

Analyse statistique et recueil de données
Guide de bonne pratique
Franck Le Duff

Vous réalisez un travail épidémiologique ou statistique et vous souhaitez un soutien méthodologique ? Avant de poursuivre (ou commencer) votre travail et avant de contacter votre référent méthodologique, assurez-vous d'avoir respecté ces quelques conseils.

Forme du fichier :

- Utiliser un tableur type Excel avec une seule feuille (la première) pour toutes les données ou une base de données Access exportée sous Excel.
- Mettre les cas (patients) en ligne avec un numéro de ligne et les variables en colonnes : chaque cas doit avoir un numéro.
- S'assurer que les noms des variables sont sur 8 caractères maximum, en minuscule, sans espace et sans caractère spécial (ex : -(' »@ 0 ••). Ces noms de variables doivent être explicites
- S'assurer qu'il n'y a pas de doublons dans le nom des variables
- Dans la mesure du possible, réaliser un tableau où toutes les données d'un patient sont sur une ligne et une seule.
- Eliminer couleurs et format de mise en page du tableau et des caractères : faire le plus simple et le plus dépouillé possible.

Codage des variables :

- En général évitez les variables de format « texte ». Préférez « numérique ».
- **Pour les variables quantitatives** (poids, taille, âge, mesure de paramètres biologiques..) respecter les mêmes échelles de valeurs. Les valeurs manquantes sont des cases vides (ni texte, ni zéro, ni autre signe, en particulier PAS de point d'interrogation ni astérisque). Pas de lettre ni de signe: <, >, +, - dans une colonne avec une variable contenant des chiffres.
- **Pour les variables qualitatives** (variable en classes: oui/non...) respecter un codage identique pour l'ensemble des patients. En particulier, ne pas coder par '0' (la lettre) si vous avez commencé par coder 0 (le chiffre zéro). Les valeurs manquantes peuvent être soit un code décidé à l'avance (99 par exemple) soit une case vide.
D'une manière générale sont conseillés : OUI et le sexe masculin = 1, NON et le sexe féminin = 2 , NON CONCERNE = 8, SANS REPONSE ou inconnu = 9.
- **Respecter les formats de dates:** « formater » les cellules du tableau d'emblée pour toutes les colonnes et choisir un format avec 4 chiffres pour l'année avec dans l'ordre : jour/mois/année (ex: 01/01/2003)
- **Respecter les formats des variables:**
 - > Ne pas mélanger des valeurs numériques à des lettres
 - > Utiliser la virgule et non le point pour les valeurs réelles pour les décimales
- **Prévoir une notice de codage** et de spécification des variables avec le nom de la variable, sa signification, les valeurs possibles et leur correspondance en clair (ex. Variable sex = sexe des patients, codé 1 pour homme et 2 pour femme).
- Au besoin, faites une différence entre des valeurs manquantes et les autres valeurs non prévues par le protocole (ex. Rythme cardiaque codé 1 si asytolie, 2 si fibrillation ventriculaire ou 3 autres. Laisser vide ou coder délibérément par 9 si non connue)
- Evitez le texte libre.
- Attention à l'utilisation des variables répondant à des questions ouvertes. Exemple: « où avez-vous eu le renseignement?».
- Attention à la classe « autre » lequel.....(texte libre)

Format de fichier :

- Sauvegardez vos données au format Excel. Si vous avez plusieurs fichiers, sauvegardez l'ensemble des fichiers dans un dossier et compactez le dans une archive unique (auto décompactable si possible).