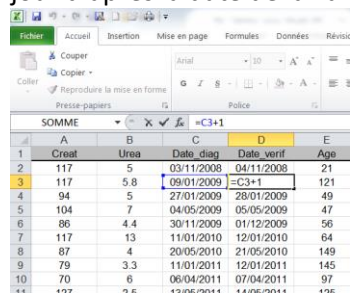


## Comment structurer un fichier de données Excel avant analyses statistiques

Voici les points que vous devez respecter pour analyser de façon rigoureuse un fichier de données. N'hésitez pas à imprimer cette feuille et cocher, manuellement, chacune des cases !...

- ☐ La première ligne, et la première ligne seulement, doit contenir le nom des variables.
- ☐ Il ne doit y avoir aucune ligne (vide ou pas) au-dessus de l'unique ligne contenant le nom des variables.
- ☐ Aucune ligne vide entre deux lignes remplies
- ☐ Aucune colonne vide à gauche de la première colonne contenant des informations.
- ☐ Chaque colonne contenant des informations doit avoir un nom, placé sur la première ligne (il s'agit du nom de la variable).
- ☐ Le nom des variables, d'une longueur maximale de 30 caractères, doit contenir exclusivement l'un des 37 caractères suivants : les 26 lettres de l'alphabet sans accent, les 10 chiffres (sans que le nom d'une variable ne commence par un chiffre), et le caractère « \_ ». Ainsi, l'espace et tout signe de ponctuation (point, virgule, ...) sont par conséquent proscrits. Voici un exemple de nom de variable : `Frequence_cardiaque_elevee`.
- ☐ Une ligne par animal, toutes les informations recueillies sur cet animal étant mises en colonnes (oui, il peut y avoir plus de 100 colonnes, mais ce n'est pas grave...).
- ☐ Les colonnes relatives aux variables qui seront utilisées dans les analyses stat' doivent uniquement contenir des chiffres ou des cases purement vides.
- ☐ Vos variables binaires devront être codées en 0/1. Le nom de la variable binaire devra correspondre à ce que représente votre « 1 ». Par exemple, si votre variable concerne le sexe de l'animal, et si vous souhaitez coder vos animaux avec un « 0 » pour les mâles, et un « 1 » pour les femelles (c'est vous qui choisissez votre codage), vous nommerez cette variable « Femelle ».
- ☐ Les colonnes relatives aux variables qui seront utilisées dans les analyses stat' ne doivent pas contenir des espaces, des « ? », ou autre signe alphanumérique. Pour vérifier ce point-là excessivement important, utiliser la fonction filtre d'Excel (<https://youtu.be/g1mlndB9wg>).
- ☐ Si une donnée est manquante pour une variable qui sera utilisée dans les analyses stat', laisser la case vide (absolument vide, c'est-à-dire sans espace, ni quoi que ce soit d'autre).
- ☐ Si une variable relative à un marqueur biologique contient une valeur indéterminée car supérieure ou inférieure à un seuil (par exemple « > 12 »), supprimez cette information pour avoir une donnée manquante (ou à la limite, rendez cette information numérique en affectant un nombre plausible – par exemple « 15 » pour toutes les valeurs « > 12 »).
- ☐ Copier toutes vos colonnes qui vont servir aux analyses stat', et uniquement celles-là, collez-les dans une nouvelle feuille Excel, faite une « Recherche » du caractère espace : vous ne devrez trouver *aucune* cellule contenant un espace.
- ☐ Si vous avez une variable relative à une date (par exemple, variable `Date_diag`), pour être sûr que cette variable contienne bien des dates (pour notamment calculer par la suite des délais), effectuer la vérification suivante (cf. figure ci-dessous) : (1) insérer une colonne vide juste après celle de la variable `Date_diag` (dans la figure ci-dessous, j'ai nommé « `Date_verif` » cette colonne), (2) tapez « `= [cellule juste à gauche en cliquant dessus]+1` », et (3) vous devez obtenir une date du jour d'après la date de la variable `Date_diag`.



	A	B	C	D	E
1	Crest	Unsa	Date_diag	Date_verif	Age
2	117	5	03/11/2008	04/11/2008	21
3	117	5.8	09/01/2009	=C3+1	121
4	94	5	27/01/2009	28/01/2009	49
5	104	7	04/05/2009	05/05/2009	47
6	86	4.4	30/11/2009	01/12/2009	56
7	117	13	11/01/2010	12/01/2010	64
8	87	4	20/05/2010	21/05/2010	149
9	79	3.3	11/01/2011	12/01/2011	145
10	70	6	06/04/2011	07/04/2011	97
11	127	2.5	13/05/2011	14/05/2011	125