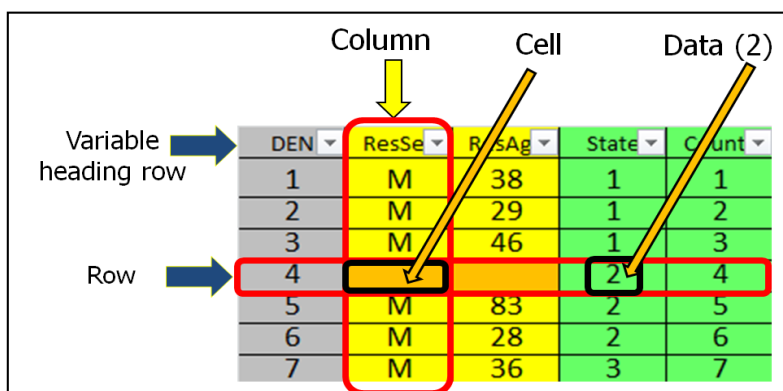


## Normes FAO

### Etude sur la sécurité semencière - CRÉATION DE BASE DE DONNÉES MS EXCEL

Quand vous ouvrez Microsoft Excel, un nouveau fichier (Classeur1) apparaît sur votre écran. Ce fichier comporte normalement trois feuilles de calcul (une nouvelle feuille de travail peut toujours être ajoutée). Chaque feuille de calcul comporte des **Colonnes** et des **lignes** et les intersections entre les colonnes et les rangées sont des Cellules (Figure 10). Les cellules sont des points dans une feuille où les données (des variables) sont saisies. Les questionnaires contiennent normalement un certain nombre de variables avec les volumes des données qui sont saisies dans des cellules sous chaque colonne de la variable.

Figure 1: Feuille de calcul MS Excel



Variable heading row	DEN	ResSe	ResAg	State	Count
1		M	38	1	1
2		M	29	1	2
3		M	46	1	3
4		M	83	2	4
5		M	83	2	5
6		M	28	2	6
7		M	36	3	7

Dans la conception d'une base de données, la première ligne est normalement consacrée aux titres des variables, où chaque variable est définie dans une colonne et les cellules en-dessous du titre sont utilisées pour la saisie des données correspondant au titre de la variable. D'autre part, une colonne (normalement la première) est utilisée pour définir le numéro de saisie de données. En règle générale;

- Colonne** - les Variables sont normalement codées comme titres de colonnes dans la rangée des titres des variables. Par exemple, une colonne est normalement utilisée pour une variable. Genre, Âge, Revenu etc.
- Ligne(s)** - Toutes les lignes au-dessous des titres des variables peuvent être utilisées pour saisir des données. Les informations d'un seul questionnaire sont entrées dans la (es) ligne(s) correspondant au numéro de saisie des données du questionnaire (NSD ou DEN en anglais). Un questionnaire ne devrait jamais avoir plus qu'un DEN, et deux ou plusieurs questionnaires ne devraient pas avoir le même DEN. Une ou deux lignes au-dessus de la ligne des titres des variables pourraient être utilisées pour entrer les numéros des questions correspondant aux variables (voir la figure 11).
- Une cellule**, une réponse

Figure 2: Tableau MS Excel avec filtres

	Qn.1.1	Qn.1.2	Qn.2.1	Qn.2.2	Qn.2.3	3.1	3.2
DEN	County	Sub-coun	Age.grou	Res.Gend	Educatio	HH_Gend	HH_Size
1	Kitui	Kitui Central	2	m	3	m	3
2	Kitui	Kitui Central	3	f	2	m	7
3	Kitui	Kitui Central	3	f	2	f	6
4	Kitui	Katulani	3	f	2	f	5
5	Kitui	Katulani	1	f	3	m	8
6	Kitui	Katulani	4	f	1	f	3
7	Makueni	Mbooni	4	m	1	m	13
8	Makueni	Mbooni	2	f	3	m	7
9	Makueni	Mbooni	3	m	3	m	4
10	Makueni	Mbooni	3	m	4	m	5
11	Makueni	Kibwezi	3	f	2	m	5
12	Makueni	Kibwezi	4	f	1	f	5

### Définition des variables et codage des réponses

Comme susmentionné, les variables d'un questionnaire sont définies dans la ligne des titres des variables. Quand on définit les titres des variables, seuls les caractères continus sont utilisés c'est-à-dire pas d'espace entre les caractères. Pour une simple réponse telle que le sexe du chef de ménage (CM ou HH, en anglais), la variable urrait être définie comme étant **HH.Gender** ou **Gender\_HH** mais pas HH Gender.

Pour des questions à réponses multiples telles que - Quelles cultures avez-vous semées la saison dernière? Ici, chaque réponse possible (par exemple, le sorgho, le maïs, les haricots) est une variable - et donc occupe une colonne - dans laquelle une valeur est saisie, dans ce cas, oui (1) ou Non (0).

Les codes sont parfois donnés dans les réponses à un questionnaire; par exemple la raison pour semer moins de superficie de la terre - 1 = manque d'accès à la terre; 2 =accès à plus de terre; 3 = manque de main d'oeuvre -- et d'autres (à spécifier) dans le questionnaire. Les codes doivent être donnés à ces autres éléments également.

**Les commis de saisie des données ESS sont fortement conseillés de faire usage de la structure de base de données qui a été conçue par l'équipe de développement ESS FAO.**

### Validation des cellules et saisie de données

Pour minimiser les erreurs lors de la saisie de données, certaines colonnes des variables doivent être validées pour restreindre l'entrée des données imprévues ou certains caractères et / ou intervalles de valeur.

Avant toute saisie de données, un numéro est attribué à un questionnaire et deux questionnaires ne devront pas avoir le même numéro. L'existence d'un numéro de questionnaire créé un lien univoque entre la copie électronique et la copie papier et sera utile dans la phase de nettoyage des données.

Les différentes variables ont différents types de données selon le type de variables.

- Variables discontinues (catégorielles, classificatoires, discrètes): les variables qui ne peuvent être divisées en fractions ou prennent des nombres finis, par exemple sexe (masculin ou féminin), présence de l'élevage (oui ou non); statut de résidence (résidents, déplacés, réfugiés, rapatriés). Ces variables peuvent être représentées par du texte, par exemple les initiales (H

pour les hommes et F pour les femmes) ou des codes numériques tels que 1 = Oui, 0 = Non; 1 = résidents, 2 = Déplacés, 3 = réfugiés et, 4 = rapatriés

- b) Variables continues - variables qui peuvent être divisées en fractions ou prennent un nombre infini de valeurs, comme par exemple le revenu, la température, l'âge, la superficie cultivée, les quantités de semences, la production, les rendements. Ces variables sont saisies sans unité de mesure, par exemple pour 10 kg de semences plantées, la quantité de semences plantées est normalement saisie comme 10 et non 10 kg.
- c) Remarque à l'intention de l'agent de saisie de données – les unités de mesure des variables continues peuvent varier d'un individu ou d'un endroit à un autre. Ceci doit être normalisé avant ou pendant la saisie des données.

Afin d'assurer la saisie des données de qualité dans la base de données, le chef d'équipe doit fournir des efforts supplémentaires pour superviser les agents de saisie des données. En contrôlant la saisie, le chef d'équipe devrait procéder à l'échantillonnage aléatoire des questionnaires qui ont été saisis par l'agent de saisie de données et à les vérifier à l'aide des numéros de saisie dans la base de données pour vérifier si les données ont été saisies correctement. Le contrôle de la saisie des données se fait normalement en fin de chaque journée pendant le processus de saisie de données. Ce processus devrait être fait en collaboration avec le personnel de saisie de données afin qu'il voie les erreurs commises et prendre des mesures pour réduire les erreurs.

#### **Erreurs de saisie de données et comment les corriger**

Les erreurs de saisie ou de codification doivent être corrigées immédiatement selon les informations contenues dans le questionnaire. Ces erreurs sont plus fréquentes pendant les premiers jours ou lorsque le travail est fait à la hâte, mais devront être réduites avec le temps. Si une plus grande fréquence de ce genre d'erreurs est enregistrée, un plus grand nombre de questionnaires devront être vérifiés.

Une autre erreur de saisie de données commune est un changement des colonnes où les données sont saisies, dans ce cas, soit une colonne a été sautée soit elle est entrée trop tôt. Dans ce genre de cas, l'ensemble du questionnaire doit être saisi à nouveau. Une façon de minimiser la colonne liée à l'erreur de saisie de données est d'utiliser des couleurs différentes pour différentes colonnes selon la section correspondante du questionnaire. Ceci permet à l'agent de saisie de données de relier la position dans le questionnaire à la position dans la base de données.

L'importance de la précision dans la saisie des données ne peut pas être surestimée, car le processus de correction peut prendre beaucoup de temps et de travail.

#### **Nettoyage des données et vérifications**

Les erreurs peuvent être introduites au cours de la collecte des données ainsi qu'à la saisie. Avant de tirer une variable supplémentaire ou faire l'analyse des données, le gestionnaire de données doit s'assurer que les données sont dépourvues d'erreurs ou de valeurs aberrantes. Dans la vérification et le nettoyage des données, toutes les variables sont vérifiées pour s'assurer qu'il n'y ait pas d'erreur, de données incohérentes ou aberrantes. L'Encadré 7 donne un exemple qui pourrait être soit une erreur ou une valeur aberrante. Toutes les entrées ou valeurs aberrantes doivent être vérifiées en contrôlant la copie dure du questionnaire et une décision appropriée ou une correction devra être prise sur les valeurs aberrantes.

#### Encadré 1: Exemple d'une erreur ou d'une valeur aberrantes

L'agent de saisie de données a entré la quantité suivante (kg) de semences de sésame cultivées par 10 ménages; 2, 3.5, 45, 2, 4, 1, 1.5, 2, 5.5 et 2.3. Il y a deux possibilités pour ces données.

- a) Une erreur introduite par l'agent de saisie pendant la saisie quand il appuie sur la touche 4 & 5 simultanément quand il voulait appuyer sur 4 ou 5, ou ne parvient pas à bien appuyer un point décimal en saisissant 4.5. Cela peut être corrigé en vérifiant la copie dure du questionnaire et en corrigeant la saisie.
- b) Une donnée aberrante – Si après vérification du questionnaire, le nombre 45 s'avère une saisie correcte, une confirmation peut être faite en appelant l'enquêteur (s'il se rappelle), les agriculteurs (si le contact téléphonique est disponible). Si après vérification à partir de toutes ces sources le nombre (45) s'avère correct, cela peut être une donnée aberrante qui pourrait influencer significativement le résultat de l'analyse.
- c) Dans l'ensemble des données ci-dessus, quand l'analyse de la quantité moyenne de semences cultivées s'effectue avec une telle valeur aberrante, la moyenne est 6.9, et lorsqu'une telle valeur aberrante est omise de la série des données, la moyenne est 2.6.

Remarque: Les données aberrantes pourraient être des données correctes mais déviantes de la distribution normale. Statistiquement, elles ont une incidence significative sur le résultat d'analyse.

Le nettoyage des données est un processus fastidieux qui exige de la patience et du temps, mais il ne doit jamais être ignoré. Il y a deux niveaux ou moyens différents d'assurer l'exactitude des données saisies et obtenues. Le **premier niveau** est le filtrage des données en utilisant la fonction de filtrage Excel.

L'activation du filtre et le nettoyage des données impliquent les mesures simples mais logiques suivantes, après que toutes les données ont été saisies.

- a) Sélectionner tous les titres des variables
- b) Aller au menu **Données** et cliquez sur l'icône Trier & Filtrer. Le Menu déroulant apparaîtra sur le côté droit de chaque titre de variable.
- c) Cliquer sur l'icône du menu déroulant et scannez pour toutes les données aberrantes et incohérentes de la liste que vous voyez.
- d) Une fois les données aberrantes ou incohérentes identifiées, d'abord **désélectionnez tout**, puis **sélectionnez** les données incohérentes ou aberrantes. Cliquez sur **OK**. Seules les valeurs sélectionnées apparaîtront sur l'écran.
- e) Vérifiez le **numéro de saisie de données(s)** correspondant à des données incohérentes ou aberrantes identifiées, revenir sur la **copie dure** du questionnaire et corriger.
- f) S'il y a de valeurs incohérentes ou aberrantes sur la copie dure, veuillez consulter l'enquêteur ou le chef d'équipe pour la correction.
- g) Dans le cas où ni la copie dure ni le chef d'équipe / enquêteurs peuvent aider, le gestionnaire de données devra porter un jugement **d'omettre (Supprimer)** les données aberrantes incohérentes si elles auront une incidence sur l'analyse finale.

Le **deuxième niveau** est de vérifier la cohérence dans le lien entre deux variables connexes, par exemple:

- 1 Un ménage ne peut pas avoir plus de surfaces cultivées avec les différentes cultures au cours d'une saison spécifique que la surface totale disponible pour la même saison;
- 2 Un ménage n'a pas d'épargne en espèces effectué au cours d'une saison, mais la variable correspondante a indiqué l'institution où l'argent a été épargné.

Ces erreurs peuvent être évitées en programmant des paramètres de sécurité dans la base de données, de sorte qu'il serait impossible de saisir des données qui ne sont pas cohérentes. A défaut, le processus de correction le plus appropriée est de vérifier à l'aide du questionnaire correspondant ou par déduction logique.

### **Dérivation des variables**

Certaines variables telles que les rendements, les taux de semis, les taux de multiplication, les unités animales ne sont pas généralement collectées directement à l'aide des questionnaires, mais sont dérivées à partir de deux ou plusieurs variables. Par exemple, le rendement qui est la quantité récoltée par unité de surface arable est dérivé de la quantité récoltée à partir d'une zone donnée plantée par les agriculteurs. Pour toute variable dérivée supplémentaire, une colonne doit être insérée et le calcul approprié doit être fait. La façon la plus appropriée et facile pour ce processus est l'utilisation des formules. Une fois les calculs effectués, vous devez vérifier les résultats, car parfois les résultats ne sont pas corrects, lorsque les données sont manquantes ou un nombre est divisé par zéro, ce qui est représenté par le symbole suivant: **#DIV/0!** dans la base de données. Ces entrées doivent être supprimées avant de procéder à l'analyse des données.

## **EXEMPLE DE DIRECTIVES POUR LE PERSONNEL DE SAISIE DE DONNEES - NEPAL**

### **Les choses à retenir pour le personnel de saisie de données**

1. Vous devez bien comprendre que le succès et la fiabilité de cette étude dépend de la qualité de votre travail en tant qu'agent de saisie de données. Vous effectuez un travail essentiel de cette étude, donc vous devriez le faire sérieusement et honnêtement.
2. Vous devriez avoir une bonne connaissance de la structure du questionnaire, les codes pour les réponses pour la saisie de données et la base de données elle-même. Vous avez reçu une formation sur ces aspects afin que vous soyez à l'aise et efficace.
3. La base de données est établie sur la feuille de calcul Microsoft Excel ; vous devez donc avoir une bonne connaissance d'Excel et vous devez savoir comment travailler avec Excel avant de commencer votre travail.
4. Une formation sur les aspects susmentionnés vous est offerte, ce qui est une condition préalable pour votre travail de saisie de données. Si vous êtes confus même après une formation, vous devriez communiquer avec vos superviseurs afin d'obtenir des précisions.
5. Vous devriez saisir les réponses du questionnaire dans la base de données suivant les codes fournis et des instructions spécifiques.
6. Pendant la saisie de données, vous devrez connaître la nature de chaque question, que ce soit une question d'option unique ou une question à réponses multiples. Les questions à option restreinte ou à réponses multiples sont mentionnées dans le questionnaire. Dépendant du type de réponse il y a soit une colonne dans la base de données pour la question de l'option unique ou plus d'une pour l'option à multiples questions. Chaque réponse doit être saisie dans une colonne différente de la base de données.
7. Il y a des questions pour lesquelles vous devez prévoir la saisie des données quantitatives. Les enquêteurs pourront recueillir les informations dans l'unité de mesure spécifiée (locale ou standard). Donc, vous devriez être très prudent sur les unités mentionnées dans le questionnaire. S'il est mentionné en unités locales, alors vous devriez les convertir en unités standard, le tableau de conversion est fourni pour vous. Les unités standard sont le kg (poids) et les hectares (pour les zones / superficie).
8. S'il y a confusion dans le questionnaire, vous devez noter ces confusions et informer votre superviseur. Il vous aidera à clarifier les réponses ou à prendre les décisions nécessaires en cas de besoin.
9. Vous devez vérifier si votre ordinateur est connecté avec un UPS (écrit sur les prises), en cas contraire vous devez enregistrer votre travail toutes les 5 minutes (à cause des coupures d'électricité). Cela vous aidera à enregistrer votre travail même lorsque l'appareil s'éteint.
10. Vous disposez d'un temps limité et précis pour accomplir votre travail et vous devez compléter dans les délais impartis;
11. Une attention particulière devrait être prêtée en cas d'absence de données ou de réponse, en principe ce sont des cellules vides, à l'exception des informations numériques (quantitatives) où on doit saisir ZERO.
12. Une attention particulière devrait également être de mise quant à la saisie des données correctes dans les colonnes correctes de la base de données, en particulier lorsque les réponses 'Non' ou 'Non applicable' impliquent des colonnes avec des cellules vides.

### **Responsabilités - Contrôle des données et superviseur**

13. Le superviseur doit contrôler environ 3% des questionnaires saisis par chaque membre du personnel de saisie de données après le premier jour. Au cas où il y a beaucoup d'erreurs l'échantillon devrait être élargi à 10%. S'il y a trop d'erreurs il est recommandé soit de former le personnel de nouveau et de lui préciser qu'il / elle doit s'améliorer, si l'on juge qu'il n'est pas la peine de le/la licencier.
14. Le superviseur doit donner des éclaircissements à chacune des difficultés ou à chacun des problèmes que de personnel de saisie de données soulève et qui concerne tout le monde, par exemple: une nouvelle mesure de poids et sa conversion, l'incompréhension d'une des questions spécifiques, etc.
15. Au cas où il y aurait un manque de clarté sur l'écriture ou les corrections apportées par les enquêteurs, le superviseur devrait aider le personnel de saisie de données à résoudre leur confusion et prendre la bonne décision.
16. Au cas où, lors du contrôle des données, les valeurs aberrantes sont identifiées (données incohérentes - en dehors d'un intervalle normal), d'abord vous devez vérifier la réponse dans le questionnaire. Si la saisie est correctement effectuée et vous trouvez qu'elle est encore aberrante, alors vous avez deux possibilités, soit l'éliminer (cellule vide), ou, dans le cas où la taille de l'échantillon est assez grande et qu'il n'y a pas trop de valeurs aberrantes, la garder, car il n'aura pas une incidence significative au cours de l'analyse des données.
17. Après chaque journée de travail, le superviseur doit obtenir un bref résumé de la saisie des données du personnel concernant leur progrès réalisés et les difficultés rencontrées. Cela permettra au superviseur de préparer le matériel pour le lendemain.
18. Le superviseur doit faire une sauvegarde sur un second support électronique des données saisies deux fois par jour (à l'heure du déjeuner et le soir) fin de réduire le risque de perte de données.

**NOUS VOUS REMERCIONS ET VOUS SOUHAITONS BON SUCCES PENDANT LE PROCESSUS**