Ecole thématique
« Gestion et analyse de données d'enquêtes épidémiologiques »

20-29 avril 2010, Rabat, Maroc

Gestion de données.
Exercices pratiques avec EpiData Analysis

Pierre Traissac

IRD

UMR 204 NUTRIPASS IRD, UM1, UM2

« Prévention des malnutritions et pathologies associées »

Montpellier, France

Gestion de données
Exercices pratiques avec EpiData Analysis

1. Modèle de données

- Source des données

Sous ensemble des données de l'enquête nationale de nutrition Tunisie 1996/97 (INNTA, Tunis, Tunisie)

- Entités

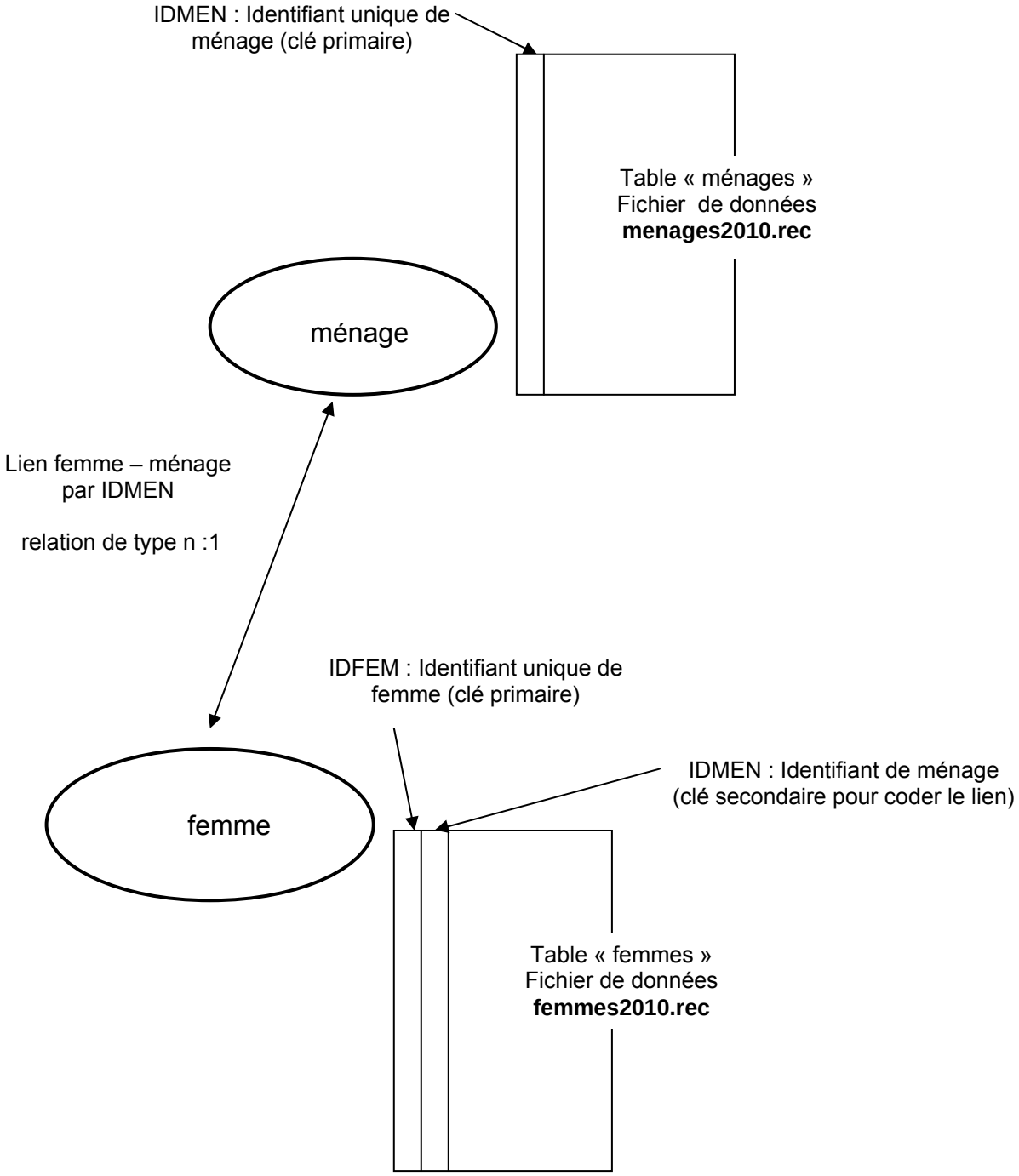
Table ménages : identifiant primaire IDMEN
Table femmes : identifiant primaire IDFEM, identifiant secondaire IDMEN

(c.f. dictionnaire des variables ci-après pour le détail des variables ménage et femme)

- Relation

Filiation : « ménage IDMEN contient femme IDFEM» – « femme IDFEM est dans ménage IDMEN»
La relation est incluse dans la table femme (identifiant secondaire IDMEN pour coder le lien)

- Schéma de la base



Gestion de données
Exercices pratiques avec Epidata Analysis

2. Dictionnaire des variables

N.B. : Les fichiers de données des exercices sont un sous-ensemble de la base de donnée de l'enquête transversale nationale de nutrition Tunisie 1996/97, propriété de l'INNTA (Institut National de Nutrition de Tunisie, Tunis, Tunisie). Ces données sont la propriété de l'INNTA et mises à disposition des utilisateurs dans le cadre des exercices proposés, à l'exclusion de toute autre utilisation.

Table « menages » : unité statistique ménage

- fichier format Epidata : **menages2010.rec** (n=1763 observations, p=39 variables)
- date de l'état : 29/03/2010

Numéro d'ordre dans le fichier	Nom de la variable	Format	Contenu	Unités ou codes	n	Minimum - maximum	Type statistique	Remarques diverses
1	**idmen**	Numérique	Clé primaire : identifiant unique de ménage	sans objet	1763	110701 à 832406	identifiant	Combine n° de gouvernorat, de délégation et de ménage
2	**gouvern**	Numérique	Gouvernorat	Détails non donnés dans le cadre de l'exercice	1763	11 à 83	qualitatif	
3	**deleg**	Numérique	Numéro de délégation dans le gouvernorat	Détails non donnés dans le cadre de l'exercice	1763	1 à 72	qualitatif	
4	**numen**	Numérique	Numéro de ménage dans la délégation	Numéro séquentiel	1763			
5	**datenq**	Date	Date de l'enquête		1763	13 jan 1996 au 28 dec 1997		Calculée à partir des variables initiales *datevisj*, *datevism*, *datevisa*
6	**stratech**	Numérique	Strates de l'échantillonnage	Détails non donnés dans le cadre de l'exercice	1763	111 à 832	qualitatif	Combine *gouvern* et *milieu*
7	**gouvdel**	Numérique	Grappes de l'échantillonnage	Détails non donnés dans le cadre de l'exercice	1763	1107 à 8324	qualitatif	Combine *gouvern* et *deleg*
8	**pond1**	Numérique	Poids de sondage lié à la stratification			316 à 2690		Car probabilités de tirage au sort ménages inégales. Est constante par stratech.
9	**region**	Numérique	Région administrative	1 : Grand Tunis 2 : Nord Est 3 : Nord Ouest 4 : Centre Ouest 5 : Centre Est 6 : Sud Ouest 7 : Sud Est	1763	1 à 7	qualitatif	Est un recodage de *gouvern*
10	**habitat**	Numérique	Type d'habitat	1 : grandes villes 2 : autres communes 3 : rural aggloméré 4 : rural dispersé	1763	1 à 4	qualitatif	Selon classification en vigueur lors de l'enquête (Institut National de Statistique Tunisien)
11	**milieu**	Numérique	Type d'habitat	1 : urbain 2 : rural	1763	1 à 2	qualitatif	Est un recodage de *habitat*
12	**typeloge**	Numérique	Type de logement	1 : villa 2 : appartement 3 : studio 4 : maison arabe 5 : gourbi 6 : autre	1703	1 à 6	qualitatif	
13	**mode**	Numérique	Mode d'occupation du logement	1 : propriétaire 2 : locataire 3 : logé à titre gratuit 4 : accédant à la propriété	1702	1 à 4	qualitatif	
14	**nbchamb**	Numérique	Nombre de chambres du logement	Nombre de chambres	1529	1 à 9	quantitatif	
15	**bain**	Numérique	Salle de bain dans logement	1 :oui 2 :non	1743	1 à 2	dichotomique	
16	**douche**	Numérique	Douche dans logement	1 :oui 2 :non	1744	1 à 2	dichotomique	
17	**salleau**	Numérique	Salle d'eau dans logement	1 :oui 2 :non	1742	1 à 2	dichotomique	
18	**toilette**	Numérique	Toilettes dans logement	1 :oui 2 :non	1735	1 à 2	dichotomique	
19	**robinet1**	Numérique	Eau par robinet dans logement	1 :oui 2 :non	1739	1 à 2	dichotomique	
20	**puit**	Numérique	Eau à partir d'un puit	1 :oui 2 :non	1739	1 à 2	dichotomique	
21	**robinet2**	Numérique	Eau par robinet public	1 :oui 2 :non	1735	1 à 2	dichotomique	
22	**egout**	Numérique	Evacuation eaux usées par égout	1 :oui 2 :non	1739	1 à 2	dichotomique	
23	**fosse**	Numérique	Evacuation eaux usées par fosse sceptique	1 :oui 2 :non	1727	1 à 2	dichotomique	

La liste des variables ménage continue à la page suivante

Numéro d'ordre dans le fichier	Nom de la variable	Format	Contenu	Unités ou codes	N	Minimum - maximum	Type statistique	Remarques diverses
24	**electric**	Numérique	Logement raccordé au réseau électrique	1 :oui 2 :non	1744	1 à 2	dichotomique	
25	**sonede**	Numérique	Logement raccordé réseau sonede (eau)	1 :oui 2 :non	1736	1 à 2	dichotomique	
26	**frigo**	Numérique	Logement équipé d'un réfrigérateur	1 :oui 2 :non	1723	1 à 2	dichotomique	
27	**phone**	Numérique	Logement équipé d'un téléphone	1 :oui 2 :non	1738	1 à 2	dichotomique	
28	**tele**	Numérique	Logement équipé d'une télévision	1 :oui 2 :non	1738	1 à 2	dichotomique	
29	**video**	Numérique	Logement équipé d'un magnétoscope	1 :oui 2 :non	1743	1 à 2	dichotomique	
30	**cuisine**	Numérique	Logement équipé d'une cuisinière	1 :oui 2 :non	1741	1 à 2	dichotomique	
31	**lavess**	Numérique	Logement équipé d'un lave vaisselle	1 :oui 2 :non	1750	1 à 2	dichotomique	
32	**clime**	Numérique	Logement équipé d'un climatiseur	1 :oui 2 :non	1745	1 à 2	dichotomique	
33	**central**	Numérique	Logement équipé du chauffage central	1 :oui 2 :non	1751	1 à 2	dichotomique	
34	**parabole**	Numérique	Logement équipe d'une parabole	1 :oui 2 :non	1750	1 à 2	dichotomique	
35	**lavlinge**	Numérique	Logement équipé d'un lave linge	1 :oui 2 :non	1745	1 à 2	dichotomique	
36	**menage**	Numérique	Présence d'une aide ménagère	1 :oui 2 :non	1751	1 à 2	dichotomique	
37	**voiture**	Numérique	Ménage possède une voiture	1 :oui 2 :non	1742	1 à 2	dichotomique	
38	**eco1**	Numérique	Indice de niveau économique du ménage	Sans unité (classement des ménages par indice de niveau de vie croissant)	1725	0 à 100	quantitatif	Est proportionnel à la coordonnée du ménage sur le premier axe d'une analyse factorielle sur un ensemble de variables d'équipement et de niveau de vie du ménage (Traissac et al RESP 1997, Delpeuch et al. EJCN 1994). Utilise *typeloge*, *mode* et les 23 variables de *bain* à *voiture*
39	**depenses**	Numérique	Dépenses alimentaires mensuelles	Dinars Tunisiens courants à la date d'enquête	1508	20 à 650	quantitatif	Déclaratif

Table « femmes » : unité statistique femme de 20 à 59 ans , non enceinte, non allaitante
- fichier format Epidata : **femmes2010.rec** (n=1849 observations, p= 36 variables)
- date de l'état : 29/03/2010

Numéro d'ordre dans le fichier	Nom de la variable	Format	Contenu	Unités ou codes	n	Minimum - maximum	Type statistique	Remarques diverses
1	**idfem**	Numérique	Clé primaire : identifiant unique de femme	sans objet	1849	1107011 à 8324061	identifiant	Combine n° de gouvernorat, de délégation, de ménage et de femme
2	**idmen**	Numérique	Clé secondaire : identifiant de ménage	sans objet	1849	110701 à 832406	identifiant	Combine n° de gouvernorat, de délégation et de ménage
3	**gouvern**	Numérique	Gouvernorat	Détails non donnés dans le cadre de l'exercice	1849	11 à 83	qualitatif	
4	**deleg**	Numérique	Numéro de délégation dans le gouvernorat	Détails non donnés dans le cadre de l'exercice	1849	1 à 72	qualitatif	
5	**numen**	Numérique	Numéro de ménage dans la délégation	Numéro séquentiel	1849	1 à 22	qualitatif	
6	**nufem**	Numérique	Numéro de femme dans le ménage	Numéro séquentiel	1849	1 à 5	qualitatif	
7	**datenq**	Date	Date de l'enquête		1849	13 jan 1996 au 24 dec 1997		Calculée à partir des variables initiales *datevisj*, *datevism*, *datevisa* (cf programme fortstatagen1.do)
8	**stratech**	Numérique	Strates de l'échantillonnage	Détails non donnés dans le cadre de l'exercice	1849	111 à 832	qualitatif	Combine *gouvern* et *milieu*
9	**gouvdel**	Numérique	Grappes de l'échantillonnage	Détails non donnés dans le cadre de l'exercice	1849	1107 à 8324	qualitatif	Combine *gouvern* et *deleg*
10	**pond1**	Numérique	Poids de sondage lié à la stratification	sans objet	1849	316 à 2690		Car probabilités de tirage au sort ménages inégales. Est constante par stratech et ménage.
11	**pond2**	Numérique	Poids pour redressement non réponses	sans objet	1849	0.56 à 1.77		Postratification : redressement de la structure d'âge et sexe par région
12	**region**	Numérique	Région administrative	1 : Grand Tunis 2 : Nord Est 3 : Nord Ouest 4 : Centre Ouest 5 : Centre Est 6 : Sud Ouest 7 : Sud Est	1849	1 à 7	qualitatif	Est un recodage de gouvern
13	**habitat**	Numérique	Type d'habitat	1 : grandes villes 2 : autres communes 3 : rural aggloméré 4 :rural dispersé	1849	1 à 4	qualitatif	Selon classification en vigueur lors de l'enquête (Institut National de Statistique Tunisien)
14	**milieu**	Numérique	Type d'habitat	1 : urbain 2 : rural	1849	1 à 2	qualitatif	Est un recodage de habitat
15	**jnais**	Numérique	Jour de naissance	sans objet	1849	1 à 31		
16	**mnais**	Numérique	Mois de naissance	sans objet	1849	1 à 12		
17	**anais**	Numérique	Année de naissance	sans objet	1849	1936 à 1977		
18	**matc3**	Numérique	Statut matrimonial	1 : célibataire 2 : mariée 3 : veuve ou divorcée	1849	1 à 3	qualitatif	Est un recodage de la variable initiale *etatcivi* à 4 modalités par regroupement de veuves et divorcées
19	**scoc4**	Numérique	Niveau scolaire	1 : analphabète 2 : primaire 3 : secondaire 4 : supérieur	1814	1 à 4	qualitatif	Est un recodage de la variable initiale *niveau* à 6 modalités
20	**actpc2**	Numérique	Activité professionnelle	1 :oui 2 :non	1842	1 à 2	dichotomique	Est un recodage de la variable initiale *profess* contenant des codes profession
21	**parc4**	Numérique	Parité	1 : 0 2 :1,2 3 : 3,4 4 : 5 et +	1813	1 à 4	qualitatif	Est un recodage de la variable initiale *nbgross* (nombre de grossesses)

La liste des variables femme continue à la page suivante

Numéro d'ordre dans le fichier	Nom de la variable	Format	Contenu	Unités ou codes	n	Minimum - maximum	Type statistique	Remarques diverses
22	**poids**	Numérique	Poids	kg	1837	32.7 à 159	quantitatif	
23	**taille**	Numérique	Taille	cm	1836	113 à 182	quantitatif	
24	**tt**	Numérique	Tour de taille	cm	1738	57 à 135	quantitatif	
25	**th**	Numérique	Tour de hanche	cm	1738	63 à 154	quantitatif	
26	**tas**	Numérique	Tension artérielle systolique	mm hg	1825	80 à 240	quantitatif	
27	**tad**	Numérique	Tension artérielle diastolique	mm hg	1823	30 à 140	quantitatif	
28	**hemoglob**	Numérique	Hémoglobine	g/dl	1789	5.3 à 22.2	quantitatif	
29	**glycemie**	Numérique	Glycémie à jeun	mmol/l	1771	2.5 à 51	quantitatif	
30	**cholest**	Numérique	Cholesterolémie totale	mmol/l	1657	0.61 à 9.96	quantitatif	
31	**trigly**	Numérique	Triglycéridémie	mmol/l	1652	0.23 à 8.18	quantitatif	
32	**cal**	Numérique	Apport énérgétique total / jour	kcal	1819	818 à 3832.73	quantitatif	
33	**calp**	Numérique	Apport énérgétique total / jour rapporté aux besoins	sans unité	1818	0.36 à 1.76	quantitatif	
34	**glucide1**	Numérique	Apport en glucides en % de l'apport total	sans unité	1819	43.2 à 71	quantitatif	
35	**protide1**	Numérique	Apport en protéines en % de l'apport total	sans unité	1819	7.1 à 17	quantitatif	
36	**lipide1**	Numérique	Apport en lipides en % de l'apport total	sans unité	1819	18.0 à 47.4	quantitatif	

Gestion de données
Exercices pratiques avec EpiData Analysis

3. Texte des exercices pratiques

- Le but des exercices suivants est de se familiariser avec le logiciel et de mettre en œuvre quelques commandes utiles pour la gestion de données dans Epidata Analysis.
- On donne certaines indications concernant les commandes utiles pour répondre aux questions posées. Il peut également être utile de vous inspirer des exemples donnés lors des exposés et la documentation EpiData Analysis qui vous a été remise.
- Lors du travail sur un fichier de données, il est nécessaire d'avoir sous les yeux le dictionnaire de variables correspondant (c.f. ci-dessus).
- Utiliser l'aide en ligne ou les documents qui vous ont été remis (impression des diaporamas de cours et/ou résumé des instructions) pour des précisions sur l'utilisation et la syntaxe des commandes.
- Dans les exercices, une importance particulière est donnée aux vérifications, contrôles etc. après chaque opération, ceci dans le cadre de la démarche « qualité des données ».
- Pour répondre aux questions posées, dans un premier temps on rentrera les commandes au clavier en mode interactif, comme indiqué lors de la démonstration. Plus avant dans les exercices il vous sera proposé d'utiliser l'éditeur de programme (fenêtre « Program Editor ») pour écrire des programmes et les utiliser avec la commande « run ».

1 – Lecture de fichiers, visualisation, vérifications (tris, statistiques élémentaires), création de nouvelles tables par sélection de lignes ou sélection de colonnes

Avant tout travail sur un fichier de données il est nécessaire de s'assurer que l'on dispose d'une copie de sauvegarde (sur disquette, CD, clé USB…) au cas où une mauvaise manipulation compromettrait son intégrité. Dans cet exercice, à défaut de pouvoir copier les fichiers sur une disquette, un CD-ROM ou une clé USB, créer un répertoire SAUVE, distinct de votre répertoire de travail et y faire une copie des deux fichiers (ce qui vous permettra d'y revenir en cas de problème pendant les exercices)

- Lire la table menages2010.rec (commande read ou menu File / Open).

- Combien a-t-il d'observations, de variables ? Quelles sont les variables contenues dans cette table ? Décrire le contenu du fichier (describe). Cela correspond-il aux informations sur le dictionnaire (vérifier effectif, minimum, maximum etc.) ?

- Afficher le contenu de la table à l'écran (commande browse ou bouton dans la barre d'outils ou Menu Window/Browse). Se déplacer dans le fichier. Y-a-t-il des données manquantes ? Pour quelles variables, quelles observations par exemple ? Voir les fonctions du menu 'clic droit' dans la fenêtre browse.

- Faire afficher les seulement variables *idmen*, *region*, *typeloge* de la table (list ou browse).

- Afficher le contenu de la table, trié par *depenses* croissantes (sort). Commenter la position des valeurs manquantes et les implications possibles ?

- Afficher le contenu de la table trié par *habitat* et *eco1.* N'afficher que les variables correspondantes et vérifier le tri.

- Vérifier la liste des modalités et leur fréquence pour chacune des variables *habitat, typeloge* et *mode* (freq). Voir l'effet de l'option /m. Noter les valeurs des modalités, effectifs et fréquences. Y a-t-il des données manquantes ? pour quelle(s) variable(s) et quel(s) pourcentage(s) ?

- Lire la table femmes2010.rec.

- Vérifier la cohérence avec le dictionnaire de variables (describe) : liste des variables, minimum et maximum pour les variables quantitatives, liste des modalités pour les variables qualitatives.

- Donner la distribution de fréquence de *actpc2* et *matc3* (freq). Noter les valeurs des modalités, effectifs et fréquences.

- Sélectionner les femmes qui ont un apport énergétique ≥ 2500 kcal par jour (select if). A combien de femmes cela correspond-il (count) ? Vérifier les valeurs de l'apport énergétique en affichant la sélection après avoir trié par *cal* (sort et browse) ou en utilisant describe

- N'afficher que les femmes qui ont un apport énergétique < 2500 kcal par jour et ayant une activité professionnelle (select if … and …). Combien de femmes sélectionnées (describe ou count). Vérifier la sélection.

- A partir la table femmes, créer une nouvelle table (femmarie.rec) contenant les variables *poids, taille, tt, th, anais, jnais et mnais* des femmes mariées (select, keep et/ou savedata).

- Ouvrir cette nouvelle table. Combien la table femmarie.rec a-t-elle d'enregistrements ?

- Afficher la liste de ses variables.

- Afficher le contenu de la table à l'écran. Commenter.

- Sélectionner seulement les femmes nées après 1960 (combien-y-en-a-t-il ?). Aurait-on obtenu le même résultat à partir de la table initiale (combien y-en aurait-il) ?

-Dans la mesure où elle ne sera pas utile dans la suite, fermer et supprimer la table femmemarie.rec (erase, ou via l'explorateur Windows).

- Si vous ne l'avez pas encore utilisée, voir le contenu de la fenêtre « history ».

2 – Concaténation de tables

- Mise en relation des tables femmes2010.rec et menages2010.rec

- Réaliser la fusion des tables femmes2010.rec et menages.2010.rec (read, merge avec option /table, voir la documentation).

- Visualiser la table résultante. Est-ce bien le résultat que nous attendions ? Combien a-t-elle d'enregistrements ? - Pourquoi les valeurs des variables ménages sont-elles parfois dupliquées ?

- Afficher la liste des variables. Réfléchir sur le nombre de variables de la table fusionnée : est ce cohérent ?

- A partir de cette table fusionnée, utiliser tables pour donner la répartition de fréquence des variables *actpc2* et *matc3*. Comparer aux valeurs obtenues précédemment (dans l'exercice à partir de femmes2010.rec). Commentaire.

- Toujours à partir de cette table fusionnée, donner la répartition de fréquence pour chacune des variables *typeloge* et *mode*. Comparer aux valeurs obtenues précédemment (dans l'exercice à partir de menages2010.rec). Commentaire.

- Donner la distribution de fréquence du mode d'occupation du logement (variable *mode*), pour les femmes qui ont une activité professionnelle et celles qui n'en ont pas (variable *actpc2*) (freq avec select ou if). Refaire le calcul en utilisant un tableau croisé (tables).
- Sauver sur disque la table fusionnée : on nommera le fichier menfem.rec (savedata). Vérifier la création du fichier (read).
- Documenter la création de la table menfem.rec (par exemple sous la forme d'un schéma précisant de quelle façon elle a été obtenue à partir des tables initiales).

3 – Changement d'unité statistique (aggrégation)

- A partir de la table femmes2010.rec

- Passage d'u.s. femme à ménage : créer un nouveau fichier de données contenant pour chaque ménage, une variable donnant le nombre de femmes enquêtées (aggregate avec option /close ou save)

- Combien a-t-il de variables ? D'enregistrements? Pourquoi ? Renommer la variable n en nbfem (gen, drop) et sauvegarder le fichier.

- Fusionner avec le tableau des ménages (on ouvrira d'abord le tableau menages2010.rec). Pourquoi certains ménages ont une valeur manquante pour nbfem ?

- Donner la distribution du nombre de femmes par ménages (freq avec et sans faire apparaitre les données manquantes)

- Sauver la nouvelle table sous le nom menage2010_2.rec

4 – Calcul de nouvelles variables, recodages

- A partir de la table femmes2010.rec

- Créer une nouvelle variable *imc* contenant les valeurs de l'indice de masse corporelle (gen).

-Vérifier la création de la variable *imc* (browse ou describe) ainsi que ses minimum et maximum (describe). Y-a-t-il des données manquantes ? Pourquoi ?

- Attribuer un label à cette nouvelle variable (label). Vérifier que la variable et le label apparaissent dans la fenêtre « variables ».

- Créer une variable *imc4* codant les 4 classes d'*imc* avec les bornes usuelles 18,5, 25 et 30 kg/m^2 (define, recode) et donner un label à cette nouvelle variable (label).

- Vérifier la cohérence des classes ainsi créées avec les valeurs initiales d'IMC (avec sort ou means)

- Donner la distribution de fréquence de la variable *imc4* (tables).

- Associer des labels « maigre, normal, surpoids, obésité » aux valeurs de imc4 (labelvalue).

- Créer une nouvelle variable (define et let ou gen) date de naissance (*datenais*) à partir des variables *jnais, mnais* et *anais* (fonction dmy). Examiner la variable (describe ou browse).

- Calculer l'âge en années à partir des variables *datenais* et *datenq,* vérifier le calcul (browse *age datenais datenq*, par exemple)

- Découper l'âge en classes de 10 ans (nouvelle variable *agec4*).

- Vérifier la cohérence des classes avec les valeurs initiales de l'âge (sort, describe).

- Attribuer un label à ces 2 nouvelles variables.

- A partir de la date d'enquête (*datenq*), donner le nombre de personnes enquêtées par année d'enquête (fonction year).

- Donner la distribution de fréquence de la variable *scoc4* (freq)

- Recoder la variable *scoc4* en *scoc3* (3 classes), en regroupant les modalités 3 et 4. Vérifier le recodage (freq ou tables).

- Recoder la variable *imc* en variable dichotomique (en 0,1) *obesite* (avec la borne 30 kg/m^2) (recode). Vérifier la cohérence des classes avec celles de *imcc4* (tables) et lui donner un label.

- Créer les indicatrices pour les modalités de la variable matc3 (ie 3 variables 0/1 matc31, matc32, matc33, qui codent les modalités de matc3). Vérifier la cohérence avec matc3 (freq, tables, browse).

- A partir des variables tas et tad, créer une variable hta codant l'hypertension artérielle (tas>=140 ou tad>=90) (attention à la gestion des données manquantes).

- Créer un nouveau fichier femmes2010_2.rec contenant toutes les variables de départ et les 7 nouvelles variables créées *imc*, *imc_c4, sco_c3, datenais, age, age_c4 obesite* et les indicatrices (savedata).

- Vérifier que la création de la table femmes2010_2.rec s'est effectuée correctement (read, browse ou describe).

- Compléter le dictionnaire de variables de femme2010.rec pour créer celui de femmes2010_2.rec (rajouter les informations concernant les variables *imc*, *imc_c4, datenais, age, age_c4, sco_c3, obesite,* indicatrices de matc3, *hta*).

- Ecrire à l'aide du « Program editor » un fichier de programme (que l'on nommera par exemple gen_femmes2.pgm) réalisant toutes les actions ci-dessus concernant la création et le codage des nouvelles variables. Utiliser ce programme (Run dans « Program Editor »). Une base pour le programme pourra être de sauvegarder le contenu de la fenêtre « history ».

N.B. : dans la pratique dès que le nombre d'opérations sur les données est important et également pour assurer la traçabilité de ces opérations, il sera indispensable d'utiliser des fichiers de programmes au lieu de travailler en interactif. Aussi penser dans les programmes à soigner la présentation et insérer suffisamment de commentaires dans le but de les rendre les plus explicités possibles (c.f. bonnes pratiques / qualité pour la gestion et l'analyse de données d'enquêtes).

- A partir de la table menages2010.rec

- Examiner la distribution de la variable *depenses* (freq avec option /CUM); recoder la variable en choisissant des bornes de classes cohérentes avec la distribution (recode), on nommera la nouvelle variable *depc3* (si 3 groupes, par exemple)

- Donner un label à la nouvelle variable

- Calculer un score de confort (variable *confort*) à partir des 12 variables "possessions" : *frigo, phone, tele, video, cuisine, lavess, clime, central, parabole, lavlinge, menage, voiture*. Les variables "possessions" devront être recodées en (0,1) et additionnées. Vérifier la création et décrire la variable ainsi obtenue.

- Donner un label à la variable *confort*.

- Découper la variable *confort* en classes en fonction de la distribution (3 classes par exemple).

- Créer une nouvelle table menages2010_2.rec contenant toutes les variables de départ et les nouvelles variables (savedata).

- Vérifier que la création de la table menages2.rec s'est effectuée correctement (read, browse, describe).

- Compléter le dictionnaire de variables de menages2010.rec pour créer celui de menages2010_2.rec (rajouter les informations concernant les variables *depc3*, *confort,* et *confc3*).

- Réaliser la fusion des tables femmes2010_2.rec et menages2010_2.rec

- Sauver sur disque la table fusionnée : on nommera le fichier menfem2010.rec (savedata). Vérifier la création du fichier.

- Ecrire (par exemple à l'aide du « Program editor ») un fichier de programme (que l'on nommera par exemple gen_menages2010.pgm) réalisant toutes les actions ci-dessus concernant la création et le codage des nouvelles variables ménages (on peut utiliser la fenêtre history pour retrouver les commandes utilisées). Utiliser ce programme (Run dans « Program editor »).

N.B. : dans la pratique dès que le nombre d'opérations sur les données est important et également pour assurer la traçabilité de ces opérations, il sera indispensable d'utiliser des fichiers de programmes au lieu de travailler en interactif.

Documentation :

- Vérifier que vous avez bien mis le dictionnaire de variables à jour.

- Documenter les opérations effectuées ci-dessus (schéma ou tableau précisant comment les tables femmes2010_2.rec menages2010_2.rec, menfem2010.rec ont été obtenues à partir des tables initiales).