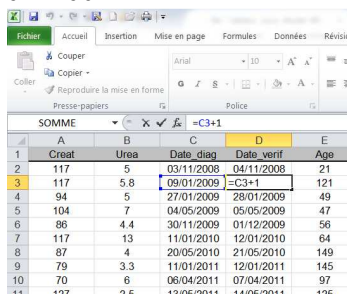


Comment structurer un fichier de données Excel avant analyses statistiques

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.
- Aucune ligne (vide ou pas) au-dessus de l'unique ligne contenant le nom des variables.
- 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.
- Le nom des variables doit contenir exclusivement l'un des 37 caractères suivants : les 26 lettres de l'alphabet, les 10 chiffres (sans que le nom d'une variable ne commence par un chiffre), et le caractère « _ ». L'espace est par conséquent proscrit. Voici un bon exemple de nom de variable : Frequence_cardiaque_elevee_ON pour une variable qui indique si l'animal a ou non une fréquence cardiaque plus élevée qu'un seuil pré-établi.
- 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.
- 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/gl1mIIndB9wg>).
- 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, ou 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 »), rendez cette information numérique si vous voulez faire des analyses statistiques sur cette variable (affectez-y un nombre ou bien supprimez l'information pour avoir une donnée manquante).
- 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 devez trouver *aucune* cellule contenant un espace.
- Si vous avez des variables relatives à des dates, pour être sûr que ce sont bien des dates (pour notamment calculer par la suite des délais), effectuer la vérification suivante : (1) insérer une colonne vide juste après celle de la variable DATE, (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. Voici un exemple ci-dessous où la variable Date_verif est la variable que vous allez créer pour vérifier que la variable Date_diag est correctement définie. Vous devrez ensuite vérifier que la variable Date_verif sera la date du jour d'après la variable Date_diag pour tous les animaux.



	A	B	C	D	E
1	Creat	Urea	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	08/04/2011	07/04/2011	97
11	197	2.5	13/05/2011	14/05/2011	195