

Traitement des données manquantes dans les enquêtes par sondage

▪ Stage proposé par

Direction/ Cire : DATA

Maîtres de stage / personne contact :

Nom : ZEGHNOUN

Prénom : Abdelkrim

Téléphone : 01.41.79.69.23 Adresse email : abdelkrim.zeghnoun@santepubliquefrance.fr

▪ Type de stage proposé

Master 1

Master 2 Professionnel

Master 2 Recherche

Extension possible au delà de la période obligatoire

Oui

Non

Commentaires :

▪ Date proposée pour le stage et durée

pas de contrainte de date

A partir de _____

Durée en mois : 6 mois

▪ Sujet proposé pour le stage

Santé publique France est l'agence nationale de santé publique française. Elle intervient au service de la santé des populations. Agence scientifique et d'expertise du champ sanitaire, elle a pour missions : (1) l'observation épidémiologique et la surveillance de l'état de santé des populations ; (2) la veille sur les risques sanitaires menaçant les populations ; (3) la promotion de la santé et la réduction des risques pour la santé ; (4) le développement de la prévention et de l'éducation pour la santé ; (5) La préparation et la réponse aux menaces, alertes et crises sanitaires ; (6) le lancement de l'alerte sanitaire. L'agence est organisée autour de 10 directions scientifiques et transversales et de 6 directions assurant le support et le soutien à l'activité.

Le stage se déroulera dans l'unité « Appui à la conception, à la mise en place et à l'exploitation des enquêtes » de la Direction Appui, Traitements et Analyses des données (DATA). Cette unité a de nombreuses missions au sein de l'agence dont le montage d'enquêtes et la participation à leur coordination, la contribution à l'élaboration du plan de sondage, le data-management des données, la réalisation des analyses statistiques prenant en compte le plan de sondage...

Malgré les efforts réalisés au moment du recueil des données d'une enquête, le statisticien est souvent confronté au problème de la non réponse. Cette dernière réduit le nombre de données à exploiter dans les analyses statistiques et a pour conséquence d'altérer la qualité des estimations (augmentation de la variance, biais). La non réponse peut être totale lorsque aucune donnée n'est disponible pour un participant ou partielle lorsque seules quelques données sont indisponibles pour ce participant.

Ce stage s'intéresse à la non réponse partielle qui est généralement traitée par imputation. Cette dernière permet d'attribuer des valeurs de remplacement aux données manquantes. Après imputation, la base de données de l'enquête devient complète et peut ensuite être utilisée dans l'estimation des paramètres d'intérêt.

L'objectif de ce stage est d'aider le statisticien d'enquête dans le choix des méthodes d'imputation à utiliser pour ses données d'enquête. Pour cela, le stagiaire devra comparer plusieurs méthodes d'imputation sur des données simulées mais aussi sur des données issues des enquêtes réalisées par Santé publique France (enquêtes ESTEBAN,

ENNS, ELFE, COSET...). Dans le cas des données simulées, le stagiaire étudiera l'effet du mécanisme de non réponse (uniforme, ignorable ou non) dans l'estimation des paramètres d'intérêt. Il comparera les performances de ces estimations en termes de biais et de variance.

▪ **Prérequis**

- Aucun
- Compétences spécifiques (préciser) : Si possible, connaissance des méthodes d'imputation
- Maîtrise d'un logiciel spécifique (préciser) : R et/ou SAS
- Autre (préciser) :

▪ **Commentaires**

Le stagiaire sera co-encadré par Abdessattar Saoudi, Statisticien à l'unité dans l'unité « Appui à la conception, à la mise en place et à l'exploitation des enquêtes » de la Direction Appui, Traitements et Analyses des données (DATA).
Coordonnées : 01 41 79 69 28 (Abdessattar.SAoudi@santepubliquefrance.fr)