

# ENSEIGNER LE RECUEIL DE DONNÉES : ÉTUDE DE CAS EN ÉPIDÉMIOLOGIE

Karen Leffondré<sup>1,2</sup>, Marthe-Aline Jutand<sup>1</sup>, Marianne Savès<sup>1</sup>, Valérie Kiewsky<sup>1</sup>

<sup>1</sup> *Université de Bordeaux, ISPED, 146 rue Léo Saignat, 33000 Bordeaux*

<sup>2</sup> *INSERM U1219, équipe biostatistique, 146 rue Léo Saignat, 33000 Bordeaux*

**Résumé.** Une bonne compréhension de la façon dont les données sont recueillies est un élément fondamental de la démarche statistique. Cette compréhension permet en effet une sélection adéquate de la méthode d'analyse des données et une interprétation juste des résultats. La méthode de recueil de données englobe non seulement l'outil et le mode de recueil, mais aussi la méthode de sélection des individus (ou unités) dans la population d'intérêt. Dans cette présentation, nous expliquerons comment nous procédons dans le Master de Santé publique à l'Université de Bordeaux pour enseigner le recueil de données, notamment dans le parcours épidémiologie. Nous présenterons en particulier trois exemples d'approches pédagogiques permettant aux étudiants de mesurer l'importance 1) de la formulation des questions dans un questionnaire, 2) de son mode administration, et 3) du choix de la méthode d'échantillonnage des sujets dans la population et de son impact sur les estimations.

**Mots-clés.** Enseignement de la statistique, recueil de données, questionnaire, échantillonnage, épidémiologie.

**Abstract.** Understanding the way the data are collected is a key component of statistical thinking. A good understanding of this key component indeed allows an appropriate selection of the statistical method to be used to analyze the data, as well as an appropriate interpretation of results. Data collection methods include not only the way the data are collected from participants but also the way the subjects are selected from the population of interest. In this talk, we will present how we teach data collection methods in the Master of Public Health at the University of Bordeaux, in particular in the Epidemiology stream. More specifically, we will present three examples of pedagogical approaches that are designed to help students understand the importance of 1) the wording of questions in a questionnaire, 2) the way the questionnaire is administered, and 3) the choice of the sampling scheme of the subjects within the population and its impact on estimates.

**Keywords.** Statistics education, data collection, questionnaire, sampling scheme, epidemiology.

# 1 Introduction

Afin de permettre aux étudiants de développer une pensée et un raisonnement statistiques, il est recommandé d'enseigner progressivement les quatre étapes fondamentales de la démarche statistique, c'est-à-dire : la formulation de la question, la collecte des données, l'analyse des données, et l'interprétation des résultats (Franklin et al. 2005, Gattuso 2011).

Une bonne compréhension de la façon dont les données sont collectées est un élément fondamental de cette démarche. Cette compréhension permet en effet une sélection adéquate de la méthode d'analyse des données et une interprétation juste des résultats.

Il existe certainement une grande variabilité intra et inter-disciplines dans la manière d'enseigner le recueil de données à l'université. Dans cette présentation, nous expliquerons comment nous procédons dans le Master de Santé publique à l'Université de Bordeaux pour enseigner le recueil de données, en particulier dans le parcours épidémiologie.

## 1.1 Définition de l'épidémiologie et particularité des formations en épidémiologie

L'épidémiologie est l'étude de la distribution des problèmes de santé dans le temps et dans l'espace, et de leurs déterminants dans les populations humaines. Le coeur de métier de l'épidémiologiste est de formuler une question à propos de la fréquence ou des causes d'un problème de santé particulier, de justifier cette question à partir des données de la littérature, de proposer un schéma d'étude permettant de collecter les données nécessaires pour répondre à la question, d'organiser le recueil des données, d'analyser les données, d'interpréter les résultats, et enfin de communiquer les résultats en les questionnant de manière scientifique. La démarche épidémiologique est donc étroitement liée à la démarche statistique.

Les formations en épidémiologie sont essentiellement dispensées au niveau du deuxième cycle universitaire. Cette discipline attire des étudiants issus de formations initiales diverses incluant, entre autres, la médecine, la pharmacie, les sciences infirmières, la biologie, les mathématiques, la statistique, la psychologie, et la géographie. Les Masters en épidémiologie doivent donc s'adapter à un public très large n'ayant pas nécessairement de base en statistique à l'entrée du Master 1. Pourtant, à l'issue du Master 2, tous les étudiants en épidémiologie doivent avoir acquis les bases de la démarche épidémiologique. Les différentes étapes de cette démarche doivent par conséquent être enseignées progressivement tout au long des deux années de Master. Le recueil des données, qui est une de ces étapes fondamentales de la démarche épidémiologique, doit par conséquent lui aussi être enseigné pas à pas.

A l'Université de Bordeaux, l'épidémiologie est l'un des huit parcours du Master en santé publique (1. Santé publique ; 2. Épidémiologie ; 3. Biostatistique ; 4. Santé internationale ; 5. Promotion de la santé et développement social ; 6. Santé, travail, environnement ; 7. Management des organisations médicales et médico-sociales ; 8. Système d'information et technologies informatiques pour la santé). Le premier semestre du Master 1 est commun à tous les parcours, et les cours dispensés pendant ce premier semestre doivent donc s'adapter à un public encore plus large que le public attiré par l'épidémiologie. Les étudiants commencent à se spécialiser dans le parcours de leur choix au cours du second semestre du Master 1 en choisissant des UE optionnelles appropriées au parcours. Le Master 2 est spécifique à chaque parcours. Le recueil de données est enseigné au cours de tous les semestres du Master 1 et du Master 2 épidémiologie.

## 1.2 Les éléments du recueil de données

La méthode de recueil de données englobe non seulement l'outil et le mode de recueil, mais aussi la méthode de sélection des individus (ou unités) dans la population d'intérêt.

Il existe différentes méthodes d'échantillonnage des sujets dans une population. En épidémiologie, on s'intéresse le plus souvent à la relation entre un problème de santé et une exposition à un facteur qui pourrait être une cause de ce problème de santé. Selon le type de problème de santé et l'exposition étudiée, et en fonction des contraintes temporelles et budgétaires de l'étude, on peut sélectionner les sujets avant ou après le développement du problème de santé, et on peut contrôler ou non l'exposition étudiée. On distingue ainsi les schémas d'études expérimentales où l'exposition d'intérêt est dite "contrôlée" (c'est-à-dire imposée par l'investigateur par tirage aléatoire des sujets dans le groupe exposé ou dans le groupe non exposé), et les schémas d'études observationnelles où rien n'est contrôlé si ce n'est la façon de sélectionner les sujets. Ces dernières incluent les études transversales et les études de cohorte où les sujets sont sélectionnés indépendamment de leur statut malade-non malade, et les études cas-témoins où les sujets sont sélectionnés conditionnellement à ce statut.

Selon la question posée et la population étudiée, il se peut que les données nécessaires pour répondre à la question soient déjà collectées dans des bases de données existantes, comme par exemple les bases de données de l'Assurance Maladie ou des réseaux de surveillance de certaines maladies. La question du recueil pour ces études se limite donc principalement à la question du schéma d'échantillonnage dans ces bases de données.

De nombreuses questions nécessitent cependant la collecte de nouvelles données. Celles-ci doivent souvent être collectées par le biais d'un questionnaire adressé aux participants de l'étude. Ce questionnaire doit être conçu de manière à pouvoir répondre précisément à toutes les questions de l'étude. Les questions doivent être formulées le plus clairement possible, de manière à minimiser les erreurs de mesure qui peuvent être plus ou moins

importantes selon les participants et le mode d'administration du questionnaire. Un même outil de recueil peut en effet être administré selon des modes de recueil bien différents : auto-administration, échange téléphonique ou face à face enquêteur-enquêté.

## 2 L'enseignement du recueil de données dans notre Master

### 2.1 L'enseignement de l'outil et du mode de recueil de données

La conception du questionnaire est abordée dans une unité d'enseignement (UE) du Master 1 dédiée au recueil de données dans les enquêtes. Les éléments fondamentaux de la conception d'un questionnaire sont d'abord présentés, puis les étudiants sont amenés à concevoir eux-mêmes un questionnaire dans le cadre d'un projet qui consiste à concevoir un protocole d'enquête fictif. Pour ce projet, les étudiants travaillent par petits groupes de trois ou quatre étudiants, et doivent proposer une méthode de recueil de données pour répondre à une question particulière. Les étudiants doivent donc non seulement proposer la façon dont les sujets vont être échantillonnés dans la population d'intérêt, mais aussi comment ces sujets vont être interrogés. Les étudiants sont ainsi amenés à rédiger un questionnaire et à tester ce questionnaire sur des participants fictifs, qui peuvent être les étudiants de la même promotion mais travaillant sur un autre protocole d'enquête. Le questionnaire proposé peut donc être évalué par les pairs. Il est également évalué par les enseignants qui accompagnent les étudiants tout au long du projet. Le questionnaire fait également l'objet de discussion lors de la restitution orale du projet devant toute la promotion. Dans la présentation, nous illustrerons cette expérience d'enseignement autour de la construction d'un questionnaire fictif (*Exemple pédagogique 1*).

Dans une autre UE du Master 1 intitulée "Projet tutoré", certains étudiants peuvent être amenés à mener une petite enquête sur le terrain. En interrogeant eux-mêmes les "vrais" participants de l'étude, les étudiants peuvent alors apprécier l'importance de a) la formulation des questions du questionnaire (fermées/ouvertes), b) l'adaptation du questionnaire à la population interrogée (jeune/âgée, éduquée/peu éduquée, active/inactive), et c) la façon dont le questionnaire est administré (auto-administration/téléphone/face-à-face). Comme pour l'UE spécifique au recueil de données du premier semestre, les étudiants travaillent en petits groupes autour d'un projet commun, et chaque projet est évalué à l'écrit et à l'oral par l'équipe pédagogique. Dans cette présentation, nous expliquerons comment nous procédons dans cette UE pour accompagner les étudiants dans leur démarche et donnerons quelques exemples d'enquêtes réelles menées par nos étudiants. Nous discuterons des avantages et des inconvénients d'une telle approche pédagogique tournée autour de la mise en oeuvre sur le terrain d'un recueil de données (*Exemple pédagogique 2*).

## 2.2 L'enseignement de la méthode d'échantillonnage

La méthode de sélection des sujets dans la population (c'est à dire le "schéma d'échantillonnage", ou encore le "schéma d'étude") est quant à elle enseignée progressivement dans différentes UE réparties sur les deux années de Master. Cette étape fondamentale de la démarche épidémiologique est enseignée au travers :

1. De présentations des différents schémas d'études en épidémiologie ;
2. De temps d'exercices où les étudiants sont amenés, à partir de questions précises, à critiquer des schémas d'études utilisés dans des articles publiés, ou à réaliser eux-mêmes l'échantillonnage des sujets dans une base de données existante ;
3. De projets où les étudiants sont amenés à proposer eux-mêmes le meilleur schéma d'étude en fonction de la question posée (voir les deux UE du Master 1 évoquées dans la section précédente) ;
4. De stages de Master 1 et de Master 2 où les étudiants sont amenés à proposer eux-mêmes un schéma d'étude, ou à défendre et discuter avec un esprit critique le schéma d'étude qui leur a été imposé pour répondre à la question de recherche. En Master 2, cet esprit critique est développé tout au long de l'année au travers de quatre ateliers où les étudiants présentent, devant toute la promotion et des membres de l'équipe pédagogique, l'état d'avancement de leur réflexion sur leur sujet de stage.

Nous sommes actuellement en plein processus de refonte de la maquette du Master, et, pour le parcours épidémiologie, nous avons en particulier décidé de restructurer le Master 2 en fonction des grands schémas d'étude utilisés en épidémiologie (études expérimentales, études de cohorte, études cas-témoins, et surveillance épidémiologique). Chaque schéma d'étude sera abordé dans une UE à part entière où le recueil de données spécifique au schéma sera explicité avant d'aborder les méthodes d'analyse des données découlant d'un tel schéma.

Par exemple, pour l'UE sur les études cas-témoins, nous prévoyons un exercice où les étudiants auront, en fonction de la question posée et des contraintes budgétaires de l'étude, à proposer le meilleur schéma d'échantillonnage des cas (les malades) et des témoins (les non malades) dans une base de données existante, et à réaliser eux-même cet échantillonnage aléatoire. Les étudiants devront ensuite analyser les données de leur propre échantillon, comparer les estimations obtenues à celles obtenues sur la base de données entière, et également à celles obtenues par les autres étudiants sur leurs propres échantillons. Cette comparaison de résultats permettra, en plus d'illustrer l'impact du schéma d'étude choisi, d'illustrer la notion de fluctuations d'échantillonnage (*Exemple pédagogique 3*).

### 3 Conclusion

Dans cette présentation, nous expliquerons donc la démarche générale que nous avons adoptée au sein de notre Master de Santé publique pour enseigner le recueil de données. Nous présenterons en particulier trois exemples d'approches pédagogiques permettant aux étudiants de mesurer l'importance de la construction du questionnaire (Exemples pédagogiques 1 et 2), de son mode d'administration (Exemple pédagogique 2), et du choix et de l'impact de la méthode d'échantillonnage des sujets (Exemple pédagogique 3).

### Bibliographie

Franklin C, Kader G, Mewborn D, Moreno J, Peck R, Perry M, and Scheaffer R. (2005) *Guidelines for Assessment and Instruction in Statistics Education (GAISE) report : A Pre-K-12 curriculum framework*. Alexandria, VA : American Statistical Association.

Gattuso L. (2011). L'enseignement de la statistique : où, quand, comment, pourquoi pas ? *Statistique et Enseignement* 2(1) :5-30.