



**JisStat**

## Tableau de données

Il est fortement recommandé de suivre les conseils de la cnil en matière d'anonymisation et de sécurité de vos fichiers. La loi est claire à ce sujet. N'hésitez pas à aller faire un tour sur leur site.

## Les fichiers

- La base de données doit être sous forme de fichier de tableur : .xls, .xlsx ou .csv

- S'il en existe plusieurs vous pouvez utiliser plusieurs fichiers voir des feuilles différentes sur excel.
- Un dictionnaire des variables, fichier ou une feuille excel, doit être présent (voir ci-dessous)
- Le nom des fichiers doit être explicite et obéir aux mêmes règles que les noms de variables. (voir ci-dessous)

## Le dictionnaire de variables

Ce fichier est un tableau qui doit contenir les colonnes suivantes (sans majuscule ni accent) :

- variable : dedans sera renseigné le nom de la variable
- description : une très courte description de ce que représente la variable. Inutile si le nom est suffisamment explicite. (*ex : vih\_pi : primo infection vih*)
- type : quantitatif, binaire, qualitatif ou date
  - quantitatif : des nombres avec lesquels on peut faire des additions et des multiplications (donc pas des paliers d'antalgique par exemple)
  - binaire : TRUE ou FALSE (en majuscule)
  - qualitatif : une variable pour laquelle il existe plusieurs possibilités et où il n'est pas pertinent de faire des additions et des multiplications
  - date : une date
- détail :
  - si la variable est quantitative : laisser vide
  - si la variable est binaire laisser vide
  - si la variable est qualitative : la liste des valeurs autorisées séparées par des virgules (*ex : bleu, vert, rouge, blanc ; ou : a, b, c*)
  - si c'est une date : le format de la date (*ex : dd/mm/yy ; mm-dd-yyyy*) avec 'd' pour day, 'm' pour month et 'y' pour year

S'il y a plusieurs tableaux (ou feuilles excel) ajouter une colonne intitulée « tableau » avec le nom du tableau. Cela permet de savoir de quel tableau vient la variable.

## La base de données

Cases :

Pas de couleur, pas de fusion, une colonne = une variable (si vous ne savez pas ce qu'est une variable posez la question), une ligne= un patient ou une mesure dans le cas de mesure répétées (patient 1 à J1, patient 1 à J2, patient 1 à J3).

## Nom des variables

Les noms de variables (nom des colonnes) doivent respecter les règles suivantes :

- tout en minuscule
- uniquement des lettres des chiffres et des 'underscores' « \_ » (tiret du 8) *donc pas d'espace ni d'accent*
- commence par une lettre
- court et autant explicite que possible (*parfois impossible donc une description sera nécessaire dans le dictionnaire des variables*)

## Contenue des variables

La base de données n'est pas un outil de communication. Seules des valeurs pour des variables y sont présentes. S'il y a nécessité d'ajouter des précisions celles-ci doivent être fournies à part.

Tout manquement aux règles ci-dessous peuvent entraîner des pertes de données et des erreurs d'analyses.

### Variable numérique

Chaque case doit contenir un nombre, un seul nombre et rien d'autre que ce nombre. Si un caractère autre qu'un chiffre, un point ou une virgule est présent cela entraîne une perte de donnée lors de l'import dans le logiciel statistique. Le choix du point ou de la virgule pour les décimales m'importe peu tant que ce choix est RESTE LE MÊME tout au long du projet. Il est cependant préférable d'**utiliser un point** comme dans les articles scientifiques qui présentent tous les résultats numériques avec un point comme séparateur entre unités et décimales.

### Variable binaire (cas particulier d'une variable qualitative à deux modalités)

Ne contient que TRUE ou FALSE. (Voir des F et des M pour masculin féminin, des oui/non. Etc.). MAIS PAS DE 0 ou de 1.

### Variable qualitative

Pour chaque variable seules les valeurs mentionnées dans le dictionnaire des variables sont autorisées. Le texte libre n'est pas exploitable statistiquement. Lors de l'import des données toutes les valeurs qui ne sont pas autorisées seront considérées comme manquantes (attention aux fautes de frappes ACCENTS et AUX ESPACES en début ou fin de texte).

Les modalités (catégories) en elles-mêmes respectent les mêmes règles que les noms de variables. (*Même quand la catégorie est un chiffre il est préférable de la préfixer avec une lettre. Par exemple pour des paliers d'antalgique on mettra « p1 », « p2 » ou « p3 ». Ainsi le logiciel de statistique ne pense pas qu'il s'agit d'une variable quantitative*).

Dans le cas d'une variable à modalités mutuellement exclusives, une variable ne peut pas ~~contenir~~ plusieurs valeurs dans une case. Si on a le cas d'une variable 'couleur' qui peut ~~prendre~~ comme valeurs 'bleu', 'jaune' ou 'rouge', on ne peut pas avoir une personne 'jaune + rouge'.

---

Si pourtant cela représente une réalité, (mutuellement non exclusif) alors il faut créer trois variables (trois colonnes) binaires : 'bleu', 'jaune' et 'rouge'. L'individu jaune et rouge aura alors FALSE, TRUE, TRUE comme valeur respectives pour ces trois variables. (Ce n'est pas obligatoire tout dépend du sens que cela apporte.)

Le plus important est le choix du nom des modalités de la variable qui doivent rester inchangées ainsi on aura agriculteur à chaque fois pour les professions et pas Agriculteur, et agricultrice et agri.

### Variable de date

Une seule date par case. Respecter toujours le même format. Je préconise dd/mm/yyyy (ex : 23/10/1994).

Si vous utilisez excel modifier le format de la colonne pour choisir « texte » ainsi excel ne modifie plus de lui-même la manière dont vous avez entré les données.

### Données manquantes

Lorsque des données sont manquantes laisser la case du tableau vide. (surtout pas 0).

## Conclusion

Ces règles peuvent paraître contraignantes mais nous permettrons de travailler efficacement. Elles évitent que l'analyse s'éternise et assure des résultats fiables.

En cas de soucis pour vous plier à ces règles ou les comprendre n'hésitez pas à contacter votre statisticien préféré que se fera un plaisir de vous aider 😊. Trouver des noms de variables peut aussi être complexe et une fois de plus il vous apportera son expertise avec joie.

Auteur et co-auteur Adrien Le Guillou, Thibaut Pressat

## Parlez de jisstat sur les réseaux





FACEBOOK JISSTAT

jisstat- **en cours de construction: arrière plan HS, logo en vacance, nombre de pages... plusieurs.**

Tous droits réservés ® 2018 jisstat

[Mentions légales](#) | [Politique de confidentialité](#) | [Politique des cookies](#) | [Plan du site](#)

[Connexion](#)

**Ce site a été conçu avec Jimdo.** Inscrivez-vous gratuitement sur <https://fr.jimdo.com>