

Rmd_traitement

Moustapha Ndiaye

2025-05-31

Préparation et traitement des données Le traitement des données s'est déroulé en plusieurs étapes afin de garantir une base propre et exploitable pour l'analyse statistique.

1. Importation et sélection des données Les données initiales provenaient de plusieurs fichiers correspondant à différentes sections de l'enquête (notamment s01, s02, s03, s04, s11). Pour chaque fichier, seules les variables pertinentes à l'étude des déterminants de santé et aux caractéristiques socio-économiques ont été extraites.
2. Fusion des bases de données Les différents fichiers ont été fusionnés en une seule base principale grâce à des clés d'identification communes (numéro de vague, grappe, ménage, individu). Cette opération a permis d'obtenir un tableau complet regroupant toutes les informations nécessaires pour chaque individu.
3. Nettoyage des données Avant toute analyse, un nettoyage des données a été effectué pour éliminer les observations non valides ou incomplètes, telles que les enregistrements sans identifiants complets ou avec des incohérences manifestes.
4. Traitement des valeurs manquantes Les valeurs manquantes ont été traitées de manière adaptée selon le type de variable :

Variables numériques : les valeurs manquantes ont été remplacées par la médiane des observations disponibles. Cette méthode est robuste aux valeurs extrêmes et permet de conserver la distribution des données.

Variables catégorielles : les valeurs manquantes ont été imputées par la modalité la plus fréquente (mode) ou, lorsque pertinent, par la création d'une nouvelle catégorie « Inconnu » pour ne pas biaiser les distributions.

5. Recodage des variables Pour faciliter l'interprétation et l'analyse, plusieurs variables ont été renommées ou recodées. Par exemple :

La variable indiquant la présence d'un problème de santé a été nommée `probleme_sante`.

Le recours aux soins est représenté par la variable `a_consulte`.

Les types de structures consultées ont été codés sous `type_structure`.

La satisfaction vis-à-vis des soins a été catégorisée sous `satisfaction_soins`.

Ces recodages permettent une meilleure lisibilité et manipulation des données lors des analyses.

6. Constitution d'un sous-échantillon Pour certaines analyses avancées, notamment l'Analyse en Composantes Principales (ACP), un sous-échantillon d'individus complets (sans valeurs manquantes sur les variables sélectionnées) a été extrait. Cela garantit la qualité et la robustesse des résultats statistiques obtenus.