

Session de formation sur STATA 14
Banque Mondiale, Nouakchott, janvier 2017

Exercices

Facilitateur : Araar Abdelkrim

PARTIE I

Exercice 1 interactif : *changement du nom des variables*

Changer les noms des variables des fichiers *bkf94l.dta* et *bkf98l.dta* par des noms français. Sauvegardez-les sous les noms *bkf94.dta* et *bkf98.dta*.

Exercice 2 interactif :

Étiqueter les deux bases de données créées en exercice 1, leurs variables, et les modalités de certaines variables.

Exercice 3 interactif : *exploration des données et statistiques descriptives et utilisation des pondérations au niveau du ménage et de l'individu.*

1. Donner le nombre d'observations de l'enquête *bkf98.dta*.
2. Donner le nombre d'observations par strate.
3. Donner la répartition des ménages à travers les strates.
4. Donner la répartition de la population à travers les strates.
5. Donner une idée sur la répartition de la population à travers les strates et les milieux de résidence.
6. Donner la moyenne des variables *dep_tete* et *dep_tete_reel*. Commenter.
7. Donner la moyenne des variables *dep_tete* et *dep_tete_reel* en tenant compte du poids de chaque observation.
8. Produire les principales statistiques de la distribution des dépenses par tête des ménages dont le chef est de sexe masculin.
9. Donner uniquement la moyenne pondérée de la variable *dep_tete* par groupe socioéconomique puis par la taille du ménage.
10. Donner uniquement la moyenne pondérée de la variable *dep_tete* par groupe socioéconomique et par la taille du ménage pour les ménages ayant une taille inférieure à 7.

Exercice 4 interactif : *étude de la corrélation entre certaines variables*

1. Analyser la corrélation entre les variables *dep_tete*, *dep_tete_reel*, et *taille* en utilisant les données du fichier *bkf98.dta*.
2. Analyser la corrélation entre les variables *dep_tete* et *dep_tete_reel* en tenant compte du poids associé à chaque observation.
3. Produire la matrice de variance-covariance des variables *dep_tete* et *dep_tete_reel* en tenant compte du poids associé à chaque observation.
4. Produire la matrice de corrélation entre les variables *dep_tete* et *dep_tete_reel* par sexe en tenant compte du poids associé à chaque observation. Vérifier si la corrélation entre ces deux variables est statistiquement significative au seuil de 1%.

Exercice 5 interactif : *tests de différence de moyenne*

L'enquête *Sample.dta* est une enquête représentative du revenu individuel en 2002 et en 2006.

1. Tester si la moyenne du revenu individuel en 2002 (*revenu02*) est égale à 70.
2. Tester si la moyenne du revenu individuel en 2002 (*revenu02*) est statistiquement différente de la moyenne du revenu individuel réel en 2006 (*revenu06r*). Consulter l'aide de STATA pour décider si l'utilisation de l'option *unpaired* avec la commande nécessaire pour effectuer ce test est justifiée dans ce cas.
3. Que faut-il faire pour tester si la moyenne du revenu individuel en 2002 dans le milieu urbain est supérieure, égale, ou inférieure à celle du milieu rural.
4. Tester si la variance de la variable *revenu02* est statistiquement différente selon la zone de résidence.
5. Quelle option faut-il introduire en réponse à la question 3 pour tenir compte du résultat obtenu en réponse à la question 4.

- Créer une variable `revenu02_u` qui prend les valeurs de `revenu02` lorsque le ménage réside en zone urbaine (`zone = 0`) et une variable `revenu02_r` qui prend les valeurs de `revenu02` lorsque le ménage réside en zone rurale (`zone = 1`).
- Tester si les deux variables `revenu02_u` et `revenu02_r` ont la même moyenne sous l'hypothèse qu'elles n'ont pas la même variance. Préciser si l'utilisation de l'option `unpaired` avec la commande nécessaire pour effectuer ce test est justifiée dans ce cas. Comparer les résultats de ce test avec ceux obtenus en réponse à la question 5.

Exercice 6 interactif : création d'une variable identifiante des ménages

- Ordonner les observations du fichier *bkf94.dta* par ordre croissant de la variable de l'unité d'échantillonnage primaire (la variable `uep`).
- Créer une nouvelle variable, appelée `numero_men`, qui indique le numéro (d'ordre) du ménage dans chaque `uep`.
- Produire les statistiques descriptives de la variable `uep`. Créer ensuite une nouvelle variable, `grappe`, telle que $grappe = (strate * 1000) + uep$.
- Produire les statistiques descriptives de la variable `numero_men`. Générer ensuite une nouvelle variable, appelée `id`, qui identifie les ménages. Vérifier que la variable générée est bien un identificateur unique.
- Étiqueter les variables créées.
- Appliquer les mêmes étapes au fichier *bkf98.dta* pour créer une variable identificatrice des ménages de cette base.

Exercices sur les commandes `drop` et `keep`

Exercice 7 : Élimination des observations dont la variable `gse` prend une valeur strictement inférieure à 5 du fichier *bkf94.dta*.

Exercice 8 : Conserver uniquement les observations dont la variable `strate` prend la valeur 3 du fichier *bkf98.dta*. Proposer une commande permettant de vérifier que vous aviez bien réussi votre tâche.

Exercice 9 : À partir de la base de données *bkf94.dta*, créer les deux fichiers suivants :

- Le premier fichier doit être sauvegardé sous le nom de *dep_men_94.dta* et il ne doit inclure que les variables `id`, `grappe`, `taille` et `dep_tete`.
- Le deuxième doit être sauvegardé sous le nom de *profil_men_94* et doit inclure toutes les variables du fichier *bkf94.dta* à l'exception des variables `taille` et `dep_tete`.

Est-il possible d'utiliser la commande **keep** pour la création du fichier *profil_94.dta* ? Si oui, faut-il privilégier l'utilisation de la commande **keep** ou **drop** dans ce cas ?

Exercices sur la commande `append`

Exercice 10 :

Concaténer les deux fichiers *bkf94.dta* et *bkf98.dta* après avoir créé une variable qui indique l'année de l'enquête. Sauvegarder la nouvelle base de données sous le nom *bkf94_98.dta*.

Exercices sur la commande `merge`

Exercice 11 :

Vous avez besoin de fusionner le fichier *dep_men_94.dta* et *profil_men_94.dta*. Toutefois, vous n'êtes pas sûr

qu'elles contiennent les mêmes observations ou que les observations soient classées de façon identique. Quel est le procédé le plus sûr qui permet de fusionner ces deux fichiers ? Fusionner les deux fichiers, puis sauvegarder la nouvelle base dans un nouveau fichier *dep_profil_94.dta*.

Exercice 12 : La commande expand

1. À partir du fichier ménage *dep_men_94.dta*, créer un fichier individuel où chaque observation est dupliquée autant de fois que la taille du ménage.
2. Ordonner les observations par ordre croissant de la variable *id*.
3. Créer une variable qui correspond au numéro de chaque individu dans le ménage appelé *numero_indv*.
4. Créer une variable qui indique le revenu total du ménage, appelé *rev_men*, sous l'hypothèse que la dépense totale est égale au revenu total (absence d'épargne).
5. Créer une variable correspondant au revenu de chaque individu au sein du ménage, appelée *rev_indv*, sous les hypothèses suivantes :
 - Le revenu individuel est égal au revenu total lorsque la taille du ménage est égale à 1.
 - Le revenu du premier membre (*numero_indv* = 1), c.-à-d. du chef du ménage, représente 60% du revenu total.
 - Le revenu du deuxième membre (*numero_indv* = 2) représente 40% du revenu total.
 - Le revenu des autres membres est nul.
7. Proposer une méthode pour vérifier si les valeurs prises par *rev_indv* n'altèrent pas l'égalité supposée entre les dépenses totales et le revenu total des ménages.
8. Sauvegarder le nouveau fichier sous le nom *rev_indv_94.dta* en ne gardant que les variables *id*, *numero_indv* et *rev_indv*.

Exercice 13 : La fonction sum()

Tracer la courbe de Lorenz de la variable *dep_tete* qui se trouve dans le fichier *bkf94.dta*.

PARTIE II

Considérations générales :

- 1) Sauf lorsqu'il est spécifiquement indiqué, utilisez comme variable de niveau de vie la variable de dépenses totales par équivalent adulte, soit **NVIE** (pour BFK94), **NVIEZ** (pour BFK98). Le seuil de pauvreté au Burkina est de 41099 FCFA au Burkina. Si un indice de pauvreté n'est pas spécifié, utilisez l'indice numérique de pauvreté.
- 2) Sauf lorsqu'il y est spécifiquement indiqué, la variable du coefficient d'expansion est la variable **POIDS** ou **WEIGHT**. Dans DAD, il faudra généralement spécifier **POIDS** ou **WEIGHT** dans la structure de l'échantillon et la variable **TAILLE** pour la taille du ménage.
- 3) Pour les exercices incluant de l'inférence statistique, n'oubliez pas de fixer dans DAD la structure d'échantillonnage en utilisant les variables **STRATA**, **PSU** (et **POIDS** ou **WEIGHT**).

Exercice I:

• Indices de prix et dynamique de la pauvreté.

1. Chargez le fichier BKF_94.daf et initialisez la structure d'échantillonnage avec les variables STRATA, PSU et POIDS.
2. Calculez l'indice numérique et l'écart moyen du seuil de pauvreté lorsque le seuil de pauvreté z_94 égale 41099 FCFA et en utilisant la variable NVIE comme variable d'intérêt (utiliser l'application : Poverty| FGT Index).
3. Chargez le fichier BKF_98.daf et initialisez la structure d'échantillonnage SD avec les variables STRATA, PSU et POIDS. Calculez l'indice numérique et l'écart moyen du seuil de pauvreté lorsque le seuil de pauvreté, z_98 égale 72690 FCFA et en utilisant la variable NVIE comme variable d'intérêt (utiliser l'application : Poverty|FGT Index).
4. Est-ce que la comparaison entre les deux indices de pauvreté trouvés pour les deux périodes est cohérente à travers les indices? Justifiez votre réponse.
5. Maintenant, calculez de nouveau l'indice numérique et le fossé de la pauvreté pour la période de 1998 lorsque la variable d'intérêt est NVIEZ.
6. Écrivez (ou utilisez) un petit programme *fgt* Stata permettant de calculer l'indice FGT et la moyenne des niveaux de vie tel que les arguments de ce programme sont, respectivement, le poids final associé à l'observation, le niveau de vie, le paramètre alpha et le seuil de pauvreté. En utilisant ce programme, faites de nouveau le calcul demandé dans la question 2.
7. Utilisez DASP pour calculer la différence dans la pauvreté entre 1998 et 1994, lorsque la pauvreté est mesurée par l'indice numérique.
8. Tracez les courbes FGT lorsque :
 - Le paramètre $\alpha = 0$.
 - L'intervalle de la ligne de pauvreté et entre 0 et 60000 F CFA.

- Les distributions sont celles des groupes composés selon le genre du chef de ménage.

Exercice II:

• *Décomposition de l'indice de pauvreté totale*

1. Chargez le fichier BKF_94.daf et initialisez la structure d'échantillonnage avec les variables STRATA, PSU et POIDS
2. Décomposez la pauvreté totale par groupe socio-économique (GSE). Commenter les résultats trouvés.
3. Répétez la question (2.) lorsque les groupes de population sont déterminés par milieu de résidence.
4. Écrivez un petit programme Stata permettant de calculer la contribution de chaque groupe à la pauvreté totale. (Utiliser le fichier bkf94I.dta).
5. Utilisez DASP, et faire les mêmes décompositions, demandées dans les questions (2.) et (3.).

Exercice III:

• *Évolution de l'inégalité au Burkina Faso entre 1994 et 1998*

1. Chargez simultanément les fichiers BKF_94.daf et BKF_98.daf et initialisez la structure d'échantillonnage de chaque fichier avec les variables STRATA, PSU et POIDS.
2. Tracez les courbes de Lorenz les distributions de niveaux de vie de 1994 et 1998 au Burkina. Que peut-on dire en comparant ces deux courbes ?
3. Tracez les courbes de Lorenz par milieu de résidence. Laquelle, des deux distributions, est plus égale ?
4. Calculez l'indice de Gini par groupe socio-économique pour l'année 1994.
5. Utilisez DASP et tracer les courbes de Lorenz par milieu de résidence et calculer aussi les indices de Gini. Sauvegardez les résultats dans un fichier *.log et le graphique sous format *.wmf (un format compatible avec les documents de type Word).

Guide sur les bases de données utilisées pour ces exercices

Liste des variables pour les enquêtes du Burkina Faso

1-BKF94.daf

Cet échantillon, représentatif au niveau national, a été tiré par échantillonnage à deux étapes et stratification. Sept strates ont été considérées. Cinq de ces strates étaient rurales et deux étaient urbaines. Les zones de dénombrement PSU ont été tirées dans une première étape à partir d'une liste des zones de dénombrement provenant du recensement de 1985. La première étape de sélection (sauf pour la strate 7: Ougadougou-Bobo-Dioulasso) a été faite avec une probabilité proportionnelle à la taille de la population dans la strate et sans remplacement. Par la suite, 20 ménages ont été systématiquement sélectionnés pour chaque zone de dénombrement sélectionnée.

POIDS : Poids d'échantillonnage

TAILLE : Taille du ménage
STRATA : Strate où le ménage vit (7strates)
PSU : Zone de dénombrement où le ménage vit
GSE : Groupe socio-économique du chef de ménage.

1. Salariés du secteur public
2. Salariés du secteur privé
3. Artisans et commerçants
4. Autres actifs
5. Agriculteurs de rente
6. Agriculteurs vivriers
7. Inactifs

SEXE : Sexe du chef de ménage
1. Homme
2. Femme

ZONE : Milieu de résidence
1. Rural
2. Urbain

DEPTOT : La dépense annuelle totale du ménage (FCFA)
DEPEQ : La dépense annuelle par équivalent adulte (FCFA)
NVIE : La dépense annuelle par tête (FCFA)
TPOND : Poids*Taille
DEPAL : La dépense alimentaire annuelle (FCFA)

2-BKF98.daf

Cet échantillon, représentatif au niveau national, a été tiré avec échantillonnage à deux étapes et avec de stratification. L'approche utilisée, pour cette enquête, est similaire à celle de 1994 avec cependant, dix strates au lieu de sept pour 1994.

PSU : Zone de dénombrement où le ménage vit
Poids : Poids d'échantillonnage
SEXE : Sexe du chef de ménage
1. Homme
2. Femme

STRATA : Strates ou régions économiques (10régions)

Zone	: Milieu de résidence <ol style="list-style-type: none"> 1. Rural 2. Urbain
GSE	: Groupe socio-économique du chef de ménage <ol style="list-style-type: none"> 1. Salariés du secteur public 2. Salariés du secteur privé structuré 3. Salariés du secteur privé non structuré 4. Indépendants ou employeurs non agricoles 5. Agriculteurs de coton 6. Agriculteurs vivriers 7. Aides familiaux, bénévoles, apprentis 8. Chômeurs 9. Chefs de ménages inactifs
Taille	: Taille du ménage
GSE94	: Groupe socio-économique du chef de ménage. <ol style="list-style-type: none"> 1. Salariés du secteur public 2. Salariés du secteur privé 3. Artisans et commerçants 4. Autres actifs 5. Agriculteurs de rente 6. Agriculteurs vivriers 7. Inactifs
STRAT94	: Strate où le ménage vit(7strates)
DEPEQ	: Dépense totale par équivalent adulte (en F CFA de 1998)
DEPEQZ	: Dépense totale par équivalent adulte (déflatée par $z_{1994}/z_{2_{1998}}$)
DEPEQI	: Dépense totale par équivalent adulte (déflatée par IPC_{1994}/IPC_{1998})
NVIEZ	: Dépense totale par tête (déflatée par $z_{1994}/z_{2_{1998}}$)
NVIEI	: Dépense totale par tête (déflatée par IPC_{1994}/IPC_{1998})