

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

Ministère de l'enseignement supérieur et de la recherche scientifique



Institut des Sciences
Vétérinaires- Blida



Université Saad
Dahlab-Blida 1-

Projet de fin d'études en vue de l'obtention du
Diplôme de Docteur Vétérinaire

**Application des méthodes épidémiologiques sur des
données réelles (non virtuelles ou fictives)**

Application of epidemiological methods on real data

Présenté par :

- ABDELTIF Abderrahmane
- TARBIT Youcef et ABDELAZIZ Oussama

Soutenu le & Juin 2020

Devant le jury :

Président(e) :	ADEL Amel	MCB	ISV-Blida1
Examineur :	MDROUH Bachir	Doctorant	ISV-Blida1
Promoteur :	SAIDANI. Khelaf	MCA	ISV-Blida1

Année : 2019/2020

Dédicaces collectives

ABDELTIF Abderrahmane

Je dédie ce travail et évènement marquant de ma vie à mes parents pour leur soutien incomparable, à mon petit frère ABDELTIF Riad.

Aux belles rencontres faites durant mon cursus de cinq inoubliables années, parmi elles : AGRANE Fares, OUKARA Mohamed, DRIASSA Roumaissa, BAGHDADI Chawki, ALLAF Azouz, CHETI Iyes , ABDI Nacer edine.

MEDJADJI Ilhem qui par marque de grande gentillesse nous a aidé sur un nombre important de prélèvements.

TARBIT Youcef

Je dédie ce modeste travail à mes parents pour leur soutien et précieux aide de mon premier jour en tant qu'étudiant jusqu'au moment où je rédige ces mots.

ABDELAZIZ Oussama

Je dédie ce mémoire de fin d'étude à mes parents, amis, ainsi qu'à mes binômes avec qui nous avons travaillé dans la joie et la bonne humeur.

Remerciements

Nous tenons à témoigner notre reconnaissance à toutes les personnes qui ont contribué au succès de notre stage et qui nous ont aidés lors de la rédaction de ce mémoire.

Dans un premier temps, nous exprimons notre gratitude à notre promoteur de mémoire SAIDANI Khelaf, maitre de conférences (A) de Parasitologie à l'université de Saad Dahleb Institut sciences vétérinaire, pour sa patience, sa disponibilité et surtout ses judicieux conseils, qui ont contribué à alimenter notre réflexion.

Nous remercions également les vétérinaires praticiens privés qui ont grandement participé à notre formation, et qui nous ont considérés comme des membres de leurs familles, parmi eux Dr.TOUDJINE Malik, Dr. CHARIF Toufik, Dr. MOUSSOUNI Naceredine.

Nous remercions aussi chaleureusement la directrice ainsi que tous les enseignants de l'institut des sciences vétérinaires de Blida.

Nos remerciements les plus sincères vont aux membres du jury, Dr ADEL Amel maitre de conférences à l'institut des sciences vétérinaires en tant que présidente de jury et Dr MEDROUH Bachir doctorant en épidémiologie comme examinateur. Le choix dudit jury est loin d'être arbitraire, mais strictement déterminé par le haut niveau scientifique notamment en biostatistique dont jouissent ces deux honorables enseignants et chercheurs.

Résumé en français

Trois types de fichiers de données réelles obtenues sur le terrain entre 2016 et 2018 nous ont servi à montrer les principaux concepts de base de biostatistique et de l'épidémiologie.

En effet, un fichier portant sur la recherche des parasites digestifs chez 148 enfants à Blida au printemps 2018, et un fichier lié à l'infestation des bovins par les tiques chez 300 bovins des wilayas de Médéa et Bouira durant la période estivale de 2016 et enfin un troisième fichier inhérent à l'épidémiologie des mammites bovines cliniques chez 403 vaches laitières dans la wilaya de Bejaia entre 2017 et 2018 ont servi pour illustrer les principaux concepts de l'épidémiologie et de la statistique.

Ont été développés les tests statistiques de base tels que le test de l'écart réduit et le test de chi-deux ainsi qu'un outil majeur en épidémiologie analytique qui est la régression logistique.

Mots-clé : Epidémiologie, statistique, données réelles, tests, régression logistique.

Abstract

Three types of real data files obtained in the field between 2016 and 2018, were used to show the main basic concepts of biostatistics and epidemiology.

Indeed, a file relating to the search for digestive parasites in 148 children in Blida in the spring of 2018, and a file linked to the infestation of cattle by ticks in 300 cattle from the wilayas of Médéa and Bouira during the summer period of 2016 and Finally, a third file inherent in the epidemiology of clinical bovine mastitis in 403 dairy cows in the wilaya of Bejaia between 2017 and 2018 was used to illustrate the main concepts of epidemiology and statistics.

Basic statistical tests such as the Z test and the chi-square test have been developed, as well as a major tool in analytical epidemiology which is logistic regression.

Keywords: Epidemiology, statistics, real data, tests, logistic regression.

ملخص

تم استخدام ثلاثة أنواع من ملفات البيانات الحقيقية التي تم الحصول عليها ميدانيًا بين عامي 2016 و2018 لإظهار المفاهيم الأساسية الرئيسية للإحصاءات الحيوية وعلم الأوبئة. في الواقع، ملف يتعلق بالبحث عن طفيليات الجهاز الهضمي لدى 148 طفلاً في البلدة في ربيع 2018، و ملف مرتبط بانتشار الأبقار عن طريق القراد في 300 رأس من ولايات المدينة والبويرة خلال فترة الصيف 2016 و وأخيراً، تم استخدام ملف ثالث متأصل في وبائيات التهاب الضرع البقري السريري في 403 بقرة حلوب في ولاية بجاية بين عامي 2017 و 2018 لتوضيح المفاهيم الرئيسية لعلم الأوبئة والإحصاءات. تم تطوير الاختبارات الإحصائية الأساسية مثل اختبار الانحراف المخفض واختبار خي مربع، بالإضافة إلى أداة رئيسية في علم الأوبئة التحليلي وهو الانحدار اللوجستي.

الكلمات المفتاحية: علم الوبائيات، علم الاحصاء، البيانات الحقيقية، الاختبارات الانحدار اللوجستي

Liste des tableaux et des figures

Nature	Titre	Pages
TABLEAU 1	Espèces de tiques identifiées	27
TABLEAU 2	Nombre de cas et pourcentages de portage selon les différents facteurs étudiés	28
TABLEAU 3	Nombre de cas et pourcentages d'atteinte par les oxyures selon les différents facteurs étudiés.	29
FIGURE 1	: Courbe sigmoïde définie par la fonction logistique , d'équation : $f(x) = \frac{\exp(x)}{1 + \exp(x)} = p$	20

Table des Matières

Titre	Pages
Dédicaces collectives	I
Remerciements	II
Résumés	III
Liste des tableaux et figures	V
Table des matières	VI
Introduction générale	1
I. Synthèse Bibliographique	3
1. Concepts de mesures, de variables et d'échelle en épidémiologie	3
1.1. La fonction des variables	3
1.2. Classification des variables en épidémiologie	3
1.2.1. Variables qualitatives	3
1.2.2. variables quantitatives	3
1.3. Notion d'échelles	3
1.4. Description des données	4
1.5. Les trois types de mesures en épidémiologie	5
2. Plans d'échantillonnages	8
2.1. Echantillonnage non probabiliste	8
2.2. Echantillonnage probabiliste	9
3. Estimations des caractéristiques d'intérêt	11
3.1. Estimation d'une proportion	12
3.2. Estimation de la moyenne	12
3.3. Estimation de la variance	13

4.	Tester nos hypothèses	14
4.1.	Définition et situations distinctes	14
4.2.	Tests de comparaisons	14
4.3.	Tests de liaison	15
4.4.	Analyse de variance	15
4.5.	Régression linéaire	16
5.	Etudes épidémiologiques et les biais	16
6.	Régression logistique comme outil majeur en épidémiologie	19
II.	Matériel et Méthodes	23
II.1.	Objectifs	23
II.2.	Les données réelles utilisées	23
II.3.	Les mesures effectuées	23
II.4.	Plan d'échantillonnage et taille d'échantillon	24
II.5.	Estimation de la prévalence des mammites bovines cliniques	25
II.6.	Présentation du logiciel R	25
III.	Résultats et discussion	27
IV.	Conclusion générale et perspectives	34
V.	Références bibliographiques	35

Introduction générale

L'épidémiologie est l'étude de la maladie dans les populations et des facteurs qui déterminent son occurrence ; le mot clé étant donc la population. L'épidémiologie vétérinaire comprend en outre les enquêtes et l'évaluation d'autres événements liés à la santé, notamment la productivité. Toutes ces enquêtes consistent à observer les populations animales et de tirer des conclusions des observations.

Une traduction littérale du terme «épidémiologie» est basé sur ses racines grecques : (epi-) = sur, (demo-) = population, et (logos-) = science, ce qui veut dire dans le langage moderne l'étude des maladies dans la population (Thrusfield, 2007).

Traditionnellement, la médecine vétérinaire s'occupait des maladies des animaux à l'échelle individuelle et l'objectif était de trouver un remède à une maladie chez un animal déjà affecté. Or, l'épidémiologie est l'étude des maladies dans les populations. Il s'agit de déterminer la fréquence de la maladie, sa survenue régie par l'interaction d'un grand nombre de différents facteurs ou déterminants, et de l'action sur ces déterminants pour réduire la fréquence d'apparition de la maladie dans une population. (Pfeiffer, 2002).

Le domaine de l'épidémiologie vétérinaire peut être divisé en différents composants. Une de ses bases essentielles est la collecte de données, qui doivent ensuite être analysées en utilisant des approches qualitatives ou quantitatives afin de formuler des hypothèses concernant les liens de causalité. Les données à analyser sont obtenues à partir d'études sur le terrain (observationnelles ou expérimentales) ou sont issues d'épidémio-surveillance ou de la modélisation, dans ce dernier cas on parle de simulation (Pfeiffer, 1999).

Initialement, l'épidémiologie était centrée sur l'étude des maladies infectieuses et des épidémies (on parle d'épizootie en médecine vétérinaire). Ainsi, progressivement, les objectifs des études épidémiologiques ont évolué. L'épidémiologie peut donc revêtir aujourd'hui plusieurs objectifs et donc types d'études (Thrusfield, 2007) :

- de surveillance
- d'investigation
- de recherche
- d'évaluation

Pour cela, différentes mesures ou indicateurs, ainsi que types d'enquêtes sont utilisés,

Introduction générale

On peut donc distinguer l'épidémiologie descriptive, l'épidémiologie analytique ou étiologique et enfin l'épidémiologie évaluative.

La finalité de l'épidémiologie descriptive est la description d'un phénomène de santé dans une population (fréquence, variations) en fonction des caractéristiques de cette population et de paramètres tels que le temps et l'espace. Elle est nécessaire pour mesurer l'importance d'un problème de santé et soulever des hypothèses étiologiques (facteurs de risque de survenue de cette maladie).

L'épidémiologie analytique a pour but de rechercher les causes des maladies et les facteurs ou marqueurs de risque influençant leurs survenues au sein d'une population. Elle permet ainsi d'établir des relations de causalité entre l'exposition à un facteur ou marqueur de risque et la maladie. Le risque est défini comme la probabilité de survenue d'un événement donné.

Les enquêtes évaluatives étudient les actions de prévention ou de traitement, par exemple évaluer l'impact d'une campagne de vaccination contre la fièvre aphteuse.

Ainsi le présent mémoire se propose d'exposer les différentes démarches suivies par l'épidémiologiste pour réaliser ses objectifs, de la collecte de données à leur analyse et leur interprétation. Ainsi, la première partie traite plusieurs points à savoir les concepts de variables, de mesures et d'échelles en épidémiologie, les méthodes d'échantillonnages aléatoires ou non, les estimations des paramètres (c'est l'inférence), les tests permettant d'éprouver nos hypothèses, les différents types d'études épidémiologiques et enfin la régression logistique avec ses différentes variantes comme outil majeur en épidémiologie, analytique en particulier.

La partie pratique, quant à elle, est destinée à l'application de tous les concepts précédents sur données réelles issues de trois recherches épidémiologiques : un fichier sur les tiques des bovins, un fichier sur les mammites bovines et enfin un fichier sur des parasites des enfants.

1. Concepts de mesures, de variables et d'échelles en épidémiologie

L'épidémiologie est fondée sur la statistique (Chalat, 2015).

1.1. La fonction des variables

Pour étudier la répartition d'une maladie et les différentes circonstances qui entourent son apparition et son développement au sein d'une population, l'épidémiologiste est amené à regarder un certain nombre de variables descriptives, souvent assez fortement reliées au problème considéré (Desquilbet, 2019).

1.2. Classification des variables en épidémiologie

1.2.1. Variables qualitatives

Une variable est qualitative si les valeurs qu'elle prend correspondent à des qualités, des attributs (Charout, 2019). Ainsi en est-il du sexe (male, femelle), des races d'un bovin ou d'un ovin, des groupes sanguins. La variable qualitative est de nature discrète. Lorsqu'elle n'admet que deux modalités, elle est dite binaire ou binomiale.

1.2.2. Variables quantitatives

Une variable est quantitative si les valeurs qu'elle prend sont d'emblée de nature numérique, des quantités. On distingue les variables quantitatives en variables discrètes et en variables continues. Le poids, la taille, le volume, la glycémie et la cholestérolémie sont des variables quantitatives continues. En revanche, le nombre d'enfants dans une famille, le nombre d'animaux dans un élevage sont des variables quantitatives discrètes (Abdi, 1987).

1.3. Notion d'échelles

Pour une variable donnée, l'ensemble des classes (ou catégories) définit ce que l'on appelle une échelle de classification. Une échelle de classification doit permettre de classer toutes les observations, chacune ne pouvant être classée que dans une catégorie. On distingue l'échelle nominale, l'échelle ordinale, l'échelle proportionnelle et l'échelle par intervalle.

1.4. Description des données

1.4.1. Paramètres de tendance centrale

Appelés aussi paramètres de positions, ils comprennent la moyenne, la médiane et le mode (Roger, 2006 ; Chalat, 2015 ; Kévin, 2018).

La moyenne arithmétique est la somme des valeurs de la variable divisée par le nombre total d'individus ou d'unités statistiques :

$$\bar{x} = \frac{\sum_1^n n_i x_i}{N}$$

La médiane est la valeur centrale qui partage l'échantillon en 2 groupes de même effectif : 50 % au-dessus et 50 % en dessous. La médiane peut avoir une valeur différente de la moyenne (Boulares, 2016)

On appelle le mode d'une variable X, la valeur de la variable qui a la plus grande fréquence et on le note Mo(X). Le mode est une importante mesure de tendance centrale pour les variables qualitatives nominales (Alloua, 2019).

1.4.2. Paramètres de dispersion

Les mesures ou paramètres ou encore indicateurs de dispersion comprennent l'étendue, la variance et l'écart type (qui n'est rien d'autre que la racine carrée de la variance), le coefficient de variation.

L'étendue est la mesure de dispersion la plus simple à calculer. Lorsqu'on a une variable quantitative X, mesurée sur un échantillon de taille n. Alors l'étendue est égale à la valeur maximale moins la minimale (Rosner, 2016).

La variance d'une variable mesurée sur un échantillon est égale à la moyenne des carrés des écarts qui séparent chaque observation de la moyenne échantillonnale, son calcul diffère selon la nature des données (Boulares, 2016).

Pour un échantillon de données dont la moyenne est non négative, on définit le coefficient de variation d'une variable X par : (Ecart type/la moyenne) x 100 (Allaoua, 2019).

1.4.3. Paramètres de forme

Le coefficient d'asymétrie (Sk, skewness en anglais) évalue le défaut de symétrie d'une distribution. Il est nul pour une distribution symétrique (par exemple une distribution normale, ou une distribution binomiale avec $p=0,5$). Il est positif pour une distribution "étalée à droite" et négatif pour une distribution, au contraire, étalée à gauche.

Le coefficient d'aplatissement (K, kurtosis en anglais) évalue la dispersion des valeurs "extrêmes" par référence à la loi normale. Il est nul pour une distribution normale, négatif pour une distribution moins "aplatie" qu'une distribution normale et positif sur une distribution plus aplatie qu'une distribution normale (Ricco, 2011).

1.4.4. Intérêt d'étudier la normalité

L'étude de la normalité ou non d'une variable est un prérequis à l'application des tests paramétriques (Champely, 2017). Il y a plusieurs tests pour éprouver l'hypothèse de normalité. Mais, très populaire, le test de Shapiro-Wilk, qui est basé sur la statistique W . est le plus robuste en comparaison des autres tests, il est particulièrement puissant pour les petits effectifs (Ricco, 2011).

1.5. Les trois types de mesures en épidémiologie

1.5.1. Les mesures de fréquences

L'une des tâches les plus fondamentales la recherche épidémiologique est la quantification de l'apparition de la maladie.

D'un point de vue mathématique, la fréquence de l'apparition de la maladie peut être exprimée par des mesures statiques ou dynamiques. Les mesures statiques comprennent les proportions et les ratios. Une proportion est une fraction dans laquelle le numérateur est inclus dans le dénominateur. Elle est sans dimension (unité), allant de 0 à 1 et est souvent exprimé en pourcentage ($\times 100$). Le ratio est une fraction dans laquelle le numérateur n'est pas inclus dans le dénominateur et il peut être avec ou sans dimension. Les mesures dynamique les mesures comprennent des taux qui changent surtout en fonction du temps, c'est une quantité variable en fonction d'une autre quantité (généralement du temps). Ils ne sont pas sans dimension et n'ont pas de limite supérieure finie (Pfeiffer, 2002).

L'incidence doit être distinguée de la prévalence, qui est en fait une mesure statique. L'incidence mesure le nombre de nouveaux cas sur le total d'individus à risque.

1.5.2. Les mesures d'association

Beaucoup d'efforts épidémiologiques visent à tenter de découvrir l'étiologie (c'est-à-dire les causes) de maladies particulières, en vue de la prévention.

En pratique, la plupart des maladies n'ont pas un seul facteur identifiable comme une infection par un micro-organisme spécifique ou une exposition à une toxine particulière ou à un traumatisme physique. En revanche, il semble que la plupart des maladies sont multifactorielles et représentent les effets d'une combinaison de facteurs génétiques, constitutionnels et environnementaux. (Alan et Gary, 2002).

En résumé, les mesures d'association permettent de tester s'il y a une différence de risque entre les individus exposés et non exposés au facteur de risque versus protecteur (Florence, 2016).

a) La cote du risque

Cette mesure est un rapport entre la prévalence de la maladie (ou l'incidence cumulative) dans deux groupes de population. Il s'agit donc d'un rapport de risques et appelé ratio de risque. Une telle mesure peut être dérivée, par exemple, d'une étude transversale.

b) Rapports de côtes

Les risques relatifs dérivés des études transversales et de cohorte sont intuitivement des mesures intéressantes à estimer car elles reflètent ce que l'épidémiologiste souhaite généralement savoir. Quelle est la variation du taux (risque) de maladie associée d'être exposé à un facteur particulier?

Dans une étude cas-témoins, deux groupes d'individus sont sélectionnés : un groupe atteint d'une maladie et un groupe non atteint. Dans de telles circonstances, il ne sera généralement pas possible de déterminer le taux (ou risque) de maladie chez les sujets exposés à un facteur particulier par rapport à ceux qui ne sont pas exposés. Ainsi, le ratio de taux (ou ratio de risque) ne peut pas être calculé.

Un rapport de cotes peut être calculé (côtes d'exposition parmi les cas / côtes d'exposition parmi les témoins). Si ce rapport est supérieur à 1, cela implique que les cas

sont plus susceptibles d'être exposés à un facteur que les contrôles, et si le rapport est inférieur à 1, l'inverse est vrai. Si ce rapport est proche ou égal à 1, cela implique que les chances d'exposition sont très similaires dans les deux groupes (Alan et Gary, 2002).

c) Risque relatif

On appelle risque relatif (RR) le rapport de l'incidence de la maladie chez les exposés sur l'incidence chez les non exposés. Il mesure l'augmentation du risque de maladie chez les sujets exposés par rapport aux sujets non exposés ; c'est une mesure de l'intensité de l'association entre le facteur d'exposition et la maladie (Labapere, 2012).

Le risque relatif s'interprète de la manière suivante :

- Un $RR=1$ signifie qu'il n'existe pas de différence d'incidence entre les sujets exposés et ceux qui ne le sont pas. Le facteur d'exposition étudié n'est pas associé à l'apparition de la maladie
- Un $RR>1$ signifie que le facteur est lié positivement à l'apparition de la maladie.
- Un $RR<1$ signifie que le facteur d'exposition est un facteur protecteur.

1.5.3. Les mesures d'impact

La fraction étiologique permet d'exprimer la relation entre l'exposition et la maladie par un indice qui fournit un renseignement important pour la prévention. En effet, pour un groupe de sujets exposés au facteur de risque (Ranque et al, 2011), la fraction étiologique fournit directement la proportion de cas attribuable au facteur. Ainsi, même pour un risque relatif assez petit, par exemple $RR=2$, la $FE=50\%$, ce qui signifie que sur 100 cas de la maladie survenus chez les exposés, 50 sont dus au facteur de risque. Autrement dit, la suppression du facteur de risque, si la relation entre exposition et maladie est causale, conduirait à une diminution de 50% du risque de maladie chez les sujets exposés. La fraction étiologique du risque chez les exposés peut être calculée à partir d'une enquête de cohorte, et également à partir d'une enquête cas-témoins si la maladie est rare. Dans ce cas, $FE= OR-1/OR$.

La fraction étiologique du risque dans la population est un paramètre fondamental en santé publique car elle permet de quantifier la proportion de cas évitables si on supprime l'exposition.

Le risque attribuable dans la population est proportion, parmi tous les cas dans la population cible, de ceux que l'on peut attribuer à l'exposition. Il mesure l'impact global du

facteur de risque dans la population en tenant compte de la proportion des personnes exposées.

2. Plans d'échantillonnages

Un échantillon peut être obtenu de manières différentes, selon des méthodes aléatoires ou non probabilistes (Tamara et Carlos, 2017). La taille de l'échantillon à prélever dépend à la fois de la précision absolue désirée, du risque d'erreur et de la taille de la population (WHO, 2001 ; Pfeiffer, 2002 ; Bonita, 2010 ; Carrat et al, 2013).

2.1. Echantillonnage non probabiliste

La différence entre l'échantillonnage probabiliste et l'échantillonnage non probabiliste tient à une hypothèse de base au sujet de la nature de la population étudiée. Dans le cas de l'échantillonnage probabiliste, chaque unité a une chance d'être sélectionnée. Dans celui de l'échantillonnage non probabiliste, on suppose que la distribution des caractéristiques à l'intérieur de la population est égale. C'est ce qui fait que le chercheur croit que n'importe quel échantillon serait représentatif et que les résultats, par conséquent, seront exacts. Pour l'échantillonnage probabiliste, la randomisation est une caractéristique du processus de sélection, plutôt qu'une hypothèse au sujet de la structure de la population.

Voici les types les plus courants des méthodes en question (Dufour et Larivière, 2012) :

- l'échantillonnage de commodité ou à l'aveuglette ;
- l'échantillonnage volontaire ;
- l'échantillonnage au jugé ;
- l'échantillonnage par quotas.

On appelle parfois **l'échantillonnage de commodité** l'échantillonnage à l'aveuglette ou accidentel. Cet échantillonnage n'est pas normalement représentatif de la population cible, parce qu'on ne sélectionne des unités d'échantillonnage dans son cas que si on peut y avoir facilement et commodément accès.

Comme son nom l'indique, ce type d'échantillonnage intervient lorsque des gens offrent volontairement leurs services pour l'étude dont il est question. Il serait, par exemple, difficile et contraire à l'éthique dans le cadre d'expériences psychologiques ou d'essais de produits

pharmaceutiques (de tests de médicaments) de recruter au hasard pour y participer des gens du grand public. En pareils cas, on prélève l'échantillon à partir d'un groupe de volontaires.

L'échantillonnage au jugé, c'est le cas où chercheur juge que l'échantillon va lui permettre d'atteindre les objectifs de la recherche. On utilise la méthode d'échantillonnage au jugé lorsqu'on prélève un échantillon en se fondant sur certains jugements au sujet de l'ensemble de la population. L'hypothèse qui sous-tend son utilisation est que l'enquêteur sélectionnera des unités qui seront caractéristiques de la population

Lorsque le chercheur veut reproduire les caractéristiques d'une population (ex. âge, sexe, Région, etc.) dans son échantillon, il s'agit d'un **échantillonnage par quota** L'échantillonnage par quotas est l'une des formes les plus courantes d'échantillonnage non probabiliste.

2.2. Echantillonnage probabiliste

L'échantillonnage probabiliste entraîne la sélection d'un échantillon à partir d'une population, sélection qui repose sur le principe de la randomisation (la sélection au hasard ou aléatoire) ou la chance. Il est plus complexe, prend plus de temps et est habituellement plus coûteux que l'échantillonnage non probabiliste.

Toutefois, comme les unités de la population sont sélectionnées au hasard et qu'il est possible de calculer la probabilité d'inclusion de chaque unité dans l'échantillon, on peut, grâce à l'échantillonnage probabiliste, produire des estimations fiables, de même que des estimations de l'erreur d'échantillonnage et faire des inférences au sujet de la population.

Voici les méthodes d'échantillonnage probabiliste les plus courantes (Dufour et Larivière, 2012) :

- l'échantillonnage aléatoire simple ;
- l'échantillonnage systématique ;
- l'échantillonnage avec probabilité proportionnelle à la taille ;
- l'échantillonnage stratifié ;
- l'échantillonnage en grappes ;
- L'échantillonnage à plusieurs degrés ;

- L'échantillonnage à plusieurs phases.

Dans un **échantillonnage aléatoire simple** (EAS), chaque membre d'une population a une chance égale d'être inclus à l'intérieur de l'échantillon. Chaque combinaison de membres de la population a aussi une chance égale de composer l'échantillon. Ces deux propriétés sont ce qui définit un échantillonnage aléatoire simple.

Parfois appelé échantillonnage par intervalles, **l'échantillonnage systématique** (SYS) signifie qu'il existe un écart, ou un intervalle, entre chaque unité sélectionnée qui est incluse dans l'échantillon. Il faut suivre les étapes énumérées ci-dessous pour sélectionner un échantillon systématique :

- Numéroté de 1 à N les unités incluses dans la base de sondage (où N est la taille de la population totale).
- Déterminer l'intervalle d'échantillonnage (K) en divisant le nombre d'unités incluses dans la population par la taille de l'échantillon qu'on désire obtenir.
- Sélectionner au hasard un nombre entre 1 et K. Ce nombre s'appelle l'origine choisie au hasard et serait le premier nombre inclus dans votre échantillon.

Si la base de données renferme de l'information sur la taille de chaque unité (comme le nombre d'animaux de chacune des exploitations) et si la taille de ces unités varie, on peut utiliser cette information dans le cadre de la sélection de l'échantillonnage afin d'en accroître l'efficacité. Cela s'appelle **l'échantillonnage avec probabilité proportionnelle** à la taille (PPT). Dans le cas de cette méthode, plus la taille de l'unité est grande, plus sa chance d'être incluse dans l'échantillon est élevée.

Lorsqu'on utilise **l'échantillonnage stratifié**, on divise la population en groupes homogènes (appelés strates), qui sont mutuellement exclusifs, puis on sélectionne à partir de chaque strate des échantillons indépendants. On peut utiliser n'importe quelle des méthodes d'échantillonnage mentionnées dans la présente section (et il en existe d'autres) pour sélectionner l'échantillon à l'intérieur de chaque strate.

La technique de **l'échantillonnage en grappes** entraîne la division de la population en groupes ou en grappes comme son nom l'indique. Suivant cette technique, on sélectionne au hasard un certain nombre de grappes pour représenter la population totale, puis on englobe dans l'échantillon toutes les unités incluses à l'intérieur des grappes sélectionnées. On n'inclut dans l'échantillon aucune unité de grappes non sélectionnées ; ces unités sont représentées par celles tirées de grappes sélectionnées.

La méthode **d'échantillonnage à plusieurs degrés** ressemble à la méthode d'échantillonnage en grappes, sauf qu'il faut dans son cas prélever un échantillon à l'intérieur de chaque grappe sélectionnée, plutôt que d'inclure toutes les unités dans la grappe. Ce type d'échantillonnage exige au moins deux degrés. On identifie et sélectionne au premier degré de grands groupes ou de grandes grappes. Ces grappes renferment plus d'unités de la population qu'il n'en faut pour l'échantillon final. Pour obtenir un échantillon final, on prélève au second degré des unités de la population à partir des grappes sélectionnées (à l'aide de l'une des méthodes d'échantillonnage probabiliste possibles). Si l'on utilise plus de deux degrés, le processus de sélection d'unités de la population à l'intérieur des grappes se poursuit jusqu'à l'obtention d'un échantillon final.

L'échantillonnage à plusieurs phases est assez différent de l'échantillonnage à plusieurs degrés, malgré les similarités entre eux sur le plan de leur appellation. Même si l'échantillonnage à plusieurs phases suppose aussi le prélèvement de deux échantillons ou plus, dans son cas, tous les échantillons sont tirés de la même base de sondage et les unités sont structurellement les mêmes à chaque phase. Comme dans le cas de l'échantillonnage à plusieurs degrés, plus l'on utilisera de phases, plus le plan d'échantillonnage et l'estimation deviendront complexes.

3. Estimations des caractéristiques d'intérêt

En raison du temps et de l'argent dépensés pour étudier et mesurer tous les éléments d'une population, on se contente d'un échantillon à partir duquel on induira des informations sur la population dont a été tiré cet échantillon (Fallas, 2012_b; Carrat et al,

2013). On parle d'estimations (Jaime et al, 2018). Une estimation peut ponctuelle ou par intervalle (Lombardardia, 2011).

3.1. Estimation d'une proportion

L'intervalle de confiance de niveau de confiance $1 - \alpha$ de la proportion, fondé sur la valeur observée p , p étant calculée sur n individus, est donnée par l'intervalle ci-dessous :

$$IC_{1-\alpha} = \left[p - u_{\alpha, \sqrt{n}} \sqrt{\frac{p(1-p)}{n}} ; p + u_{\alpha, \sqrt{n}} \sqrt{\frac{p(1-p)}{n}} \right]$$

On remarque que c'est un intervalle symétrique par rapport à la valeur ponctuelle p .

Exemple : $n = 100$, risque d'erreur $\alpha = 0,05$ ou bien niveau de confiance $95\%=0,95$, $p = 12\%=0,12$

Conditions de validité : $n \times p \geq 5$ et $n \times (p-1) \geq 5$

3.2. Estimation de la moyenne

Le calcul de l'intervalle de confiance diffère selon qu'il s'agit de grands ou petits échantillons.

a) Cas de grands échantillons où $n \geq 30$

Dans ce cas, le calcul de l'intervalle de confiance de la moyenne « vraie » dite de la population, ne nécessite aucune autre condition à remplir. Calcul est basé sur le théorème central limite.

Si la variance vraie est connue, IC est donnée par :

$$IC_{1-\alpha} = \left[m - u_{\alpha} \frac{\sigma}{\sqrt{n}} ; m + u_{\alpha} \frac{\sigma}{\sqrt{n}} \right]$$

Si la variance de la population (variance vraie) n'est pas connue, on la remplace dans l'expression précédente par son estimation ponctuelle s^2 .

Rappelons que $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$

L'intervalle de confiance de la moyenne est calculé selon cette formule :

$$IC_{1-\alpha} = \left[m - u_{\alpha} \frac{s}{\sqrt{n}} ; m + u_{\alpha} \frac{s}{\sqrt{n}} \right]$$

Où m et s sont, respectivement la moyenne et l'écart type estimés sur l'échantillon de n individus.

b) Cas de petits échantillons où $n < 30$

Le calcul de l'intervalle de confiance, dans ce cas, n'est possible que si la variable étudiée obéit à la loi normale. Il est calculé selon la formule :

$$IC_{1-\alpha} = \left[m - t_{\alpha} \frac{s}{\sqrt{n}} ; m + t_{\alpha} \frac{s}{\sqrt{n}} \right]$$

Où t_{α} est associé à une nouvelle distribution, dite de Student, à $(n-1)$ degrés de liberté. Par conséquent, la valeur t_{α} se lit dans la table de Student à l'intersection de la ligne $(n-1)$ et de la colonne correspondant au risque α choisi, le plus souvent 0,05.

3.3. Estimation de la variance

L'intervalle de confiance de la variance vraie estimée sur un échantillon de taille n est donnée par l'expression :

$$\frac{(n-1)s^2}{\chi^2_{\frac{\alpha}{2}}} \leq \sigma^2 \leq \frac{(n-1)s^2}{\chi^2_{(1-\frac{\alpha}{2})}}$$

$\chi^2_{\frac{\alpha}{2}}$ Est la valeur qui se lit dans la table de khi-deux à l'intersection du risque $\frac{\alpha}{2}$ et du degré de liberté $(n-1)$.

$\chi^2_{(1-\frac{\alpha}{2})}$ est la valeur qui se lit dans la table de khi-deux à l'intersection du risque $1 - \frac{\alpha}{2}$ et du degré de liberté $(n-1)$, s c'est l'écart type estimé sur l'échantillon.

Pour l'intervalle de confiance de l'écart type, il suffit juste de prendre la racine carrée de l'expression précédente.

4. Tester nos hypothèses (Fallas, 2012)

Toute démarche en épidémiologie se base sur des tests d'hypothèse (Labapere, 2012). Le type de données joue un rôle très important. Il circonscrit le cadre d'application des techniques.

Pour un même objectif, selon le type de données, nous serons amenés à mettre en œuvre des tests différents (Champely, 2017). Par exemple, pour mesurer l'association entre deux variables : si elles sont quantitatives, nous utiliserons plutôt le coefficient de corrélation de Pearson ; si elles sont qualitatives nominales, ce coefficient de corrélation n'a pas de sens, on utilisera plutôt le coefficient de corrélation de Spearman (Wikipédia, 2020).

4.1. Définition et situations distinctes

Un test statistique d'hypothèse est définie comme une procédure qui permet, avec un risque d'erreur connu, d'effectuer un choix entre deux hypothèses complémentaires (H_0 et H_1) au vu des observations réalisées sur un échantillon. On distingue :

- Les tests de comparaison de deux ou plusieurs groupes entre eux ou bien d'un groupe avec une valeur de référence dite aussi de théorique (Champely, 2017).
- Les tests de liaison (Laoussati, 2016).

4.2. Tests de comparaison

Deux situations sont possibles :

- Comparer un échantillon observé à une population de référence (théorique).

Exemple : la distribution par sexe observée dans une maternité est-elle conforme à la distribution nationale ?

- Comparer deux ou plusieurs échantillons entre eux.

Exemple : comparer la fréquence du paludisme dans deux régions d'Afrique (Laoussati, 2016).

En pratique, on peut comparer deux ou plusieurs proportions, par exemple la composition raciale d'un élevage de ruminants (Meyer, 2011 ; Saidani et al, 2019), comparer deux moyennes par le test de Student ou test Z selon la taille de l'échantillon (Carrat et al,

2013). La comparaison de plus de deux moyennes se fait par l'analyse de variance (Boran, 2013).

4.3. Tests de liaison

Dans ce type de problème on se demande s'il existe une liaison entre une ou plusieurs variables étudiées sur un échantillon.

Exemple : Etudier le lien entre la consommation de glace au chocolat et la survenue de la gastroentérite (Laoussati, 2016).

Le choix du test statistique (Millot, 2018) est fonction des données du problème et de la variable(s) étudiée (s) :

- La nature de la variable (qualitative ou quantitative)
- Le nombre d'échantillons d'observations : si plus d'un échantillon, sont-ils indépendants ?
 - La taille des échantillons : grands ($n \geq 30$) ou petits ($n < 30$)

Pour étudier le lien versus indépendance entre deux variables qualitatives, on recourt au test de khi-deux. Le lien entre deux variables quantitatives se fait par la régression linéaire (Delyon, 2020).

4.4. Analyse de variance

L'analyse de la variance (ANOVA) a pour objectif d'étudier l'influence d'un ou plusieurs facteurs sur une variable quantitative (Dagnino, 2014).

C'est la comparaison de moyennes pour plusieurs groupes (> 2). Il s'agit de comparer la variance intergroupe (entre les différents groupes : écart des moyennes des groupes à la moyenne totale) à la variance intragroupe (somme des fluctuations dans chaque groupe).

S'il n'y a pas de différence entre les groupes, ces deux variances sont (à peu près) égales. Sinon, la variance intergroupe est nécessairement la plus grande. L'ANOVA se résume à une comparaison multiple de moyennes de différents échantillons constitués par les différentes modalités des facteurs. Les conditions d'application du test paramétrique de comparaison de moyennes s'appliquent donc à nouveau.

L'analyse de variance (analysis of variance ou ANOVA) peut être vue comme une généralisation du test de Student (Carbon, 2015). Les variables indépendantes sont appelées facteurs, chacune des modalités des facteurs est appelée niveau.

L'analyse de variance, tout comme la régression linéaire, fait partie des modèles linéaires (Chavent, 2013). Il existe plusieurs modèles d'analyses de variance, modèles à un facteur, à deux ou plusieurs facteurs (Bertrand et Maumy, 2011).

Si on veut étudier l'effet de la race sur la production laitière, il s'agit de l'analyse de variance à un facteur, chacune des races (Locale, Holstein, Normande) représente une modalité (un niveau) du facteur race. L'ANOVA à deux facteurs permet d'étudier, par exemple, l'effet du facteur race et du facteur fourrage sur le rendement laitier. De même, si on propose d'étudier à la fois les effets de la race, du fourrage et du numéro de lactation, il s'agira d'une ANOVA à 3 facteurs.

4.5. Régression linéaire

En présence de deux mesures prises sur les mêmes unités statistiques, on peut s'intéresser à la relation entre ces variables. Lorsqu'elles sont toutes deux quantitatives, la méthode la plus classique repose sur le coefficient de corrélation linéaire (Champely, 2017 ; Fernando, 2017). Lorsqu'il y a plus de deux variables quantitatives, il s'agit de régression linéaire multiple (Ricco, 2015 ; Boukrit, 2016 ; Lévy-Leduc, 2016 ; Fernando, 2017).

Dans le langage courant le terme de corrélation est utilisé pour désigner la liaison (relation, association) entre deux variables quelconques. Cependant, en statistique, ce terme est réservé pour désigner la liaison entre deux variables quantitatives, le plus souvent continues. Si on peut permuter entre les variables X et Y, il s'agit d'un rôle symétrique, on parle de corrélation. Si le statut des deux variables est asymétrique, Y variable dépendante (=expliquée) et X indépendante (explicative), il s'agit de la régression (Labapere, 2012a).

5. Etudes épidémiologiques et les biais

On distingue deux grands types d'études selon qu'il s'agit d'une étude expérimentale ou non, dite dans ce dernier cas observationnelle (Institut d'élevage, 2011).

5.1. Les études expérimentales

Le chercheur intervient sur le statut d'exposition des sujets. Il peut intervenir sur

- Le ou les facteurs d'exposition,
- Le moment d'exposition,
- Les individus exposés.

Dans un premier temps, les groupes vont être suivis puis comparés entre eux. En pratique, la réalisation d'études expérimentales est souvent limitée en raison de contraintes éthiques.

Le principe de randomisation

L'étude est dite randomisée si l'affectation des sujets à un groupe ou à un autre est tirée au sort. C'est le hasard qui détermine donc l'appartenance du sujet à un groupe. Ainsi, seul le groupe d'appartenance des sujets semble pouvoir différencier les sujets entre eux. La randomisation permet de limiter les biais des études

Les études ou essais randomisés sont qualifiés :

- D'ouverts quand le traitement est connu de tous,
- en simple aveugle quand seuls les patients ignorent la nature du traitement,
- en double aveugle quand patients et médecins ignorent la nature du traitement.

5.2. Les études non expérimentales ou observationnelles

5.2.1. Etudes de prévalence

Les études de prévalence permettent d'observer la fréquence de survenue d'un phénomène de santé, dans une population, à un moment précis. Il s'agit d'enquêtes transversales. Le recueil d'information s'effectue sur une période brève, « un jour donné ». La notion de suivi des patients dans le temps est absente. C'est un indicateur « statique » de morbidité.

5.2.2. Etudes d'incidence

Les études d'incidence nécessitent l'observation sur une période déterminée d'un ou de groupes d'individus pour mesurer les modifications de l'état de santé des populations. Il s'agit d'études longitudinales

5.3. Étude de cohorte

Une cohorte est un groupe de sujets suivis dans le temps. Les études de cohortes sont également appelées études exposés / non exposés. En pratique, deux groupes sont établis :

- Les individus exposés au facteur de risque
- Les individus non exposés au facteur de risque

Les deux groupes vont être suivis (études longitudinales) puis comparés entre eux. On parlera de cohorte historique si la survenue de l'exposition au(x) facteur(s) et de la maladie, a déjà eu lieu au moment où le chercheur débute son enquête (Goldberg et Zins, 2012 ; Mrabet, 2017).

5.4. Étude cas témoin

Deux groupes de sujets vont être comparés :

- Des sujets malades : « les cas »
- Des sujets non malades : « les témoins »

Le recueil d'information est toujours rétrospectif dans ce cas. Les groupes sont ensuite comparés. Les cas et les témoins doivent être choisis dans la même population. Les caractéristiques du groupe « témoin » doivent être les plus proches possible du groupe « cas ». La seule différence observable entre les groupes devrait être théoriquement l'absence de signes de la maladie.

5.5. Les biais

Les biais sont des erreurs méthodologiques ayant comme conséquence systématique de produire des estimations supérieures ou inférieures à la valeur réelle des paramètres étudiés. Les biais sont indépendants de la taille de l'échantillon. Le biais de sélection (ou de population) est celui qui discrimine la population (Problèmes dans la constitution des échantillons, perdus de vue, non répondants, etc.).

Le biais de mesure (d'investigation, ou d'information) est dû à une mesure incorrecte du facteur d'exposition (exemple : défaut de mémorisation différente chez les cas et les témoins, défaut de randomisation). Le biais de confusion concerne une mauvaise analyse ou un facteur incriminé qui n'a aucun lien de causalité avec le phénomène observé.

Toute étude présente des biais, il faut avant tout essayer de les prévoir pour les minimiser (Mrabet, 2016).

6. Régression logistique comme outil majeur en épidémiologie

6.1. Régression logistique binaire

La régression logistique binaire (Genin, 2015) est un outil statistique qui peut être employé pour évaluer et caractériser les relations entre une variable réponse de type binomiale, dite aussi variable de Bernoulli (Chesneau, 2016) comme par exemple : Vivant / Mort, Malade / Non malade, succès / échec), et une, ou plusieurs, variables explicatives, qui peuvent être de type catégoriel (le sexe par exemple), ou numérique continu (l'âge par exemple). Ainsi, la variable indépendante peut être catégorielle, quantitative discrète ou quantitative continue telle qu'une variable obéissant à la loi normale (milan, 2015).

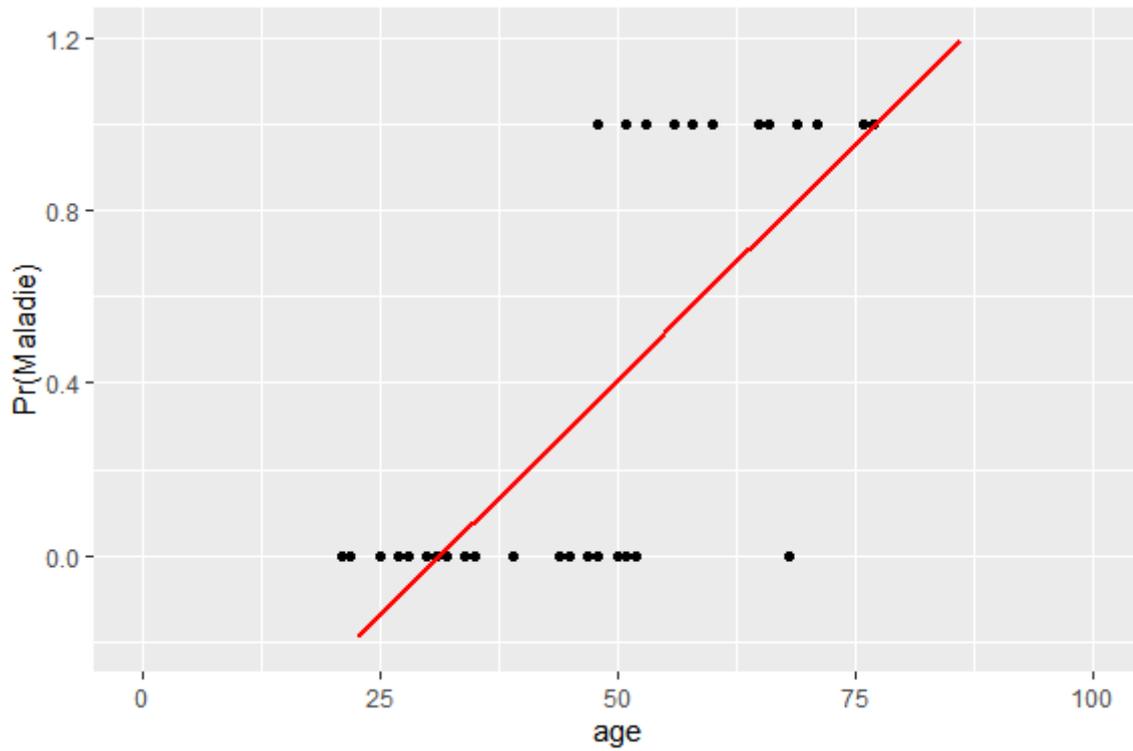
Tout comme la régression de Poisson, la régression logistique (Berlanga-Silvente et Vilà-Baños, 2014) appartient aux modèles linéaires généralisés (Abraira, 2014). Pour rappel, il s'agit de modèles de régression qui sont des extensions du modèle linéaire (Ballesteros, 2008) et qui reposent sur trois éléments :

- un prédicteur linéaire
- une fonction de lien
- une structure des erreurs

Les principaux éléments de la régression logistique, modélisation de la probabilité

Dans la régression logistique, ce n'est pas la réponse binaire (malade/pas malade) qui est directement modélisée, mais la probabilité (Rau, 2011) de réalisation d'une des deux modalités (être malade par exemple). Il est donc évident que la variable réponse n'est pas directement introduite dans le modèle linéaire (Lévy-Leduc, 2016).

Cette probabilité de réalisation ne peut pas être modélisée par une droite car celle-ci conduirait à des valeurs <0 ou >1 . Ce qui est impossible puisqu'une probabilité est forcément bornée par 0 et 1.



Cette **probabilité**, est alors modélisée par une **courbe sigmoïde**, bornée par 0, et 1 :

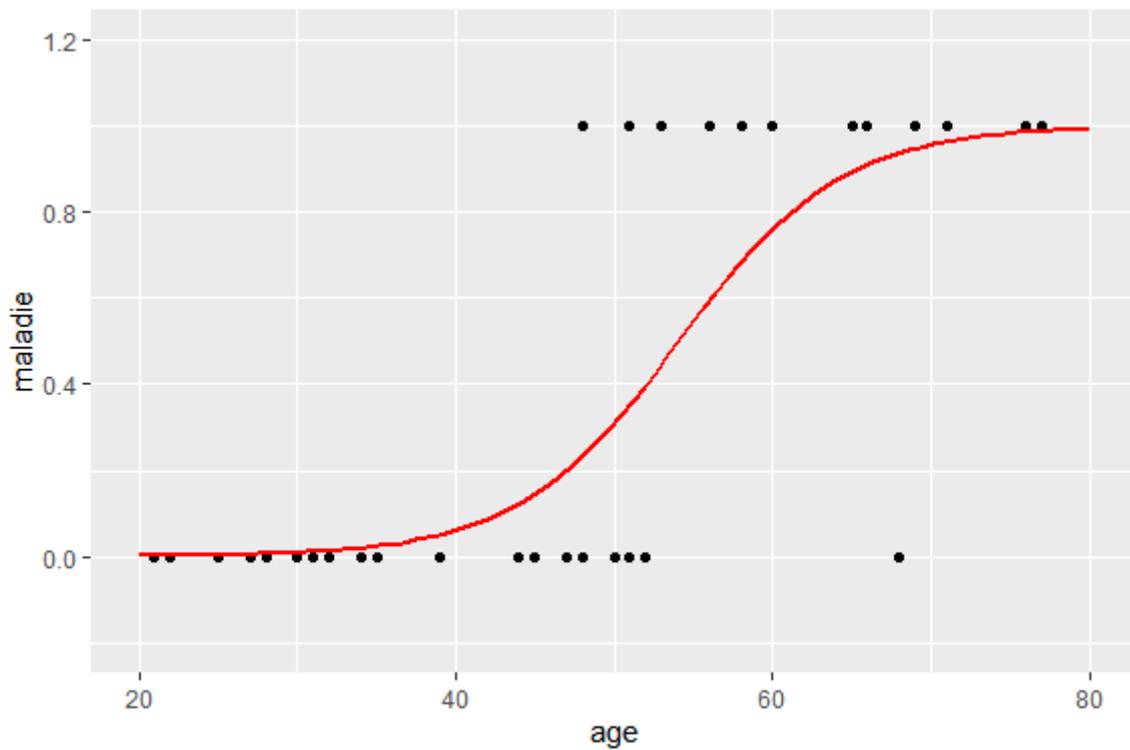


Figure 1 : Courbe sigmoïde définie par la **fonction logistique**, d'équation :

$$f(x) = \frac{\exp(x)}{1 + \exp(x)} = p$$

La fonction logistique

Lorsque la fonction logistique est ajustée à des données observées, la forme de la courbe sigmoïde s'adapte à ces données, par l'estimation de paramètres. Dans le cas d'une seule variable explicative (X), l'équation de la courbe logistique est alors :

$$P(X) = \frac{\exp(\beta_0 + \beta_1 X)}{1 + \exp(\beta_0 + \beta_1 X)}$$

La fonction de lien Logit

Le modèle précédent n'est pas linéaire dans l'expression des paramètres β_j puisque la probabilité de réalisation ne s'exprime pas comme une addition des effets des différentes variables explicatives.

Pour obtenir un tel modèle (linéaire dans ses paramètres), il est nécessaire de passer par une transformation logit : $\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \sum_{j=1}^n \beta_j X_{ij}$

Cette transformation logit est la fonction de lien qui permet de mettre en relation la probabilité de réalisation (bornée entre 0 et 1), et la combinaison linéaire de variable explicatives (Mahideb, 2012).

La structure d'erreur

Les données de base employées dans une régression logistique sont des données binaires (oui/non). Celles-ci sont distribuées selon une **loi binomiale B(1,p)**. Il en est alors de même pour les erreurs : elles sont distribuées selon une **loi binomiale B(1,p)**.

Les coefficients estimés sont des log odds ratio

Le terme $p/(1-p)$ est un rapport de cote (RC) ou Odds Ratio (OR), en anglais. Ce paramètre permet de mesurer la relation entre la variable explicative (X) et la réponse Y (vivant par exemple).

Les coefficients β_j issus de la régression logistique sont donc des log odds ratio. Un odds ou cote est le rapport de deux probabilités complémentaires : la probabilité P de survenue d'un événement (risque), divisé par la probabilité (1-P) que cet événement ne survienne pas (non risque, c'est-à-dire sans l'événement) (Jordi, 2017).

6.2. Régression à effets mixtes

Une approche bien établie de la modélisation des données groupées introduit des effets aléatoires dans le modèle d'intérêt. Des modèles de régression logistique à effets mixtes peuvent être utilisés pour prédire des variables-réponses discrètes lorsque les observations sont corrélées. Ce modèle règle donc les problèmes d'observations corrélées et d'erreur de mesure dans la variable dépendante (Jeroen, 2005).

6.3. Régression logistique polynomiale

Lorsque la variable dépendante prend K ($K > 2$) modalités, nous sommes dans le cadre de la régression logistique polytomique. Dans cette partie, nous considérons qu'elle est nominale c.-à-d. il n'y a pas de relation d'ordre entre les modalités, ou tout du moins nous souhaitons ne pas en tenir compte si elle existe. On parle de régression logistique multinomiale. On peut la voir comme une forme de généralisation de la régression logistique binaire (Ricco, 2017).

6.4. Régression logistique ordinale

La régression logistique ordinale (Benítez, 2018) un cas particulier de la régression logistique polynomiale où la variable dépendante est ordinale (Ricco, 2017).

II.1. Objectifs

Parmi les objectifs principaux de notre travail appliqué sur trois fichiers de données réelles obtenues entre 2016 et 2018 :

- Savoir planifier une étude épidémiologique
- Identifier les variables d'intérêt, leurs natures et les codifier
- Mesurer les paramètres pertinents
- Effectuer les mesures épidémiologiques de fréquences et d'association
- Analyser les données obtenues à partir des tableaux Excel en faisant appel au logiciel R et ses extensions le cas échéant.
- Exploiter les résultats d'analyses pour construire un modèle.

II.2. Les données réelles utilisées

Le premier fichier porte sur 148 prélèvements de matières fécales chez des enfants de la wilaya de Blida prélevés et examinés entre mars et avril 2018 en vue de la recherche de parasites digestifs.

Le deuxième fichier fait partie d'une étude en 2016 portant sur les tiques des bovins dans les wilayas de Bouira et Médéa.

Le troisième fichier est inhérent à étude portant sur les mammites bovines cliniques, une enquête réalisée entre 2017 et 2018, laquelle a porté sur 403 vaches laitières appartenant à quelques élevages de la wilaya de Bejaia.

II.3. Les mesures effectuées

Les mesures effectuées sur les trois types sont :

- La prévalence du parasitisme chez les enfants
- La fréquence de diagnostic de chaque parasite chez les enfants
- Fréquences des espèces de tiques bovines dans la wilaya de Blida, étude faite en 2018
- La moyenne d'intensité d'infestation par les tiques chez les bovins
- Les prévalences des mammites selon plusieurs facteurs intrinsèques et extrinsèques
- Les facteurs de risque des mammites bovines cliniques et leurs Odds ratios correspondant à travers la régression logistique standard

II.4. Plan d'échantillonnage et taille d'échantillon

Un échantillon s'obtient de manières différentes, selon des méthodes aléatoires ou non probabilistes (Tamara et Carlos, 2017). La taille de l'échantillon à prélever dépend à la fois de la précision absolue désirée, du risque d'erreur et de la taille de la population WHO, 2001 ; Pfeiffer, 2002 ; Bonita, 2010 ; Carrat et al, 2013).

Pour les besoins de notre travail, on a recouru aux méthodes aléatoires mais aussi aux méthodes non probabilistes

L'échantillonnage probabiliste entraîne la sélection d'un échantillon à partir d'une population, sélection qui repose sur le principe de la randomisation (la sélection au hasard ou aléatoire) ou la chance. Il est plus complexe, prend plus de temps et est habituellement plus coûteux que l'échantillonnage non probabiliste.

Toutefois, comme les unités de la population sont sélectionnées au hasard et qu'il est possible de calculer la probabilité d'inclusion de chaque unité dans l'échantillon, on peut, grâce à l'échantillonnage probabiliste, produire des estimations fiables, de même que des estimations de l'erreur d'échantillonnage et faire des inférences au sujet de la population.

Voici les méthodes d'échantillonnage probabiliste les plus courantes (Dufour et Larivière, 2012) :

- l'échantillonnage aléatoire simple ;
- l'échantillonnage systématique ;
- l'échantillonnage avec probabilité proportionnelle à la taille ;
- l'échantillonnage stratifié ;
- l'échantillonnage en grappes ;
- L'échantillonnage à plusieurs degrés ;
- L'échantillonnage à plusieurs phases.

Dans le cas des mammites bovines à Béjaia, on s'est basé sur un échantillon aléatoire simple pour estimer la prévalence des mammites bovines cliniques dans la wilaya de Bejaia

Dans le cas des parasites des enfants et celui des tiques bovines, il a été recouru à un échantillonnage de commodités.

On appelle parfois **l'échantillonnage de commodité** l'échantillonnage à l'aveuglette ou accidentel. Cet échantillonnage n'est pas normalement représentatif de la population cible, parce qu'on ne sélectionne des unités d'échantillonnage dans son cas que si on peut y avoir facilement et commodément accès.

II.5. Estimation de la prévalence des mammites bovines

L'intervalle de confiance de niveau de confiance $1 - \alpha$ de la prévalence, fondé sur la valeur observée p , p étant calculée sur n individus, est donnée par l'intervalle ci-dessous :

$$IC_{1-\alpha} = \left[p - u_{\alpha, N} \sqrt{\frac{p(1-p)}{n}} ; p + u_{\alpha, N} \sqrt{\frac{p(1-p)}{n}} \right]$$

II.6. Présentation du logiciel R version 3.6.0.

R est un système qui est communément appelé langage et logiciel. Il permet, entre autres, de réaliser des analyses statistiques. Plus particulièrement, il comporte des moyens qui rendent possibles la manipulation des données, les calculs et les représentations graphiques. R a aussi la possibilité d'exécuter des programmes stockés dans des fichiers textes.

En effet R possède :

- un système efficace de manipulation et de stockage des données,
- différents opérateurs pour le calcul sur tableaux (et spécialement les matrices)
- un grand nombre d'outils pour l'analyse des données et les méthodes statistiques,
- des moyens graphiques pour visualiser les analyses,
- un langage de programmation simple et performant comportant : conditions, boucles, moyens d'entrées sorties, possibilité de définir des fonctions récursives.

R est un langage orienté objet, c'est à dire que le résultat des instructions peut être stocké dans des objets qui pourront être, à leur tour, utilisés par la suite. Pour pouvoir donner une instruction, l'utilisateur doit "avoir la main", c'est à dire que le logiciel doit être prêt à la recevoir. C'est le cas lorsque le prompt (matérialisé par le signe ">") est visible dans la console. Lorsqu'une instruction comporte des erreurs, le message **Error** :

syntax error apparait. Si elle est incomplète, le signe “+” apparaît, indiquant que **R** attend la fin d’une instruction

III.1. Résultats et discussion liés aux tiques des bovins

Sur les 300 bovins, d'âges et de sexe différents examinés dans les 2 wilayat, 28 se révélés infestés, ce qui représente une prévalence globale d'infestation de 9,33%

Mais, il convient de faire remarquer que chacun bovin n'a été examiné qu'une seule fois au cours de l'étude.

Ici statistiquement parlant, il s'agit d'une variable qualitative binaire, qui n'admet que deux modalités, soit le bovin est infesté soit il n'est pas porteur de tiques.

On rappelle que la prévalence (P) représente le rapport en pourcentage du nombre d'hôtes infestés (N) par une espèce de parasite donnée sur le nombre d'hôtes examinés ; ici c'est la rapport d'animaux porteurs de tiques sur le nombre total examinés.

L'intensité parasitaire moyenne (I) correspond au rapport du nombre total d'individus d'une espèce parasite (n) dans un échantillon d'hôtes sur le nombre d'hôtes infestés ; ici, 465 tiques sur 28 bovins infestés, ce qui représente une intensité ou une charge parasitaire moyenne de 465/28 tiques par bovin infesté, soit 16,6 tiques par bovin infesté.

Dans le tableau 1, sont consignées les différentes espèces ayant pu être identifiées.

Tableau 1 : espèces de tiques

Genre	Espèce	Mâles	Femelles	Total
<i>Rhipicephalus</i>	<i>R. bursa</i>	8	6	14
	<i>R. turanicus</i>	3	5	8
<i>Hyalomma</i>	<i>H.dromedari</i>	9	12	21
	<i>H.m.rufipes</i>	19	17	36
	<i>H. impeltatum</i>	11	15	26
<i>Ixodes</i>	<i>I. ricinus</i>	1	1	2
<i>Boophilus</i>	<i>Boophilusspp.</i>	3	4	7
	Total	54	60	114

On se demande si la fréquence d'infestation par les tiques diffère selon le sexe de celles-ci.

Le test X^2 de conformité appliqué sur les données du tableau 3 a révélé qu'il n'y a de différence significative quant au sexe de la tique. La même question peut se traiter par le test de l'écart réduit étant donné qu'il y a que deux groupes à comparer, le groupe des tiques mâles et celui des tiques femelles.

Maintenant, on se propose de comparer les distributions des sept espèces identifiées toujours en appliquant le test X^2 de conformité, qui a montré que les trois espèces de *Hyalomma* sont les plus fréquemment identifiées (p-value = 8.904e-05) parmi celles-ci prélevées.

III.2 Résultats et discussion liés aux parasites des enfants

Dans le tableau 2 sont consignés le nombre de cas positifs et négatifs de parasitisme selon les différents facteurs étudiés, catégorie d'âges, sexe, provenance, niveau socioéconomique et ainsi qu'en fonction de l'altération des selles.

Tableau 2 : Nombre de cas et pourcentages de portage selon les différents facteurs étudiés

Facteur	Niveaux	N de Négatifs	N de positif	p-value
Catégories d'âges	Moins de 2 ans	6	3	0.007107**
	De 2 à 4 ans	5	4	
	De 4 à 6 ans	6	15	
	De 6 à 8 ans	6	17	
	Plus de 8ans	55	33	
Sexe de l'enfant	Fille	40	39	0.5894
	Garçon	38	31	
Provenance du prélèvement de selles	Crèche 1	9	3	0.001144**
	Crèche 2	2	9	
	Laboratoire	1	9	

Résultats et discussion

	Ecole primaire	66	49	
Aspect et consistance des selles	Ferme, normale	23	12	0.14
	Molle	19	15	
	Pâteuses	36	30	
	Liquides	10	13	
Niveau socioéconomique	Bon	42	44	0.2673
	Mauvais	36	26	
Total	148	78	70 (47,29%)	0.5108

Dans le **tableau 3** sont consignés le nombre de cas et le pourcentage du parasitisme par les oxyures selon les différents facteurs étudiés, catégorie d'âges, sexe, provenance.

Tableau 3 : Nombre de cas et pourcentages d'atteinte par les oxyures selon les différents facteurs étudiés.

Facteur	Niveaux	N de Négatifs	N de positif	p-value
Catégories d'âges	Moins de 2 ans	3	0	0.6206
	De 2 à 4 ans	14	3	
	De 4 à 6 ans	3	1	
	De 6 à 8 ans	2	0	
	Plus de 8 ans	1	1	
Sexe de l'enfant	Fille	12	3	0.7505
	Garçon	11	2	
Provenance du	Crèche 1	1	0	0.2065

prélèvement de selles	Crèche 2	13	3	
	Laboratoire	7	0	
	Ecole primaire	2	2	
Total	28	23	5 (17,85%)	0.01108*

III.3. Résultats et discussion liés aux mammites bovines cliniques

Le fichier concernant les mammites bovines cliniques nous a servi à montrer comment réaliser une régression logistique standard, sous quelles conditions et comment calculer les Odds ratio à partir des estimateurs tout en identifiant les facteurs de risque versus facteurs protecteurs ?

Dans ce qui suit, on voit les résultats de la régression logistique binaire appliquée sur le fichier des mammites bovines cliniques.

La formule qui suit indique au logiciel R, en ligne de commande, quelles sont les variables (avec leurs modalités) à tenir en compte, en spécifiant le statut de celles-ci

```
glm(formula = status ~ teatshape + system + parity + milking + housing + calfseparation + breed, family = binomial(logit), data = khelaf)
```

Dans l'encadré suivant, le dernier chiffre de chaque ligne représente la P-value du niveau du facteur considéré.

Coefficients :				
	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.557	2.907	-1.224	0.221081
teatshape[T.B]	6.713	1.545	4.346	0.0000139 ***
system[T.B]	6.065	1.521	3.988	0.0000666 ***
parity[T.B]	0.983	1.172	0.839	0.401701

```

milking[T.B]    -2.683    1.553 -1.728 0.083965 .
milking[T.C]    -8.140    2.398 -3.394 0.000689 ***
housing[T.T]     8.029    1.922  4.178 0.0000294 ***
calfseparation[T.B] -6.133    1.836 -3.339 0.000839 ***
breed[T.B]       1.308    2.139  0.612 0.540861
breed[T.C]      -0.850    1.944 -0.437 0.662024
breed[T.D]     -24.113  1534.783 -0.016 0.987465
breed[T.E]      -6.743    3.314 -2.034 0.041910 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 375.233 on 402 degrees of freedom

Residual deviance: 57.114 on 391 degrees of freedom

AIC: 81.114

Number of Fisher Scoring iterations: 15

```

> exp(coef(GLM.1)) # Exponentiated coefficients ("odds ratios")
      (Intercept)  teatshape[T.B]  system[T.B]  parity[T.B]
2.852167e-02    8.231262e+02    4.305473e+02    2.672343e+00
milking[T.B]    milking[T.C]    housing[T.T] calfseparation[T.B]
6.835483e-02    2.916676e-04    3.067317e+03    2.170970e-03
breed[T.B]      breed[T.C]      breed[T.D]    breed[T.E]

```

Résultats et discussion

3.697932e+00	4.274343e-01	3.373512e-11	1.179101e-03
--------------	--------------	--------------	--------------

Dans le dernier encadré, on remarque que les Odds ratio ne sont rien d'autre que l'exponentiel des estimateurs ou paramètres du modèles de régression logistique, ce qui est logique puisque les paramètres du modèle Logit représentent le logarithme népérien des rapports de côtes.

IV conclusion et perspectives

Par les trois types de fichiers liés respectivement au parasite des enfants, les tiques des bovins et les mammites bovines cliniques à Bejaia, il a été démontré comment utiliser des données réelles dans le domaine médical et vétérinaire pour illustrer les principaux concepts de l'épidémiologie et de la statistique.

Notamment à travers le dernier fichier, la régression logistique binaire comme outil majeur en épidémiologie analytique a été mise en œuvre pour identifier les facteurs de risques les plus influents sur la survenue des mammites bovines cliniques. Le calcul des Odds ratio a permis de quantifier les facteurs de risque versus facteurs protecteurs.

Le premier et le deuxième fichier ont servi à illustrer les variables, leurs natures et les tests d'hypothèse de base, liés aux comparaisons.

Dans un travail ultérieur, il serait plus utile d'utiliser des fichiers réels plus volumineux et appliquer sur eux les méthodes statistiques et épidémiologiques les plus récentes et plus puissantes en vue d'une meilleure exploitation des jeux de données réelles. A titre indicatif, utiliser la régression logistique polytomique pour une variable réponse qui admet plus de deux modalités, exemple statut sanitaire de la glande mammaire, indemne, mammite clinique et mammite subclinique.

Aussi, allons-nous montrer comment planifier différents type d'enquêtes épidémiologiques comment codifier les variables.

Références Bibliographique

- 1) Abdi, H. 1987. Introduction au traitement statistique des données expérimentales, PUG, 1987.
- 2) Abaira Víctor, 2014. Modelos de regresión multinivel. Hospital universitario de Ramón y Cajal Madrid Unidad de Bioestadística Clínica. ftp://ftp.hrc.es/pub/bioest/Multinivel/Modelos_multinivel.pdf
- 3) Alan J. Silman, Gary J. Macfarlane, 2002. Epidemiological Studies. A Practical Guide. Cambridge University Press.
- 4) Allaoua Noua, 2019. Techniques d'échantillonnage et analyse de données. Polycopié Université Larbi Ben M'Hidi, Oum El-Bouaghi Faculté des Sciences Exactes et des Sciences de la Nature et de la Vie. Département des Sciences de la Nature et de la Vie. <http://www.univ-oeb.dz/fsesnv/wp-content/uploads/2019/02/polycopier-Allaoua-Noua-Technique-d%C3%A9chantillonnage-pdf.pdf>
- 5) Ballesteros Sebastian, 2008. Le modèle linéaire avec R : fonction lm ().Ecole Normale Supérieure Paris France. http://rug.mnhn.fr/semin-r/PDF/semin-R_lm_SBallesteros_110308.pdf
- 6) Bastuji-Garin S, About F, Richard F, 2015. Interprétation d'une enquête épidémiologique. Santé Publique.
- 7) Benítez Miguel Arias, 2018. Regresión ordinal y sus aplicaciones. Tesis doctoral universidad de Sevilla Facultad de Matemáticas Departamento de Estadística e investigación operativa. <https://idus.us.es/bitstream/handle/11441/77492/Arias%20Ben%C3%ADtez%20Miguel%20TFG.pdf?sequence=1&isAllowed=y>
- 8) Berlanga-Silvente, V. y Vilà-Baños, R., 2014. Cómo obtener un Modelo de Regresión Logística Binaria con SPSS. [En línea] REIRE, Revista d'Innovació i Recerca en Educació, 7 (2), 105-118. Accesible en: <http://www.ub.edu/ice/reire.htm>
- 9) Bertrand Frédéric et Maumy Myrian, 2011. Quelques modèles d'analyse de variance. Institut de recherches mathématiques avancée université de Strasbourg France. http://irma.math.unistra.fr/~fbertran/enseignement/Master1_2011/Master1_Poly.pdf
- 10) Bonita Ruth, 2010. Eléments d'épidémiologie/R. Bonita, R. Beaglehole, T. Kjellström, 2^e Édition. I. Epidémiologie. 2. Manuels. I. Beaglehole, Robert. II. Kjellström. Tord. III. Organisation mondiale de la Santé. https://apps.who.int/iris/bitstream/handle/10665/44055/9789242547078_fre.pdf?sequence=1&isAllowed=y
- 11) Boran Altincicek, 2013. Step by step practical guide with Statistics (from ANOVA to survival analysis) in Biological Sciences: Or: Help, how can I analyze my "damned"

scientific data correctly and in an easy way with free R! Smashwords Edition
Copyright 2013 Boran Altincicek.

- 12) Boukrit Nouara, 2016. Régression Linéaire simple et multiple. Polycopié cours. Université de Bejaia. Faculté des sciences économiques, Commerciales et des Sciences de gestion Département des sciences de Gestion. http://univ.ency-education.com/uploads/1/3/1/0/13102001/secg_lessons06-regression_lineaire.pdf
- 13) Boulares Hamid, 2016. Polycopié de cours Mathématiques I. Université 08 Mai 1945 – Guelma Faculté de Mathématiques et de l'Informatique et Sciences de la Matière Département de Mathématiques. <http://dspace.univ-guelma.dz:8080/xmlui/bitstream/handle/123456789/635/polycopies.pdf?sequence=1&isAllowed=y>
- 14) Brunet Anne-Claire, 2016. Développement d'outils statistiques pour l'analyse de données transcriptomiques par les réseaux de co-expression de gènes. Bio-informatique [q-bio.QM]. Université Paul Sabatier -Toulouse III, 2016. Français. ffNNT : 2016TOU30373ff. fftel-01895908f
- 15) Carbon Michel, 2015. Cours d'analyse de variance. Département de Mathématiques et statistique de l'université de Laval. <http://www2.mat.ulaval.ca/uploads/media/polyANOVA.pdf>
- 16) Carrat F. Mallet A., Morice V., 2013. Biostatistique. Faculté de médecine. Université de Pierre et Marie Curie. France. 183 pages.
- 17) Chabanet Claire, Dessaint Fabrice, 2015. Déterminer la taille des échantillons. Notion sous-jacente : puissance d'un test. <http://wiki.inra.fr/wiki/cascisdi/download/Main/WebHome/puissance.pdf>
- 18) Chalât Adel, 2015. Introduction à la Biostatistique. Polycopié université Mohamed Khider Biskra Algérie.
- 19) Charout Hacene, 2019. Cours de Biostatistique au profil des étudiants des Troncs Communs Sciences de la Nature et de la Vie. Université d'Abderrahmane Mira de Bejaia Faculté des Sciences de la Nature et de la Vie. Département des Troncs Communs Sciences de la Nature.
- 20) Chesneau Christophe, 2016. Introduction aux lois de probabilité avec R. Licence. France. 2016. ffccl-01389942. <https://cel.archives-ouvertes.fr/cel-01389942/document>
- 21) Chavent Marie, 2013. Chapitre III : Modèles d'analyse de variance. http://www.math.u-bordeaux.fr/~mchave100p/wordpress/wp-content/uploads/2013/10/SlidesModStat_C3.pdf
- 22) Champely Stéphane, 2017. Tests statistiques paramétriques : taille d'effet, puissance et taille d'échantillon (package pwr). <https://clarolineconnect.univ-lyon1.fr/clarolinepdfplayerbundle/pdf/525984>

- 23) Dagnino Jorge S, 2014. Bioestadística y Epidemiología: Análisis de Varianza. Rev Chil Anest, 43: 306-310.
- 24) Delyon Bernard, 2020. La régression. Cours deuxième année de Master. Université de Rennes 1 campus de Beaulieu. <https://perso.univ-rennes1.fr/bernard.delyon/regression.pdf>
- 25) Desquilbet Loïc, 2019. Biostatistique en Médecine Vétérinaire. Ecole Nationale vétérinaire d'Alfort France.
- 26) Dufour Christine et Vincent Larivière. 2012. Principales techniques d'échantillonnage probabilistes et non-probabilistes. http://reseauconceptuel.umontreal.ca/rid=1J3BCT9WW-NJP6NT-8VW/sci6060_fiche_echant.pdf
- 27) Fallas Jorge, 2012_a. Pruebas de hipótesis. Rechazar o no H_0 : he ahí el dilema. http://www.ucipfg.com/Repositorio/MGAP/MGAP-05/BLOQUE-ACADEMICO/Unidad-2/complementarias/prueba_hipotesis_2012.pdf
- 28) Fallas Jorge, 2012_b. Intervalos de confianza. Cuantificando la variabilidad muestral. http://www.ucipfg.com/Repositorio/MGAP/MGAP-05/BLOQUE-ACADEMICO/Unidad-2/complementarias/intervalo_de_confianza_2012.pdf
- 29) Fernando San Segundo, 2017. Grado en Biología. Universidad de Alcalá Madrid España.
- 30) Ferrer Manuel Enrique Fuentes y Náyade del Prado González, 2013. Medidas de frecuencia y de asociación en epidemiología clínica. An Pediatr Contin. 2013; 11(6):346-9.
- 31) Florence Carpentier, 2015. Présentation des concepts de l'épidémiologie quantitative.
- 32) Genin Michael, 2015. Régression logistique. Epidémiologie et qualité des soins ; université de Lille 2. http://cerim.univlille2.fr/fileadmin/user_upload/statistiques/michael_genin/Cours/Modelisation/Regression_logistique.pdf
- 33) Goldberg Marcel, Zins Marie, 2012. Les études de cohorte : principes et méthode. <https://www.hcsp.fr/Explore.cgi/Telecharger?NomFichier=ad781418.pdf>
- 34) Institut d'élevage, 2011. Les différents types d'enquêtes. Séminaire d'épidémiologie animale. http://idele.fr/?eID=cmis_download&oID=workspace://SpacesStore/0b3ba25a-6458-4034-9eea-30950c70601e
- 35) Jaime Vázquez, Lizbeth Naranjo, Ruth Fuentes, Margarita Chávez, 2018. Proyecto PAPIME UNAM PE107117 "Estimación por intervalo. Estadística para estudiantes de ciencias". <http://jaimevazquezalamilla.com/wp-content/uploads/2018/05/Estimaci%C3%B3n-por-intervalo-0218.pdf>

- 36) Jeroen K. Vermunt, 2005. Mixed-Effects Logistic Regression: Models for Indirectly Observed Discrete Outcome Variables.
https://www.researchgate.net/publication/254837840_Mixed-Effects_Logistic_Regression_Models_for_Indirectly_Observed_Discrete_Outcome_Variables
- 37) Jordi Merino Noé, 2017. La potencialidad de la Regresión Logística Multinivel. Una propuesta de aplicación en el análisis del estado de salud percibido. Revista de Metodología de Ciencias Sociales. 36, enero- abril, 177-211.
- 38) Jost Christian, 2011. Analyses statistiques de base avec R et Rcmdr comme interface graphique.
- 39) Kévin Polisano, 2018. Cours de Statistique niveau L1-L2. Licence. France. 2018. ffcel-01787365f.
- 40) Labapere José, 2011. Epidémiologie étiologique.
http://campus.cerimes.fr/media/disquemiroir/2015-06-09/UNF3Smiroir/paces2011/labarere_jose/labarere_jose_p13/labarere_jose_p13.pdf
- 41) Labapere José, 2012. Chapitre 2 : Tests de comparaison de pourcentages.
http://unf3s.cerimes.fr/media/paces/Grenoble_1112/labarere_jose/labarere_jose_p06/labarere_jose_p06.pdf
- 42) Labapere José, 2012a. Chapitre 8 : La corrélation et la régression simple.
http://unf3s.cerimes.fr/media/paces/Grenoble_1112/labarere_jose/labarere_jose_p08/labarere_jose_p08.pdf
- 43) Laoussati M. 2016. Principes des tests d'hypothèse. Cours 1^{ère} année médecine université de Bejaia.
http://univ.ency-education.com/uploads/1/3/1/0/13102001/stat1an06-principe_tests-laoussati.pdf
- 44) Laurent Thibault, 2014. Les tests statistiques.
<http://www.thibault.laurent.free.fr/cours/tests.pdf>
- 45) Lévy-Leduc Céline, 2016. Notes de cours sur les bases statistiques du modèle linéaire.
https://www6.inra.fr/mia-paris/content/download/3995/38380/file/notes_cours_mod_lin_site_web.pdf
- 46) Lombardardia Cortiña María José, 2011. Estimación por intervalo de confianza. Departamento de Matemáticas universidad de Coruña.
http://eio.usc.es/eipc1/base/BASEMASTER/FORMULARIOS-PHP/MATERIALESMATER/Mat_41855_T3IC.pdf
- 47) Mahideb Saadia, 2012. Comparaison des modèles Logit et Probit et applications en Epidémiologie. Mémoire de magistère en mathématiques option probabilités et statistique. Université des frères Mentouri Constantine faculté des sciences exactes.
<https://bu.umc.edu.dz/theses/math/MAH6214.pdf>

- 48) Meyer Nicolas, 2011. Comparaisons de proportions. Laboratoire de Biostatistique et Informatique Médicale Faculté de Médecine de Strasbourg. <http://udsmed.u-strasbg.fr/labiostat/IMG/pdf/proportions-2.pdf>
- 49) Milan Paul, 2015. Lois de probabilités à densité, loi Normale. https://www.lyceedadultes.fr/sitepedagogique/documents/math/mathTermS/11_lois_densite_loi_normale/11_cours_lois_densite_loi_normale.pdf
- 50) Millot Gaël, 2018. Comprendre et réaliser les tests statistiques à l'aide de R. Manuel de Biostatistique. 4e Édition | Janvier 2018 | 960 pages.
- 51) Mrabet M. 2016. Les Biais en Epidémiologie. Université Mohamed V Rabat. Maroc. <http://fmp.um5.ac.ma/sites/fmp.um5.ac.ma/files/LES%20BIAS%20EN%20EPIDEMIOLOGIE.pdf>
- 52) Mrabet M. 2017. Les différentes études en épidémiologie. Université Mohamed V Rabat Maroc. <http://fmp.um5.ac.ma/sites/fmp.um5.ac.ma/files/Diff%C3%A9rentes%20%C3%A9tudes%20%C3%A9pid%C3%A9miologiques.compressed.pdf>
- 53) Pfeiffer U. Dirk, 1999. Veterinary epidemiology: An introduction. Department of Veterinary Clinical Sciences. The Royal Veterinary College, University of London.
- 54) Pfeiffer U. Dirk, 2002. Veterinary epidemiology: An introduction. Department of Veterinary Clinical Sciences. The Royal Veterinary College, University of London.
- 55) Ranque B, Mechtouff L, Grabar S, 2011. Épidémiologie étiologique : du facteur de risque à la cause. Sang Thrombose Vaisseaux 2011 ; 23 (5) : 242-52. doi:10.1684/stv.2011.0607
- 56) Rau Clement, 2011. Probabilités (Année spéciale). Université Paul-Sabatier-Toulouse-3. <http://www.math.univ-toulouse.fr/~rau/as/cours%20proba%20as.pdf>
- 57) Ricco Rakotomalala, 2011. Tests de normalité. Techniques empiriques et tests statistiques. http://eric.univ-lyon2.fr/~ricco/cours/cours/Test_Normalite.pdf
- 58) Ricco Rakotomalala, 2015. Pratique de la Régression Linéaire Multiple : Diagnostic et sélection de variables. Université Lumière Lyon 2. http://eric.univ-lyon2.fr/~ricco/cours/cours/La_regression_dans_la_pratique.pdf
- 59) Ricco Rakotomalala, 2017. Pratique de la Régression Logistique. Régression Logistique Binaire et Polytomique. Université Lumière Lyon 2. http://eric.univ-lyon2.fr/~ricco/cours/slides/regression_logistique.pdf
- 60) Roger Noel, 2006. Statistique descriptive. http://amphimaths.chez-alice.fr/N1/stats_desc_poly.pdf

- 61) Rosner Bernard, 2016. Fundamental Biostatistics. Harward University. 8th Edition. <http://galaxy.ustc.edu.cn:30803/zhangwen/Biostatistics/Fundamentals+of+Biostatistics+8th+edition.pdf>
- 62) Saidani Khelaf, López-Sández C, Ziam H, Hamiroune M, Righi S, Díez-Baños P, Panadero-Fontán R y Fernández-Rodríguez G 2018: La mastitis bovina clínica en el norte de Argelia: factores de riesgo y plan de control. *Livestock Research for Rural Development*. Volume 30, Article #139. Retrieved September 29, 2018, from <http://www.lrrd.org/lrrd30/8/kamel30139.html>
- 63) Saidani Khelaf, Ziam Hocine, Hamiroune Mourad, Righi Souad, Benakhla Ahmed, 2019. Small ruminant rearing in Kabylia, Algeria, and prospects for its development. *Rev. Elev. Med. Vet. Pays Trop.*, 72 (2): 00-00, doi: 10.19182/remvt.31745.
- 64) Tamara Otzen & Carlos Manterola, 2017. Técnicas de Muestreo sobre una Población a Estudio. *Int. J. Morphol.*, 35 (1):227-232.
- 65) Thrusfield Michael, 2007. Veterinary epidemiology. Third edition. Blackwell Publishing. http://dvmbooks.weebly.com/uploads/2/2/3/6/22365786/1_veterinary_epidemiology_thrush_filled.pdf
- 66) WHO, 2001. World Health Organization (WHO). Epidemiology: as a tool of assessment of risk. Water Quality: Guidelines, Standards and Health. Edited by Lorna Fewtrell and Jamie Bartram. Published by IWA Publishing, London, UK. ISBN: 1 900222280. https://www.who.int/water_sanitation_health/dwq/iwachap7.pdf