

**Critères à respecter pour les banques de données Excel**

**1. Versions**

- a. Il est préférable d'utiliser la dernière version disponible de MS Excel.
- b. Il est préférable d'utiliser la suite Office pour PC plutôt que Mac.

**2. Appellation**

**a. Fichier**

- i. Le nom doit être court.
- ii. Terminer le nom par la date ou un autre identifiant (ex. : 20150610 ou v1) afin de pouvoir distinguer chaque version du fichier (et en modifier le nom à chaque modification apportée aux données).

**b. Feuilles**

- i. Le nom doit être court (20 caractères ou moins).

**c. Variables**

- i. Le nom doit être court (moins de 32 caractères). Il est recommandé d'utiliser des noms plus courts d'environ 15 caractères.
- ii. Le premier caractère doit être une lettre ou un *underscore*.
- iii. Ne pas mettre d'accents (é, è, à, â, î, û, etc.), d'espaces ou de symboles (tels que %, ?, &, ), (+, :, :)), @, ", etc.)
- iv. Le seul caractère accepté dans le nom est le *underscore*, qui peut être utilisé pour diviser en 2 mots le nom de variable, comme « Date\_chx ».
- v. Le nom entier doit être contenu dans une seule cellule.
- vi. La banque ne doit pas contenir 2 variables portant le même nom (PH\_USI = ph\_usi), si elles ne contiennent pas la même information (donner des noms spécifiques si des mesures sont semblables).

**3. Structure du fichier**

- a. Idéalement, tout mettre les données dans un même fichier ou même feuille.
- b. Si des données se trouvent dans plusieurs feuilles ou fichiers, il doit toujours y avoir une variable d'identification des participants afin de pouvoir faire le lien.
- c. Si des données se trouvent dans plusieurs feuilles ou fichiers, donner le même nom à une même variable apparaissant dans chaque feuille ou fichier.
- d. Si des données se trouvent dans plusieurs feuilles ou fichiers, utiliser les mêmes termes pour désigner les valeurs des variables.

**4. Structure d'une feuille**

- a. Une ligne par observation (par exemple, le patient) et une colonne par variable.
- b. Ne pas laisser de ligne vide entre des observations.
- c. Ne pas fusionner ou fractionner les cellules.
- d. Bien délimiter la zone de données pertinentes.

### **Saisie de variables numériques**

- e. Écrire les nombres en format numérique et non en caractère.
- f. Ne pas mettre de symboles (unités) dans les valeurs des variables continues (ex.: kg, m)
- g. Ne pas écrire de texte supplémentaire, détails ou commentaires dans la cellule.
- h. Écrire les dates en format date (définies dans Excel dans « Format de cellule ») et toutes dans le même format.
- i. En présence de données manquantes, laisser la cellule vide.
- j. En présence de données non applicables, laisser également la cellule vide et préciser dans le dictionnaire de codification, les conditions particulières d'applicabilité pour les variables concernées.
- k. Toujours utiliser le même symbole pour séparer les décimales (utiliser la décimale par défaut dans votre chiffrier, soit la virgule ou le point).

### **5. Saisie de variables alphanumériques**

- a. Toujours utiliser la même orthographe pour désigner une catégorie (ex : Homme ≠ Masculin ≠ H ≠ h ≠ M ≠ m, Chirurgie pédiatrie ≠ Chirurgie pédiatrique, Yes ≠ y ≠ O).
- b. Éviter idéalement les lettres accentuées.
- c. En présence de données manquantes, laisser la cellule vide.
- d. En présence de données non applicables, laisser également la cellule vide. Vous pouvez préciser dans le dictionnaire de codification, les conditions particulières d'applicabilité pour les variables concernées.
- e. Pour la codification de variables dichotomiques ayant un ordre (Exemple : Oui/Non; Grand/Petit; Présence/ Absence; etc.) il est préférable de codifier :
  - i. 0 pour définir la petite valeur : Non; Petit; Absence.
  - ii. 1 pour définir la grande valeur : Oui; Grand; Présence.
- f. Pour la codification de variables dichotomiques nominales (Exemple : Homme/Femme; Blanc/Noir; Groupe Expérimental/Groupe Témoin; Traitement Expérimental/ Traitement Standard; etc.) il est préférable de codifier :
  - i. 0 pour définir la modalité de référence: Groupe Témoin; Traitement Standard.
  - ii. 1 pour définir la valeur d'intérêt : Groupe Expérimental; Traitement Expérimental.
- g. Pour la codification de variables génétiques, il est préférable d'utiliser la codification suivante selon le type de modèle :
  - i. Modèle additif en 3 catégories : 0 pour récessif homozygote, 1 hétérozygote, 2 pour dominant homozygote.
  - ii. Modèle dominant en 2 catégories : 0 pour récessif homozygote, 1 pour hétérozygote et dominant homozygote;
  - iii. Modèle récessif en 2 catégories : 0 pour récessif homozygote et pour hétérozygote, 1 pour dominant homozygote.
- h. Si des nombres sont utilisés pour représenter des catégories, fournir un dictionnaire de codification.

6. **Vous êtes libres de :**

- a. Ordonner les variables dans l'ordre que vous préférez.
- b. Mettre différents symboles de couleur pour vous retrouver. Attention : nous ne pouvons utiliser les couleurs dans les analyses. Créer une variable supplémentaire si nécessaire.
- c. Saisir des titres pour regrouper des catégories de variables ou mettre de longues étiquettes, dans les lignes au-dessus du nom des variables. Dans l'exemple ci-dessous, les données peuvent être lues seulement à partir de la ligne 3 où les noms des variables sont définis, donc le contenu des lignes 1 et 2 est libre.

	A	B	C	D	E	F	G	H
1		Variables sociodémographiques			Antécédents médicaux			Outcomes
2	Numéro d'ide	Sexe	Âge	Scolarité	Diabète	Hypertensior	Date diagnostic diabète	diabète
3	Id	sex	age	scol	diab	htn	diab_date	deces
4	1	F	52	Primaire	1	0	2015-06-10	0
5	2	F	46	Cegep	0	1		1
6	3	H	85	Maitrise	0	1		0
7	4	H	14	Cegep	0	0		0
8	5	F	36	Secondaire	0	0		1
9								

« **NB : Pour respecter les normes éthiques et les bonnes pratiques cliniques, il est important** de nous faire parvenir une banque de données dans laquelle les noms des patients, leurs coordonnées, leur numéro de dossier du CHU, d'assurance maladie ou d'assurance sociale **n'apparaissent pas**. Un numéro d'identification propre à l'étude **doit** être attribué à chacun des patients pour les distinguer. **(Les données confidentielles : nom, coordonnées et autres numéros, incluant le lien avec le numéro d'identification dans la base de données, doivent être conservées dans un autre fichier auquel vous seul avez accès).** »

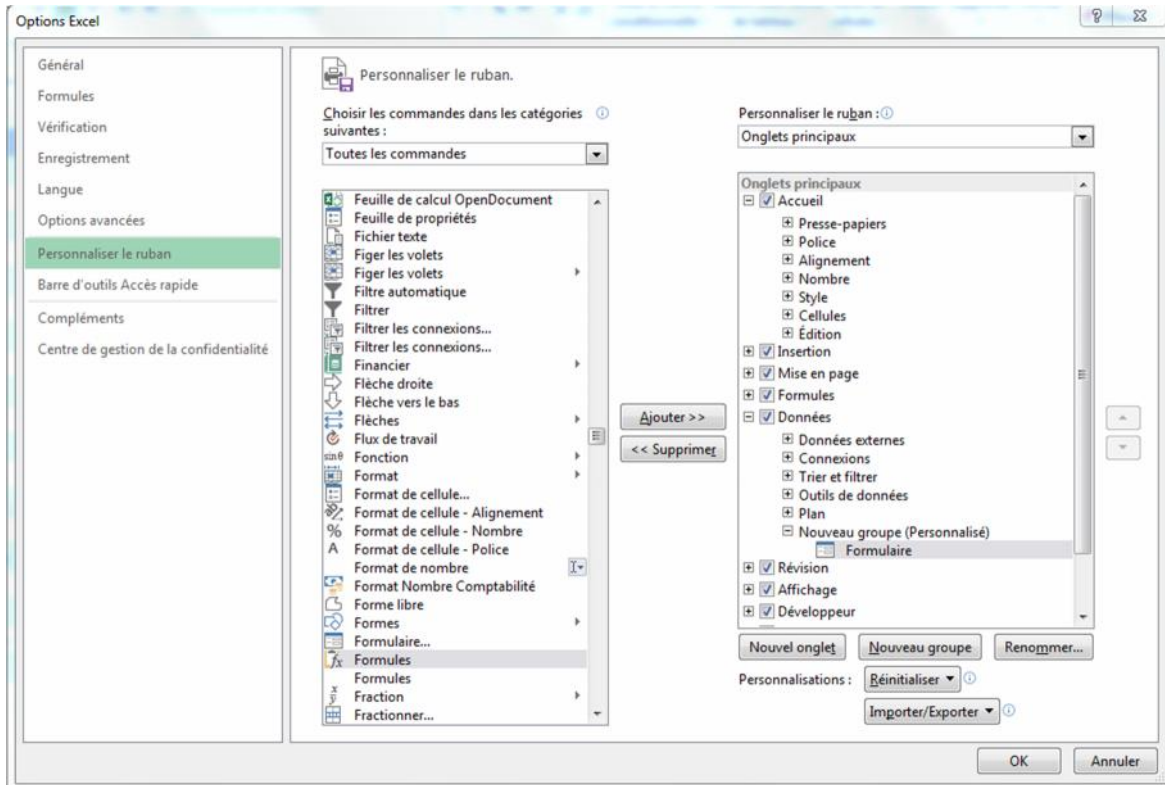
ANNEXE :

**Trucs et astuces dans Excel**

Afin de vous faciliter la vie lors de la saisie, et d'éviter les problèmes lors de l'importation pour les analyses statistiques, il est possible de :

1. Figurer des volets afin d'éviter de recopier le nom des variables à chaque 20 lignes ou le numéro d'identification des patients à chaque 10 colonnes afin que vous les gardiez en vue.
  - a. Affichage
  - b. Figurer les volets
  - c. Figurer la ligne supérieure
  - d. Figurer la première colonne
  
2. Générer des menus déroulants pour les variables alphanumériques.
  - a. Dans une feuille que vous nommerez « Menu », spécifier dans la première ligne le nom des variables visées, puis sous chacune les choix de valeurs possibles.
  - b. Sélectionner l'ensemble des choix de réponses d'une des variables, cliquer sur le champ « Nom » à gauche de la barre de formule et y inscrire le nom de la variable.
  - c. Dans la feuille de saisie des données, cliquer dans la première cellule pour les observations d'une variable. Dans l'onglet Données, cliquer Validation des données, puis Options.
  - d. Dans Autoriser, sélectionner Liste, puis dans source, indiquer « = nomdelavariablenom ». Un menu déroulant apparaîtra sur la cellule sélectionnée.
  - e. Cliquer sur le coin inférieur droit de cette cellule en la glissant vers le bas pour en copier la définition pour les observations suivantes.
  - f. Si vous saisissez autre chose que les choix du menu dans une cellule, un message d'erreur apparaîtra. En allant dans Validation des données, Alerte d'erreur, il est possible de permettre la saisie de données non valide en générant un avertissement, au lieu de l'empêcher complètement.
  
3. Implanter des conditions de saisie pour les variables numériques.
  - a. Dans la feuille de saisie des données, cliquer dans la première cellule pour les observations d'une variable. Dans l'onglet Données, cliquer Validation des données, puis Options.

- b. Par exemple, pour l'âge, Autoriser un nombre entier, avec des données comprises entre 0 et 130. Un message d'erreur apparaîtra si un nombre négatif, des décimales, un âge impossible ou du texte est saisi.
  - c. Par exemple, pour l'IMC, Autoriser Décimal, avec des données supérieures à 0. Un message d'erreur apparaîtra si un nombre négatif ou du texte est saisi.
  - d. Par exemple, pour la date de naissance, Autoriser Date, avec données inférieures à la date de fin maximale possible, comme le 10 juin 2015 en date d'aujourd'hui. Si une date postérieure ou du texte est saisi, un message d'erreur apparaîtra. Attention : Si un chiffre est saisi, il sera transformé en date à partir du nombre de jours depuis le 1 janvier 1900 : ex (15 -> 15 janvier 1900).
  - e. Par exemple, pour le site hospitalier, Autoriser Longueur du texte inférieur à 5. Si le site est saisi au long au lieu de CHUL, HSFA, HDQ, HEJ, HSS, un message d'erreur apparaîtra. Il est toutefois préférable d'utiliser un menu déroulant dans ce cas pour ne pas faire d'erreur de syntaxe dans les abréviations.
4. Créer un formulaire de base Excel pour la saisie des données : Pour afficher le formulaire, nous avons besoin d'un bouton qui, dans la version 2013, n'apparaît pas par défaut. Pour le faire apparaître, allez dans l'onglet **Fichier > Options > Personnaliser le ruban**. Dans la partie droite, sélectionnez l'un de vos onglets et cliquez sur le bouton **Nouveau groupe**, puis, dans la partie gauche, dans le menu déroulant, choisissez *Toutes les commandes* et recherchez *Formulaires...*, cliquez sur **Ajouter >>**.

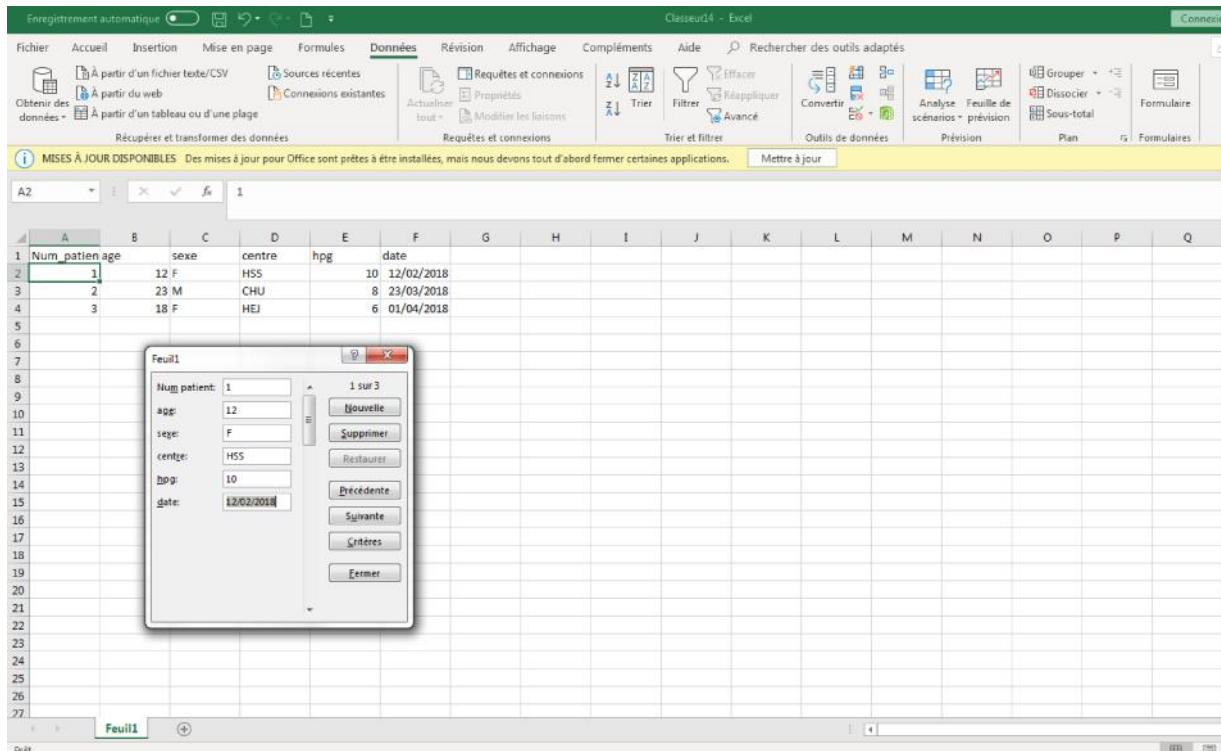


Source : <https://www.excel-plus.fr/vba/excodes/creer-formulaire-personnalise/>

Désormais, dans l'onglet que vous avez sélectionné pour votre nouveau bouton, vous avez une nouvelle icône :

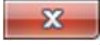


Pour utiliser le formulaire classique, cliquez sur ce bouton **Formulaire**. Il prend en charge automatiquement les variables inscrites dans la première ligne de la feuille Excel.



**Attention :** Si vous cliquez dessus mais qu'aucune variable n'existe dans votre fichier, le formulaire n'apparaîtra pas. Aussi, si vous avez trop de variables (plus de 32), vous ne pouvez pas utiliser le formulaire de base.

Vous obtenez un formulaire avec toutes vos variables qui sont reprises sous forme de champ de saisie. Si vous avez déjà des observations, elles s'affichent sous forme d'enregistrements que vous pouvez parcourir dans votre formulaire et compléter. À droite des champs de saisie, on a :

- Le **nombre de fiches** et votre position ;
- **Nouvelle** : vous ajouterez un nouvel enregistrement ;
- **Supprimer** : vous supprimerez la fiche en cours ;
- **Restaurer** : vous restaurerez les modifications effectuées sur un enregistrement ;
- **Précédente** : vous vous déplacerez vers la fiche précédente ;
- **Suivante** : vous vous déplacerez vers la fiche suivante ;
- **Critères** : vous permettra de faire une recherche. Saisissez votre élément à rechercher dans le champ correspondant et appuyez sur *Entrée*. Cliquez sur **Grille** pour rebasculer en mode normal si vous n'avez fait aucune recherche ;
- **Fermer (ou la croix rouge  )** : vous fermerez le formulaire.