travailler */

```
if GI_problems = 1 then delete;
```
/* cette ligne permet de ne sélectionner que certaines données du tableau initial : ici, on ne garde que les animaux sains en supprimant les petits qui ont été malades */

```
inter age sex = age*sex;
age_etoile = max(0,age-28);
inter_age_etoile_sex = sex*age_etoile;
```
/* ces lignes permettent de créer les variables pour l'analyse, notamment les termes d'interaction */

```
run;
```
/* la commande « run » est l'instruction qui clot cette étape de création d'un nouveau tableau de données */

PROCÉDURE « PROC MIXED » POUR L'ANALYSE DU MODÈLE Á DEUX PENTES AVEC LA VARIABLE « SEXE »

```
proc mixed data = weights_all_healthy;
```
/* on définit sur quel fichier on souhaite mener l'analyse */

```
class name;
```
/* il faut lister après « class » les variables correspondant aux niveaux de hiérarchie souhaités dans les analyses */

```
model weight = age sex inter_age_sex age_etoile inter_age_etoile_sex/ solution covb;
```
/* on définit ensuite le modèle (ici il suffit de nommer les termes. On indique au logiciel que l'on souhaite qu'il calcule les coefficients ("a", "b", "c", "d", "e", et "f") avec la commande « solution » et que l'on souhaite pouvoir visualiser la matrice de variance covariance des coefficients à l'aide de la commande « covb » */

```
random intercept age age_etoile/ subject = name type = VC solution;
```
/* la commande « random » permet de définir les effets aléatoires : on indique donc après « subject » que les effets aléatoires portent sur le niveau individuel (correspondant ici au nom de l'animal « name ») et que l'on souhaite les faire porter sur les 3 paramètres du modèle : l'ordonnée à l'origine (« intercept »), et les deux pentes (« age » et « age_etoile ») */.

```
repeated / subject = name Type = AR(1);
```
/* cette ligne permet de spécifier que l'on souhaite prendre en compte les données répétées sur chaque panthère dans une matrice carrée symétrique des résidus */

```
run;
```
/* la commande « run » est l'instruction qui clot cette procédure */

Fit Statistics

-2 Res Log Likelihood	24280.9
AIC (smaller is better)	24288.9
AICC (smaller is better)	24288.9
BIC (smaller is better)	24294.7

Lecture de l'indice d'Akaike

Solution for Fixed Effects

Lecture des solutions pour les coefficients « a », « b », « c », « d », « e », « f »

	Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
« a » →	Intercept	218.77	33.7974	29	6.47	<.0001
« b » →	Age	28.0217	2.2266	29	12.59	<.0001
« c » →	Sex	16.9251	41.0641	2295	0.41	0.6803
« d » →	inter_age_sex	-2.1584	2.7056	2295	-0.80	0.4251
« e » →	age_etoile	2.7042	2.5027	29	1.08	0.2888
« f » →	inter_ae_sex	9.7147	3.0485	2295	3.19	0.0015

Covariance Matrix for Fixed Effects

Lecture de la matrice variance-covariance entre les différents paramètres

Row	Effect	Col1	Col2	Col3	Col4	Col5	Col6
1	Intercept	1142.26	-35.5276	-1142.26	35.5276	32.7574	-32.7574
2	Age	-35.5276	4.9575	35.5276	-4.9575	-2.8427	2.8427
3	Sex	-1142.26	35.5276	1686.26	-52.4338	-32.7574	48.3034
4	inter_age_sex	35.5276	-4.9575	-52.4338	7.3205	2.8427	-4.2067
5	age_etoile	32.7574	-2.8427	-32.7574	2.8427	6.2635	-6.2635
6	inter_ae_sex	-32.7574	2.8427	48.3034	-4.2067	-6.2635	9.2931

ANNEXE 3 : NOTIONS DE VARIANCE, COVARIANCE ET DÉTERMINATION DE LA SIGNIFICATIVITÉ D'UNE SOMME

La **variance** permet de caractériser la dispersion des mesures d'un échantillon autour de la moyenne. Une variance de zéro signifie que toutes les valeurs mesurées sont identiques. Une petite variance indique que les valeurs sont proches les unes des autres alors qu'une variance élevée indique que la dispersion des mesures est importante.

Pour calculer la variance, on calcule la moyenne des carrés des écarts entre la série de mesures et la moyenne. La formule de la variance de la variable X de moyenne m prenant les valeurs x_1, x_2, ..., x_n est la suivante :

$$V(X) = \frac{1}{n}\sum_{i=1}^{n}(x_i - m)^2.$$

La racine carrée de la variance s'appelle l'écart type : $SE = \sigma = \sqrt{V(X)}$

Afin de statuer sur la significativité d'une somme, il faut calculer la variance de cette somme, qui s'écrit (dans le cas où il n'y a que deux variables) :

$$V(X + Y) = V(X) + V(Y) + 2\,cov(X\,;\,Y)$$

Intuitivement, la covariance caractérise les variations simultanées pour un même individu de deux variables aléatoires : elle est positive lorsque les écarts entre les variables et leurs moyennes sont de même signe, négative dans le cas contraire. Lorsque les variables sont indépendantes, leur covariance est nulle.

La matrice de variance-covariance d'un vecteur **X =** (X_1, X_2, …, X_p) de p variables aléatoires (dont chacune possède une variance) est la matrice carrée des termes génériques $a_{i,j}$ telle que

$$a_{i,j} = \mathrm{Cov}\,(X_i, X_j)$$

En développant les termes :

$$\mathrm{Var}(\vec{X}) = \begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_p) \\ \mathrm{Cov}(X_2, X_1) & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_p, X_1) & \cdots & \cdots & \mathrm{Var}(X_p) \end{pmatrix} = \begin{pmatrix} \sigma^2_{x_1} & \sigma_{x_1 x_2} & \cdots & \sigma_{x_1 x_p} \\ \sigma_{x_2 x_1} & \ddots & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{x_p x_1} & \cdots & \cdots & \sigma^2_{x_p} \end{pmatrix}$$

C'est cette matrice qui est fournie par le logiciel (cf. Annexe 2).

Le calcul de la variance de la somme de deux variables aléatoires à partir des valeurs des variances et covariances contenues dans la matrice permet ensuite de trouver l'écart type SE (en en prenant la racine carrée) puis l'intervalle de confiance à 95 % calculé tel que IC 95 % = ± 1.96*SE. Si l'intervalle de confiance de la somme contient la valeur « 0 », alors l'estimation « X+Y » n'est pas significativement différente de zéro. Dans le cas où l'intervalle de confiance à 95 % ne contient pas la valeur zéro, on peut conclure à la significativité de la somme « X+Y ».

ANNEXE 4 : ANALYSE DE L'EFFET BRUT DE LA DURÉE DE LA MALADIE DES JEUNES PANTHÈRES SUR LEUR CROISSANCE PONDÉRALE

CRÉATION D'UN NOUVEAU TABLEAU DE DONNÉES SUR LEQUEL EFFECTUER LES ANALYSES

```
data duree_maladie;
set cl.cl_data;
if length_sick < 8 then long = 0;
if length_sick > 7 then long = 1;
inter_age_long = age*long;
age_etoile = max(0,age-28);
inter_age_etoile_long = long*age_etoile;
run;
```

/* définition de la variable « long » utilisée dans l'analyse */

PROCÉDURE « PROC MIXED » POUR L'ANALYSE DU MODÈLE Á DEUX PENTES AVEC LA VARIABLE « LONG »

```
proc mixed data = duree_maladie;
class name;
model weight = age long inter_age_long age_etoile inter_age_etoile_long/ solution covb;
random intercept age age_etoile/ subject = name type = VC solution;
repeated / subject = name Type = AR(1);
run;
```

ANNEXE 5 : ANALYSE AJUSTÉE SUR LES DEUX VARIABLES « SEXE » ET « LONG » DE LA CROISSANCE PONDÉRALE DES JEUNES PANTHÈRES

CRÉATION D'UN NOUVEAU TABLEAU DE DONNÉES SUR LEQUEL EFFECTUER LES ANALYSES

```
data ajust;
set cl.cl_data;
if length_sick < 8 then long = 0;
if length_sick > 7 then long = 1;
inter_age_sex = age*sex;
inter_age_long = age*long;
age_etoile = max(0,age-28);
inter_age_etoile_sex = sex*age_etoile;
inter_age_etoile_long = long*age_etoile;
run;
```

PROCÉDURE « PROC MIXED » POUR L'ANALYSE DU MODÈLE AJUSTÉ Á DEUX PENTES

```
proc mixed data = ajust;
class name;
model weight = age long sex inter_age_sex inter_age_long age_etoile inter_age_etoile_sex inter_age_etoile_long/ solution
covb;
random intercept age age_etoile/ subject = name type = VC solution;
repeated / subject = name Type = AR(1);
run;
```

ANNEXE 6 : PROGRAMMES UTILISÉS POUR LA CRÉATION DU TABLEAU DE DONÉEES ET LA REPRÉSENTATION EN NUAGE DE POINTS

```
proc sort data = microcebe.final out = final;
by nom agepesee;
run;  /* On ordonne le fichier de données « final » du répertoire « microcebe »par nom d'animal et par age*/
```

```
data fusion_certaines;
merge final (in=fichier_info) microceb.donnees_certaines (in=fichier_microcebe);
by nom agepesee;
if fichier_info = 1 and fichier_microcebe = 1 then output;
run;  /* On fusionne deux tables de données avec la commande « merge », en supprimant les doublons */
```

```
data microceb.pour_nuage_de_pts;
set fusion_certaines;
by nom;
retain compteur 0;
if first.nom = 1  then compteur = compteur+1;
if compteur > 151 then delete;
run;  /* On sélectionne les 150 premiers individus de la table de données (afin de tracer le graphique en nuage de points) */
```

ANNEXE 7 : MACRO PERMETTANT DE TESTER LES DIFFERENTS AGES DE RUPTURE DE PENTE

Création d'une macro pour obtenir les résultats pour plusieurs âges de rupture de pente (boucle qui tourne de « âge mini » à « âge maxi », spécifiés par l'opérateur). Il suffit ensuite de regarder l'indice d'Akaike pour chacun des âges de rupture testés.

```
%macro rupture(age_mini=,age_maxi=);

%do age_rupt = &age_mini. %to &age_maxi.;


data pour_proc;
set fusion_certaines;
agepesee_etoile = max(0,agepesee-&age_rupt.);
run;

title "on en est à un age de rupture égal à &age_rupt.";


proc glimmix data = pour_proc;
class nom mere portee;
model poids = agepesee agepesee_etoile / solution;
random intercept agepesee /G type = un(2) subject = mere;
random intercept agepesee /G type = un(2) subject = portee(mere);
random agepesee agepesee_etoile /G  type = un(2) subject = nom(portee mere);
run;

%end;

%mend;


%rupture (age_mini=20,age_maxi=80); /* Appel de la macro, avec spécification des âges */
```

ANNEXE 8 : EXEMPLE DE PROGRAMME PERMETTANT DE CONNAITRE LES POURCENTAGES DE MICROCÈBES DANS LES DIFFÉRENTES CATÉGORIES

```
data pour_pourcentages;
set fusion_certaines;
by nom;
if first.nom;
run;  /* Sélection de la première mesure de chaque individu, et obtention d'une table dans laquelle chaque individu n'est présent qu'une fois */


proc freq data = pour_pourcentages;
table male nb_petits_eleves nombre_accouch_precedents;
run;  /* La commande « table » donne les pourcentages dans chacune des catégories spécifiées ensuite, sous forme de tableau */
```

ANNEXE 9 : PROGRAMMES PERMETTANT L'ANALYSE BRUTE DE L'EFFET DU NOMBRE D'AUTRES PETITS SUR LA CROISSANCE PONDÉRALE DES MICROCÈBES

Création des variables utiles à l'analyse

```
data pour_effet_nap_binaire;
set fusion_certaines;
agepesee_etoile = max(0,agepesee-46);
nb_autres_petits = nb_petits_eleves - 1;
if nb_autres_petits = 0 then petit_unique = 1;
if nb_autres_petits > 0 then petit_unique = 0;
inter_ap_pu = petit_unique*agepesee;
inter_ape_pu = petit_unique*agepesee_etoile;
run;
```

Analyse à l'aide de la procédure glimmix pour la variable binaire « petit_unique » (modèle linéaire à une rupture de pente à 46 jours)

```
proc glimmix data = pour_effet_nap_binaire;
class nom mere portee;
model poids = agepesee agepesee_etoile petit_unique inter_ap_pu inter_ape_pu / solution covb;
random intercept agepesee /G type = un(2) subject = mere; /* effet aléatoire maternel */
random intercept agepesee /G type = un(2) subject = portee(mere); /* effet aléatoire portée*/
random agepesee agepesee_etoile /G  type = un(2) subject = nom(portee mere); /* effet aléatoire individuel */
random _residual_ /subject = nom type =cs; /* prise en compte des données répétées dans la procédure glimmix, syntaxe particulière*/
run;
```

Analyse à l'aide de la macro %nlinmix pour la variable binaire « petit_unique » (modèle non linéaire de type Gompertz)

```
%nlinmix ( data = pour_effet_nap_binaire,
model= %str (

wf = wfa + effet_aleat_wfa;
k = ka + effet_aleat_ka;

predv=(w0*(1+alpha*petit_unique))*EXP((1-EXP(-(k*(1+beta*petit_unique))*agepesee))*LOG((wf*(1+delta*petit_unique))
/(w0*(1+alpha*petit_unique)))); /* équation du modèle */
),

parms = %str (
w0 = 7
wfa = 59
ka = 0.04
alpha = 0.01
beta = 0.01
delta = 0.01
),

stmts = %str (
class nom;
model pseudo_poids = d_w0 d_wfa d_ka d_alpha d_beta d_delta / noint notest solution cl;
random d_effet_aleat_wfa d_effet_aleat_ka / subject = nom type = VC solution; /* effet aléatoire individuel */
Repeated / subject = nom Type = vc; /* prise en compte des données répétées */
),
expand=eblup
);
```

Création des variables utiles à l'analyse

```
data pour_effet_nb_autres_petits;
set fusion_certaines;
agepesee_etoile = max(0,agepesee-46);
nb_autres_petits = nb_petits_eleves - 1;
inter_ap_naupt = nb_autres_petits*agepesee;
inter_ape_naupt = nb_autres_petits*agepesee_etoile;
run;
```

Analyse à l'aide de la procédure glimmix pour la variable quantitative « nombre_autres_petits » (modèle linéaire à une rupture de pente à 46 jours)

```
proc glimmix data = pour_effet_nb_autres_petits;
class nom mere portee;
model poids = agepesee agepesee_etoile nb_autres_petits inter_ap_naupt inter_ape_naupt/ solution covb;
random intercept agepesee /G type = un(2) subject = mere;
random intercept agepesee /G type = un(2) subject = portee(mere);
random agepesee agepesee_etoile / G  type = un(2) subject = nom(portee mere);
random _residual_ /subject = nom type =cs;
run;
```

Création de classes pour la vérification de la linéarité

```
data pour_effet_nap_classes;

set fusion_certaines;
agepesee_etoile = max(0,agepesee-46);
nb_autres_petits = nb_petits_eleves - 1;
inter_ap_naupt = nb_autres_petits*agepesee;
inter_ape_naupt = nb_autres_petits*agepesee_etoile;

if nb_autres_petits = 0 then classe0 = 1;
if nb_autres_petits  = 0 then classe1 = 0;
if nb_autres_petits  = 0 then classe2 = 0;

if nb_autres_petits  = 1 then classe1 = 1;
if nb_autres_petits  = 1 then classe2 = 0;
if nb_autres_petits  = 1 then classe0 = 0;

if nb_autres_petits  >1 then classe1 = 0;
if nb_autres_petits  >1 then classe2 = 1;
if nb_autres_petits  >1 then classe0 = 0;

agepesee_cl1 = agepesee*classe1;
agepesee_cl2 = agepesee*classe2;

ape_cl1 = agepesee_etoile*classe1;
ape_cl2 = agepesee_etoile*classe2;

run;
```

Analyse à l'aide de la procédure glimmix pour la variable représentant le nombre d'autres petits en classes

```
proc glimmix data = pour_effet_nap_classes;
class nom mere portee;
model poids = agepesee agepesee_etoile classe1 classe2 agepesee_cl1 agepesee_cl2 ape_cl1 ape_cl2/ solution;
random intercept agepesee / G type = un(2) subject = mere;
random intercept agepesee / G type = un(2) subject = portee(mere);
random agepesee agepesee_etoile / G  type = un(2) subject = nom(portee mere);
random _residual_ /subject = nom type =cs;
run;
```

ANNEXE 10 : PROGRAMME PERMETTANT L'ANALYSE BRUTE DE L'EFFET DU SEXE DU JEUNE MICROCÈBE SUR SA CROISSANCE PONDÉRALE

Exemple de sélection de données pour une analyse ponctuelle au premier jour de vie chez les femelles

```
data analyse_ponctuelle_femelles;
set fusion_certaines;
if agepesee=0;
if male = 0;
run;
```

Calcul de la moyenne de poids chez les femelles au premier jour de vie

```
proc means data = analyse_ponctuelle_femelles;
class agepesee;
var poids;
run;
```

Création des variables utiles à l'analyse

```
data pour_effet_sexe;
set fusion_certaines;
agepesee_etoile = max(0,agepesee-46);
inter_ap_sexe = agepesee*male;
inter_ape_sexe = agepesee_etoile*male;
run;
```

Analyse à l'aide de la procédure glimmix pour la variable sexe (modèle linéaire à une rupture de pente à 46 jours)

```
proc glimmix data = pour_effet_sexe;
class nom mere portee;
model poids = agepesee agepesee_etoile male inter_ap_sexe inter_ape_sexe/ solution covb;
random intercept agepesee /G type = un(2) subject = mere; /* Effet aléatoire niveau maternel*/
random intercept agepesee /G type = un(2) subject = portee(mere); /* Effet aléatoire niveau portée*/
random agepesee agepesee_etoile/G type = un(2) subject = nom(portee mere); /* Effet aléatoire niveau individuel*/
random _residual_ /subject = nom type =cs; /* Prise en compte des données répétées dans la procédure glimmix, syntaxe particulière*/
run;
```

ANNEXE 11 : PROGRAMME PERMETTANT L'ANALYSE BRUTE DE L'EFFET DE LA PARITÉ (PRIMIPARE *VS* MULTIPARE) DE LA MÈRE DU JEUNE MICROCÈBE SUR LA CROISSANCE PONDÉRALE DE CE DERNIER

Création des variables utiles à l'analyse

```
data pour_effet_parite_binaire;
set fusion_certaines;
agepesee_etoile = max(0,agepesee-46);
if nombre_accouch_precedents = 0 then primipare =1;
if nombre_accouch_precedents >0 then multipare=1 ;

if nombre_accouch_precedents =0 then multipare = 0;
if nombre_accouch_precedents >0 then primipare=0;

inter_ap_multipare = multipare*agepesee;
inter_ape_multipare = multipare*agepesee_etoile;
run;
```

Analyse à l'aide de la procédure glimmix pour la variable binaire « multipare » (modèle linéaire à une rupture de pente à 46 jours)

```
proc glimmix data = pour_effet_parite_binaire;
class nom mere portee;
model poids = agepesee agepesee_etoile multipare inter_ap_multipare inter_ape_multipare/ solution covb;
random intercept agepesee /G type = un(2) subject = mere;
random intercept agepesee /G type = un(2) subject = portee(mere);
random agepesee agepesee_etoile / G  type = un(2) subject = nom(portee mere);
run;
```

ANNEXE 12 : PROGRAMME PERMETTANT L'ANALYSE AJUSTÉE DE LA CROISSANCE PONDÉRALE DES JEUNES MICROCÈBES

Création des variables utiles à l'analyse

```
data pour_ajust;
set fusion_certaines;
agepesee_etoile = max(0,agepesee-46);

nb_autres_petits = nb_petits_eleves - 1;
if nb_autres_petits = 0 then petit_unique = 1;
if nb_autres_petits > 0 then petit_unique = 0;
inter_ap_pu = petit_unique*agepesee;
inter_ape_pu = petit_unique*agepesee_etoile;

inter_ap_sexe = agepesee*male;
inter_ape_sexe = agepesee_etoile*male;

if nombre_accouch_precedents = 0 then primipare =1;
if nombre_accouch_precedents >0 then multipare=1 ;
if nombre_accouch_precedents =0 then multipare = 0;
if nombre_accouch_precedents >0 then primipare=0;
inter_ap_multipare = multipare*agepesee;
inter_ape_multipare = multipare*agepesee_etoile;

run;
```

Analyse ajustée à l'aide de la procédure glimmix

```
proc glimmix data = pour_ajust;
class nom mere portee;
model poids = agepesee agepesee_etoile male petit_unique multipare inter_ap_sexe inter_ape_sexe inter_ap_pu inter_ape_pu
inter_ap_multipare inter_ape_multipare/ solution covb;
random intercept agepesee / G type = un(2) subject = mere;
random intercept agepesee / G type = un(2) subject = portee(mere);
random agepesee agepesee_etoile / G  type = un(2) subject = nom(portee mere);
run;
```

GUIDE PRATIQUE DE MODÉLISATION DE LA CROISSANCE PONDÉRALE DE JEUNES MAMMIFÈRES SAUVAGES À TRAVERS DEUX EXEMPLES

Nom et prénom : VIROLLE Laurie

RÉSUMÉ

L'objectif de ce guide pratique est de fournir une démarche étape par étape permettant de modéliser la croissance pondérale de jeunes mammifères sauvages.

La première partie de cette thèse présente la théorie et les étapes de construction d'un modèle adapté à la croissance pondérale de jeunes panthères nébuleuses (*Neofelis nebulosa*) dans leurs 90 premiers jours de vie. La démarche pour construire une base de données et trouver un modèle en adéquation avec les données de poids recueillies est détaillée, en s'appuyant pas à pas sur un exemple concret.

La seconde partie de ce travail est un retour d'expérience sur l'analyse de la croissance pondérale de jeunes microcèbes murins (*Microcebus murinus*) à l'aide de deux modèles. La prise en compte de la non-indépendance des données à plusieurs niveaux est discutée, ainsi que le choix d'un modèle linéaire à effets mixtes à une rupture de pente plutôt qu'un modèle non linéaire plus complexe. Les associations brutes et ajustées entre la croissance pondérale et certaines caractéristiques des jeunes microcèbes (leur sexe, la parité de leur mère et le nombre d'autres petits dans leur portée) sont analysées et interprétées.

Ce manuscrit montre que l'utilisation de modèles à effets mixtes autorise la prise en compte de la non-indépendance des données, et permet de choisir le type de modélisation le plus adapté (linéaire ou non linéaire, avec ou sans rupture(s) de pente(s)) pour décrire la croissance pondérale de jeunes mammifères sauvages dans leurs premiers jours de vie.

Mots-clés : MODÉLISATION, GUIDE, CROISSANCE, POIDS CORPOREL, MAMMIFÈRE SAUVAGE, *NEOFELIS NEBULOSA*, *MICROCEBUS MURINUS*, MODÈLES À EFFETS MIXTES.

Jury

Président du jury :
Directeur : Dr. Loïc DESQUILBET
Assesseur : Pr. Valérie CHETBOUL

TUTORIAL IN MODELLING BODY MASS DYNAMICS OF YOUNG WILD MAMMALS: APPLICATION TO TWO EXAMPLES

Name : VIROLLE Laurie

SUMMARY

The aim of this tutorial is to provide a step by step approach to build and use models characterizing the body mass dynamics in young wild mammals.

The first part of this manuscript presents the theory and the construction steps to model the body mass dynamics of young clouded leopards (*Neofelis nebulosa*) during their first 90 days of life. This part explains in details the approach to build a database and find a model fitting with the collected weights, while referring to a practical example.

The second part of this manuscript comments the analysis of the body mass dynamics of young gray mouse lemurs (*Microcebus murinus*) with two different models. The non-independence of the data on several levels and the choice of a linear mixed-effect model with one change in slope instead of a more complicated nonlinear model are discussed. The unadjusted and adjusted associations between the body mass dynamics and some individual characteristics of the young gray mouse lemurs (such as their sex, their mother's parity and the number of other offsprings in their litter) are analyzed and interpreted.

This thesis shows that using mixed-effect models allows to take into account the non-independence of the data and to choose which type of regression models (linear or nonlinear, with or without change(s) in slope(s)) is the best to describe the body mass dynamics of young wild mammals in their first days of life.

Key words : MODELLING, TUTORIAL, GROWTH, BODY MASS, WILD MAMMALS, *NEOFELIS NEBULOSA*, *MICROCEBUS MURINUS*, MIXED-EFFECT MODELS.

Jury

President :
Director : Dr. Loïc DESQUILBET
Assessor : Pr. Valérie CHETBOUL