

INITIATION A LA CONCEPTION DE MASQUE DE SAISIE AVEC CSP_{ro}

Rédigé par :

NAYO Ankouvi Mawoudoudji

Ingénieur Statisticien Economiste

Assistant de recherche à l'Ecole Nationale Supérieure de
Statistique et d'Economie Appliquée (ENSEA) d'Abidjan

e-mail : ankouvi@hotmail.com

Octobre 2009

Table des matières

Liste des figures	2
Liste des encadrés et exemple	2
I. Consignes d'installation	5
I.1. Configuration requise.....	5
I.2. Installation.....	5
I.3. Vue d'ensemble de CSPro au lancement	6
II. Création d'un masque de saisie	6
II.1. Définition d'un dictionnaire de variables.....	7
II.1.1 Création et définition des records	10
II.1.2 Création et définition des items (questions ou variables).....	11
II.1.3 Déclaration des valeurs légales.....	13
II.2. Génération du masque de saisie	16
II.3. La spécification des procédures de contrôle	18
II.4. Création de l'application de saisie	20
II.5. Quelques options.....	21
II.5.1 Création de liste de choix	21
II.5.2 Génération de numéro séquentiel et utilisation de certaines fonctions.....	21
III. Fusion des fichiers et redressement des données	22
III.1. Fusion de fichier	22
III.2. Apurement des données saisies par l'application Batch	23
III.2.1 Correction d'une information erronée.....	24
III.2.2 Redressements et corrections des données.....	25
III.3. Exportation de fichier de données.....	25
III.4. Les options d'exportation de fichier de données.....	26
IV. Création de tableaux sous CSPro	29
ANNEXES	33

Liste des figures

Figure 1 : assistant permettant la création d'une nouvelle application ou l'ouverture d'une application existante.....	6
Figure 2 : exemple de présentation schématique de questionnaires (ou données) hiérarchiques.....	7
Figure 3 : Création d'un dictionnaire de variable avec l'aide de l'assistant.....	8
Figure 4 : support du dictionnaire des variables.....	9
Figure 5 : Création d'une application de saisie avec l'aide de l'assistant.....	10
Figure 5 : procédure d'ajout d'un record.....	11
Figure 6 : procédure d'ajout d'un item.....	11
Figure 7 : Définition des modalités d'une variable.....	14
Figure 8 : Déclaration des modalités associées à une variable.....	15
Figure 9 : exemple de saisie de données.....	20
Figure 10 : module de fusion de données.....	22
Figure 11 : création de l'application BATCH.....	23
Figure 12 : module d'exportation des données vers (SPSS, SAS, STATA...).....	26
Figure 13 : création de tableaux de données.....	29

Liste des encadrés et exemple

Exemple 1: Définition des valeurs légales et création des classes.....	14
Encadré 1 : définition de données hiérarchiques et données individuelles.....	7
Encadré 2 : Caractéristiques d'un record :	11
Encadré 3 : Création d'un identifiant pour les répondants.....	12

Remerciements

La réalisation du travail qui a abouti à la rédaction de ce document, qui est en réalité un résumé, a nécessité la participation de nombre de personnes envers qui nous tenons à manifester notre reconnaissance.

Nous adressons nos remerciements à notre collègue et aîné Innocent GAMLI, fonctionnaire au bureau du PAM/Sénégal, qui a semé en nous la passion pour CSPro et qui nous a aidé à faire nos premiers pas dans la conception de masque de saisie sous CSPro en 2004.

Nous remercions Monsieur KOFFI N'GUESSAN, DIRECTEUR de l'ENSEA, pour la confiance qu'il met en nous et pour le cadre de travail qu'il nous offre afin que les activités de recherche se déroulent dans de bonnes conditions.

Nous remercions aussi Messieurs KOUADIO Kouassi Hugues et KOUAKOU N'goran Jean Arnaud directeurs des études à l'ENSEA pour leur encadrement dans tous nos travaux.

C'est le moment pour nous, de remercier Monsieur HANDLEY Josh, consultant au UNFPA/Rwanda dont les critiques nous ont aidé à mettre à jour ce document dans le cadre de l'Atelier sous Régional de Formation sur le logiciel CSPro qui s'est tenu à l'ENSEA du 05 au 23 Octobre 2009.

Introduction

Ce guide est une initiation à l'utilisation du logiciel CSPro dont les applications les plus utilisées seront présentées. Votre curiosité et votre travail personnel vous permettront d'approfondir vos connaissances. Cette initiation ne représente que 10% des capacités de ce logiciel. Ce guide a été rédigé sur la base du HELP de CSPro.

CSPro signifie « Census and Survey Processing System » ou « Système de traitement des données issues des questionnaires de recensement ». C'est un logiciel conçu par le Bureau du recensement des États-Unis d'Amérique. Il est destiné à **la saisie et au traitement des données issues d'enquêtes et de recensements**. Il puise ses origines du logiciel CENTS. Il incorpore des caractéristiques de IMPS, ISSA et EpiInfo. En définitif, il remplace aussi bien IMPS que ISSA. Il contient plusieurs modules exécutables à savoir :

- Application designer (CSPRO.EXE)- création des applications de saisie, application batch (pour apurement) et production des tableaux simples ou croisés,
- Data Entry System (CSENTRY.EXE)

Comme tout logiciel de saisie de données, il permet **plusieurs niveaux de contrôle de cohérence de l'information saisie par les limitations de champs, les contrôles calculés, les sauts automatiques...** De plus ce logiciel permet à tout moment de produire un rapport d'étape permettant d'évaluer la qualité et l'avancement de la saisie ainsi que les erreurs intervenues à ce stade.

Il présente plusieurs avantages :

- ◆ facile d'utilisation : avec une aide bien détaillée et agrémentée d'exemples ;
- ◆ traitement des fichiers hiérarchiques ;
- ◆ gratuit ; ce logiciel, ainsi que sa documentation, sont gratuits en ligne, quoiqu'une inscription en ligne puisse être exigée¹.
- ◆ complet : il propose tous les contrôles et corrections relatifs à la saisie et aux phases postérieures à la saisie (redressements, tabulation des données, etc.).

¹ Des détails peuvent être obtenus du Bureau de recensement des Etats Unis
<http://www.census.gov/ipc/www/cspro/index.html>.

- ♦ Le transfert des données saisies vers un logiciel de traitement est souple, sans complication et facile à mettre en œuvre. Le transfert n'implique pas de travaux supplémentaires tels l'introduction des dictionnaires de variables (recodification) et la labellisation des variables. CPro offre ses fonctionnalités de transfert vers SPSS, STATA, SAS ...

Quatre applications de CPro seront présentées dans ce document à savoir :

- création d'un dictionnaire de variable sous CPRO (Data Dictionary) ;
- génération du masque de saisie et contrôle de cohérence (Data Entry Applications)
- apurement des données saisies (Batch Edit Applications)
- La phase d'analyse sera sommairement abordée (Cross Tabulation Applications).

Comme tout logiciel, Cpro requiert un certain nombre de consignes d'installation qui sont présentées ci-dessous.

I. Consignes d'installation

I.1. Configuration requise

Le logiciel CPro tourne sur un minimum de matériels relatifs à la version du système d'exploitation installé sur l'ordinateur, à la vitesse du processeur, à taille du disque dur...

Pour plus de détails, consultez le site : www.census.gov/ipc/www/cspro.

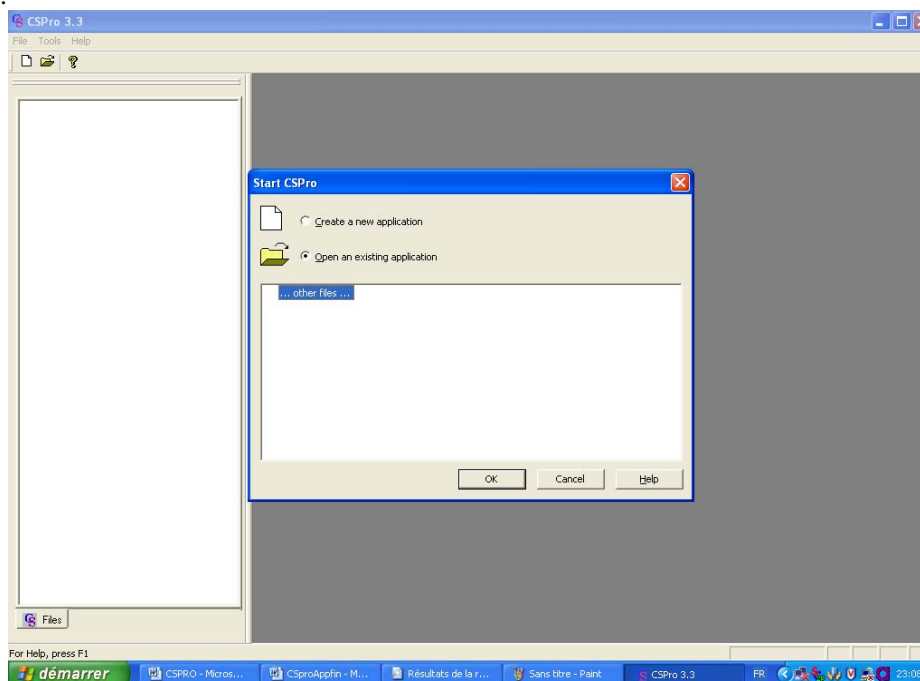
I.2. Installation





CSPRO version 4.0 est disponible sur Internet sous le site www.census.gov/ipc/www/cspro et tient aisément sur une clé USB de 64 Mo. Il est techniquement commercialisé par le département International Programs Center du Bureau of census U.S

I.3. Vue d'ensemble de CSPro au lancement

Au lancement du logiciel, la page ci-dessous s'affiche suivie généralement de la boîte de dialogue en avant plan. Cette dernière permet soit d'accéder à un fichier existant («Open an existing application») ou d'en créer un nouveau («Create a new application »).

Figure 1 : assistant permettant la création d'une nouvelle application ou l'ouverture d'une application existante.



La barre de menu donne accès aux menus ci-après **File Tools Help**. Le menu «File» permet d'atteindre les boutons   et permet de quitter (fermer) l'application CSPro. Ces boutons que l'on retrouve dans la barre d'outils permettent soit d'ouvrir  un fichier existant ou d'en créer  un nouveau.

II. Création d'un masque de saisie

La création d'un masque de saisie sous CSPro implique l'exécution de 4 étapes précises qui sont :

- La définition d'un dictionnaire de variables (ou données) ;
- La génération du masque de saisie ;
- La spécification des procédures de contrôle ;

- La création de l'application de saisie ;

II.1. Définition d'un dictionnaire de variables.

Dans cette section, on abordera d'une part les dictionnaires relatifs aux «questionnaires dits individuels» et d'autre part les dictionnaires de données hiérarchiques.

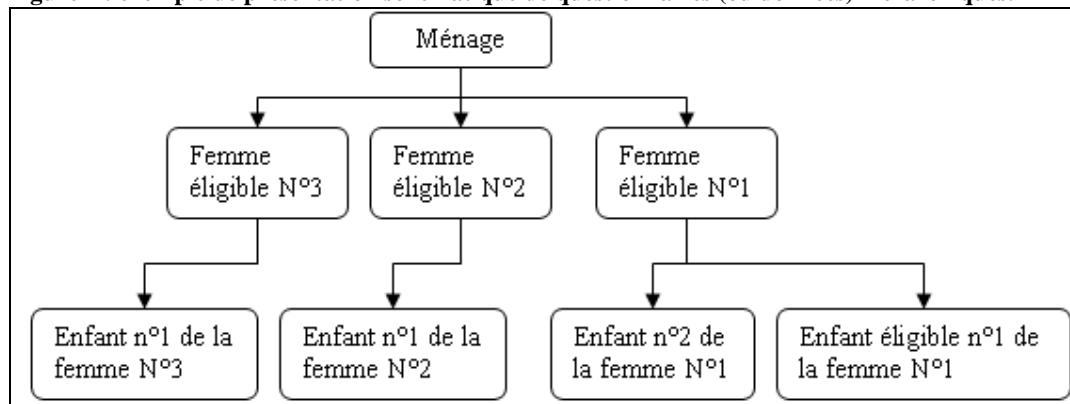
Encadré 1 : définition de données hiérarchiques et données individuelles

Un questionnaire sera dit hiérarchique s'il permet de prendre plusieurs niveaux d'information. Considérons l'exemple d'une enquête démographique de santé (EDS). Supposons que, dans cette enquête, l'on s'intéresse aux ménages. Pour chacun d'entre eux, on renseigne certaines de leurs caractéristiques (taille, sexe du chef de ménage, son âge, type de logement occupé ...) sans oublier de s'intéresser à ses membres. On interviewe ensuite toute femme ayant au moins 12 ans révolus. On enregistre enfin les enfants, ayant moins de 5 ans révolus, de la femme éligible. On dispose ainsi de trois questionnaires qui sont :

- un questionnaire ménage,
- un questionnaire femme éligible
- enfin d'un questionnaire enfant éligible.

Ce type d'organisation est dit hiérarchique. On parle alors de données hiérarchiques.

Figure 2 : exemple de présentation schématique de questionnaires (ou données) hiérarchiques.



A l'opposé des données hiérarchiques, les données ou questionnaires seront dits individuels lorsque l'information est prise sur un seul et même niveau. Par exemple une enquête relative aux habitudes alimentaires des étudiants de l'ENSEA (on suppose dans le cas d'espèce que l'information prise concerne les caractéristiques de l'étudiant « âge, sexe, nationalité, le statut matrimonial » et ses habitudes alimentaires).

Supposons que l'on dispose d'un questionnaire (dit individuel) dont on veut créer le masque de saisie. Avant de se lancer dans une telle entreprise, il faut au préalable faire une **analyse statistique du questionnaire** pour identifier clairement les unités d'observations, les sections, les différents niveaux d'observations et les variables. Sous CSPro, cette analyse consiste à la définition d'un dictionnaire de variables. Le dictionnaire des variables décrit l'organisation de celles-ci tout en précisant la nature, le type et la taille **des records**² et **des variables** (libellé, type, modalités, taille, etc....)

Le dictionnaire est le socle des différentes étapes et travaux qui seront effectués sur les données saisies. Il convient donc de le concevoir en ayant à l'esprit les différentes analyses que l'on voudrait faire des données.

Deux approches peuvent être utilisées pour créer l'application de saisie sous CSPro.

La première approche est présentée ci-dessous.

Approche 1


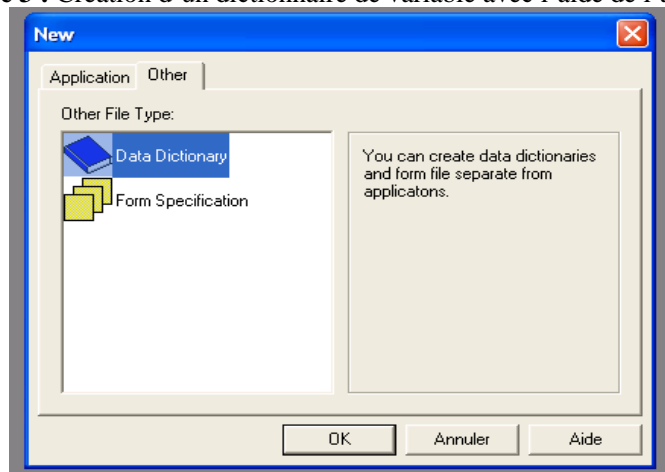
Après la mise en route du logiciel, la création du dictionnaire de variables est possible en cliquant sur le bouton  dans la barre d'outils ou sur l'option « Create a new application » dans la boîte de dialogue en avant plan sur la **figure N°1**. Cette action affiche la boîte de dialogue ci-dessous.

Figure 3 : Création d'un dictionnaire de variable avec l'aide de l'assistant

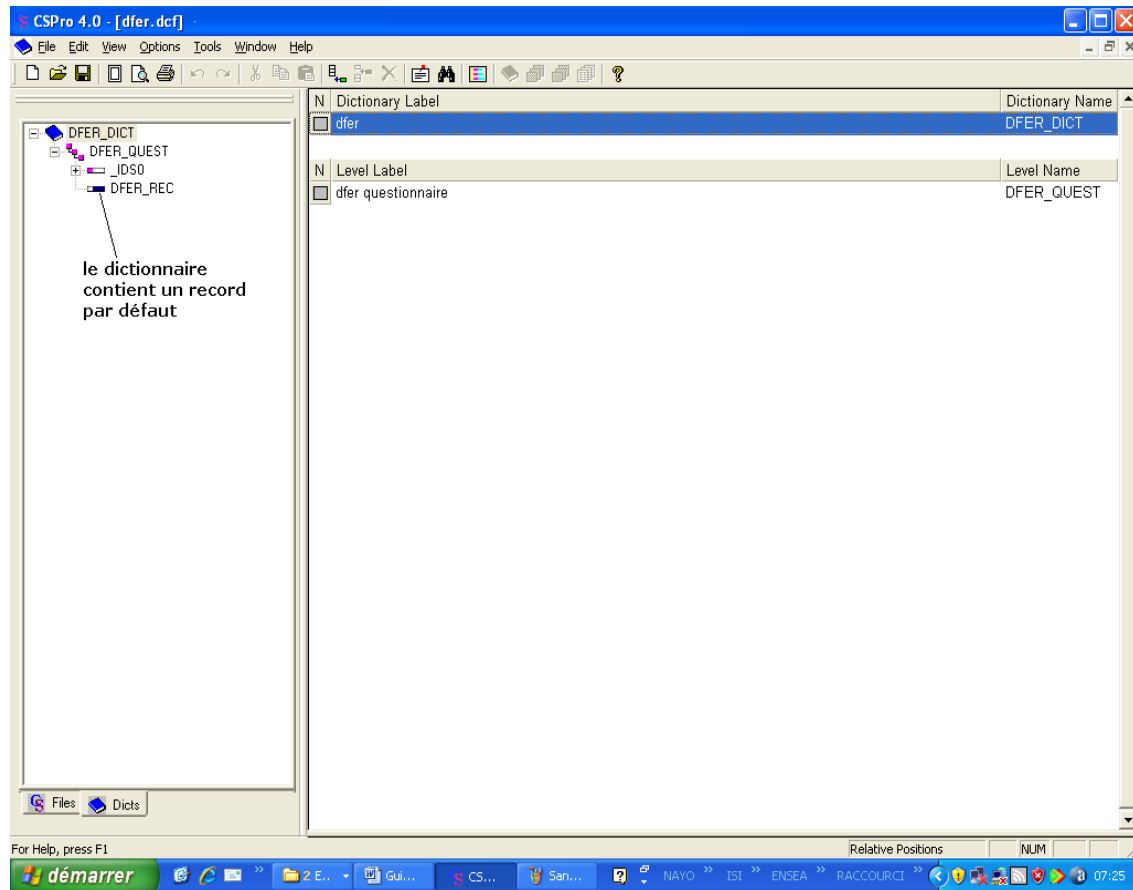


² **Un record** correspond à une section du questionnaire. Il permet d'enregistrer les réponses relatives aux questions traitant d'un thème commun.

Le type d'objet correspondant au dictionnaire de variables est «**Data Dictionnaire**» (sélectionnez le et cliquez ensuite sur Ok). Il convient maintenant de choisir le chemin d'accès³, de donner un nom au fichier, puis de cliquer sur « Create/créer ». (Le dictionnaire est un fichier de format (extension) « dcf »)

Cette dernière action donne accès à l'environnement de travail suivant :

Figure 4 : support du dictionnaire des variables

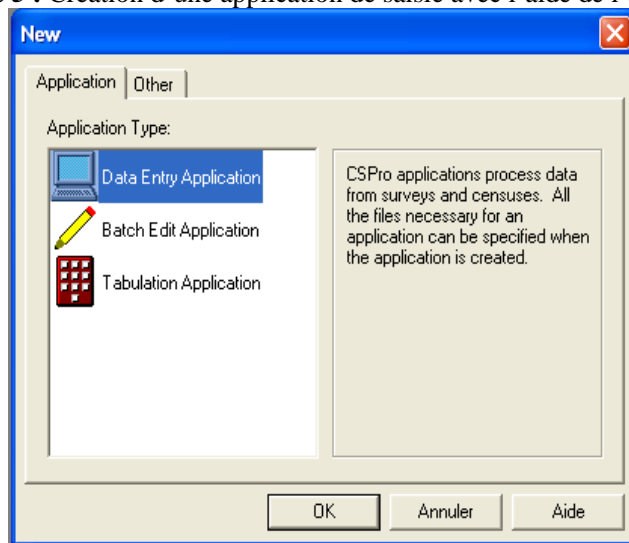


Approche 2

Dans cette approche, on choisit plutôt l'application de saisie « Data entry application ». Un clic sur « Ok » donne accès à une autre boîte de dialogue permettant de créer le dictionnaire. Dans ce cas, le dictionnaire porte par défaut le nom de l'application de saisie est en cours de création. La démarche sera présentée en détail dans la suite.

³ Chemin : Il s'agit du chemin donnant accès au répertoire ou au dossier dans lequel l'on désire enregistrer le fichier.

Figure 5 : Création d'une application de saisie avec l'aide de l'assistant



Le dictionnaire présenté à la figure 4 est le dictionnaire initial (à la création), *il contient, par défaut, un records* et peut être modifié en fonction de la structure du questionnaire. Le premier élément «**Id items**» n'est destiné qu'à contenir l'identifiant⁴ du questionnaire. «**Id items**» *appartient à tous les records du questionnaire*. Le second élément qui est un record est vide.

Il faut noter que la création de dictionnaire passe d'abord par la création des records et celle des items (variables) vient ensuite.

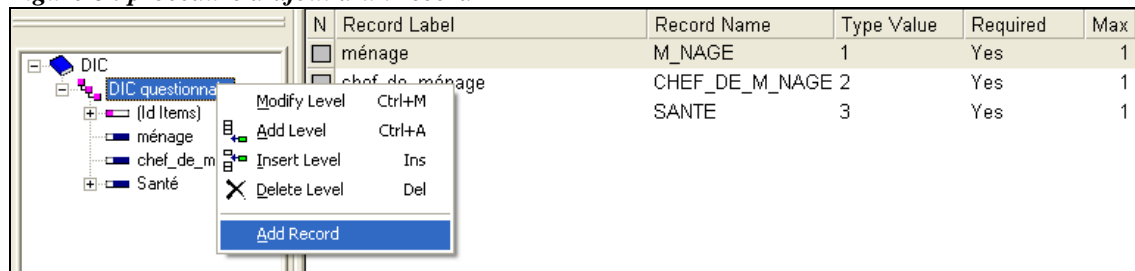
II.1.1 Création et définition des records

Comme dit plus haut, un record correspond à une section du questionnaire. Il permet d'enregistrer les réponses relatives aux questions (items) traitant d'un thème commun. Par conséquent, si le questionnaire est divisé en plusieurs sections, il est préférable de créer autant de records que de sections (*Cela n'est pas obligatoire*).

Pour ajouter de nouveau records, faites un clic droit, dans la fenêtre de gauche, sur le bouton en forme d'escalier à 3 niveaux ou sur n'importe quel record puis cliquer sur «**Add Record**». Par cette même démarche, on a aussi la possibilité de modifier ou de supprimer un record. Un record est caractérisé par des propriétés qui sont présentées dans le tableau ci-dessous.

⁴ Chaque questionnaire ou enregistrement doit être marqué par un identifiant unique et sans redondance.

Figure 6 : procédure d'ajout d'un record



Encadré 2 : Caractéristiques d'un record :

Record Label : il s'agit du libellé que l'on désire donner au record. Ce Nom permet d'identifier le record. Il est utilisé par défaut dans le masque de saisie

Record Name : Il correspond au nom logique du record. Il sera utilisé le cas échéant dans les procédures de contrôle pour faire référence à ce dernier. Il est composé d'au plus 32 caractères. Les caractères possibles sont : A-Z, 0-9, et le _.

Remarque : la première position est réservée pour les lettres de A à Z. Par ailleurs, la dernière position ne saurait être le _. Il faut noter pour finir que CSPro possède des noms réservés⁵ qui ne peuvent pas être utilisés au niveau du dictionnaire.

Type value : C'est un alphanumérique qui permet d'identifier les lignes du fichier des données qui représentent le record.

Required : prend deux valeurs possibles : Yes/No. Il s'agit de dire si pour chaque questionnaire, on a au moins une occurrence de ce record. Dans le cas des conditions d'éligibilité on mettra la valeur No pour cet attribut.

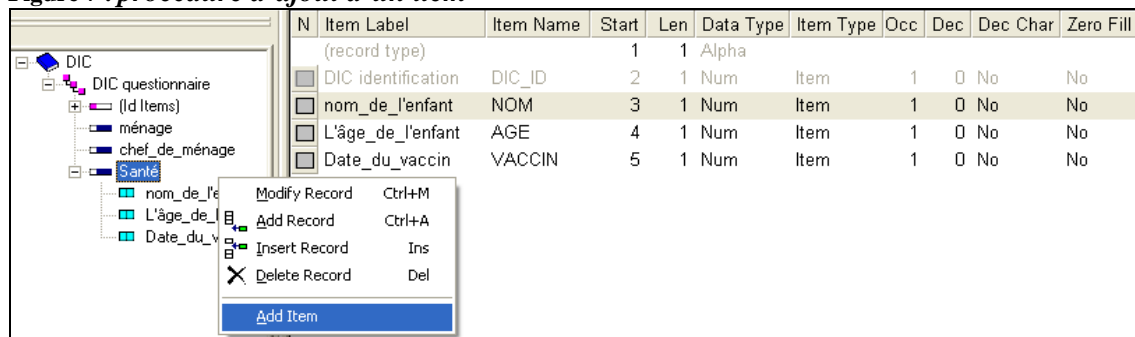
Max : Précise le nombre maximum d'occurrences possibles pour une section donnée. Si ce nombre est supérieur à 1, on pourra générer des **rosters (tableaux)** au niveau du masque de saisie. (Confère exemple)

II.1.2 Création et définition des items (questions ou variables)

Une fois les records créés, il ne reste plus qu'à ajouter/créer les **items** (questions/variables) qui composent cette section (à l'intérieur de chaque **record**). Pour ajouter un item dans un record, faites un clic droit sur ce record, dans la fenêtre de gauche, et cliquer sur «**Add Item**».

L'environnement de travail prend alors l'apparence ci-dessous.

Figure 7 : procédure d'ajout d'un item



⁵ Voir annexes

Ce menu contextuel permet entre autre de supprimer ou modifier un item.

Avant la création des items, il convient de définir les identifiants. Un identifiant peut être composé d'une succession de variables. Nous avons ci-dessous un exemple de création d'identifiant.

Encadré 3 : Création d'un identifiant pour les répondants

Considérons toujours notre EDS avec 3 niveaux d'information donc 3 types de questionnaires.

- questionnaire ménage,
- questionnaire femme éligible
- questionnaire enfant éligible.

Intéressons nous seulement au questionnaire santé de l'enfant pour lequel l'on désire concevoir un masque de saisie. On définit à cet effet un dictionnaire de variable selon la démarche définie plus haut. Dans le dictionnaire, chaque enfant est indexé par :

- le numéro du ménage auquel il appartient,
- le numéro d'ordre de la femme éligible dont il est l'enfant,
- enfin son numéro d'ordre parmi les enfants de cette femme.

Ces trois composantes constituent pour chaque enfant un identifiant unique sans redondance.

Dans le record réservé à la création des identifiants⁶, il convient de créer ces 3 variables (items) :

- numéro du ménage,
- numéro d'ordre de la femme éligible,
- numéro d'ordre de l'enfant.

En général, on créera dans ledit record autant de variables que de composantes de l'identifiant.

Précisons que chaque item est caractérisé par des propriétés qui doivent forcément être renseignées. Les propriétés des items sont différentes de celles des records. On ne s'étendra pas ici sur les propriétés **Item label** et **Item Name** qui rejoignent le cas record label et record Name.

⁶ On peut remarquer, dans le cadre de ce guide, que ce record est nommé «Id Items»

Start	Champ très important, précise dans fichier la position à partir de laquelle la donnée saisie sera stockée. Il est incrémenté par défaut. Dans le cas des SubItems (à voir plus loin) il doit être modifié.
Len	Indique la taille du champ que va occuper la donnée saisie en terme de position
Data Type	Identifie le type de la variable du champ en cours (numérique ou alphanumérique)
Item Type	Par défaut, ce champ prend la valeur Item . Il existe des questions dont les réponses peuvent être subdivisées en sous réponses. L'exemple typique c'est celui de la date qui est composé du jour, du mois et de l'année. On peut demander à CSPRO de gérer le jour, le mois et l'année comme des Subitems de la variable date. Lorsqu'on choisit Subitem, il faut modifier les propriétés Start et Len de telle sorte que la position du Subitem coïncide avec la partie de l'item qu'il représente.
Occ	permet de préciser le nombre d'occurrences de l'item dans le record. Si l'on décide, par exemple, que cette valeur vaut 5 pour une variable dont la donnée est saisie sur deux (02) positions (Len), CSPRO va réserver 10 positions pour la saisie des informations sur ce champ (5*2=10). Si la variable est à modalités exclusives, le nombre d'occurrences est égal à 1. Si elle est polytomique ⁷ à réponses non exclusives, le nombre d'occurrence correspond au nombre de réponses possibles
Dec	Fait référence au nombre de décimale d'une variable de type numérique.
Dec Char	Vaut «Yes » ou «No ». Il précise pour les variables numériques avec décimale s'il faut une virgule pour séparer la partie entière de la partie décimale.
Zero Fill	Prend deux valeurs possibles : Yes ou No. Lorsque la valeur choisie est Yes, CSPRO complète (le cas échéant) les saisies par des zéros à gauche.

II.1.3 Déclaration des valeurs légales

Dans cette partie, nous apprenons à définir les valeurs légales afin d'une part d'éviter la saisie des valeurs non admises et d'autre part de créer des classes de valeur. Les exemples ci-dessous nous permettrons de mieux appréhender la notion de valeurs légales.

⁷ On entend par variable polytomique une variable ayant plus de deux modalités de réponses. Les réponses ou modalités sont dites non exclusives lorsque l'on a la possibilité de choisir plus d'une modalité à la fois

- Concernant les variables dont les modalités sont entièrement connues, il est préférable de définir des valeurs légales (modalités ou si possible des codes associés à ces modalités).
- Pour les besoins de l'analyse, certaines variables numériques peuvent être regroupées en classes. C'est le cas, par exemple, de la variable âge. Au cours de l'interview, on suppose qu'on a donné aux enquêteurs l'instruction de noter l'âge en année révolu du répondant. Cependant on est aussi intéressé par la répartition des individus par groupe d'âge quinquennaux.

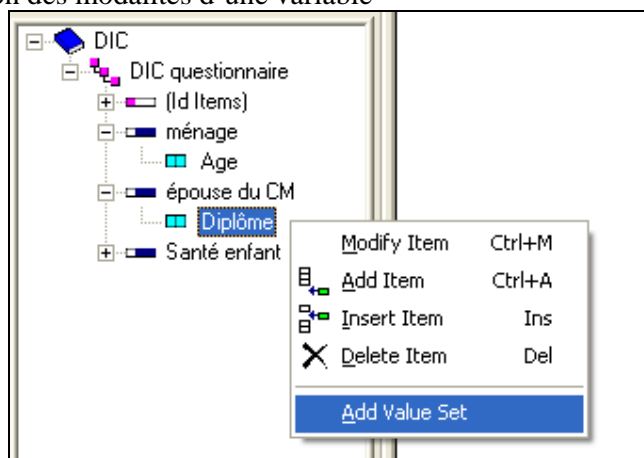
Exemple 1: Définition des valeurs légales et création des classes

Durant notre E.D.S, des informations en été prises sur l'âge et le niveau d'instruction des répondants. Dans cet exemple, le niveau d'instruction est représenté par la variable « DIPLOME » et l'âge par la variable « AGE ».

Pour chaque individu l'âge inscrit sur le questionnaire est l'âge en année révolue. Quant au niveau d'instruction on s'est plutôt intéressé au dernier diplôme obtenu par le répondant au cours de son cursus scolaire. Les diplômes retenus sont le CEPE, le BEPC et le BAC (Pour des raisons liées à l'étude, tout individu ayant un diplôme supérieur au BAC sera classé parmi ceux qui ont le bac).

Pour les besoins de l'analyse, les répondants seront classés par groupe d'âges quinquennaux. A partir du dictionnaire des variables, sur la variable (item) concernée (DIPLOME par exemple), faites un clic droit. Le menu contextuel s'affiche comme le montre la figure ci-dessus et cliquer sur « Add Value Set » pour définir les valeurs légales.

Figure 8 : Définition des modalités d'une variable



Cette dernière action donne accès à une page où est inscrit par défaut le nom de la variable concernée. Appuyez deux fois sur « entrer/enter », saisissez ensuite les différentes modalités en prenant le soin d'attribuer à chaque modalité le code correspondant. Ce code sera utilisé pour la saisie des données. La figure 8 (ci-dessous) en donne une bonne illustration.

Value Label : doit être renseigné par le libellé de la modalité concernée

From : doit être renseigné par le code correspondant à la modalité. En outre, pour les variables numériques dont les modalités sont en regroupés en classe, on saisit dans le champ « From », la borne inférieure

To : est réservé à la borne supérieure

Figure 9 : Déclaration des modalités associées à une variable

N	Value Set Label	Value Set Name	Value Label	From	To	Special
	Diplôme	DIPL_ME_VS1				
			CEPE	1		
			BEPC	2		
			BAC			

Value Set Label	Value Set Name	Value Label	From	To	Special
AGE	AGE_VS1				
		<5 ans	0	4	
		5ans à 9 ans	5	9	
		10 ans à 14 ans	10	14	
		15 ans à 19 ans	15	19	
		20 ans à 24 ans	20	24	

Le code associé à la modalité CEPE est «1». Ainsi au cours de la saisie pour tout répondant ayant le niveau CEPE, on saisira le code «1».

En ce qui concerne la variable âge, tout répondant ayant plus de 5 ans révolus et moins de 10 ans révolus (par exemple 7 ans) sera dans la classe « 5 ans à 9 ans ». Ainsi au cours de la saisie, on saisira 7 ans pour l'individu concerné et il sera automatiquement considéré dans ladite classe.


Remarque

Par ailleurs, pour les variables numériques regroupées en classe, les valeurs légales peuvent être générée. Pour y arriver, il suffit de faire un double clic sur la variable concernée, positionner le curseur de la souris dans l'environnement de travail de droite. Faites ensuite

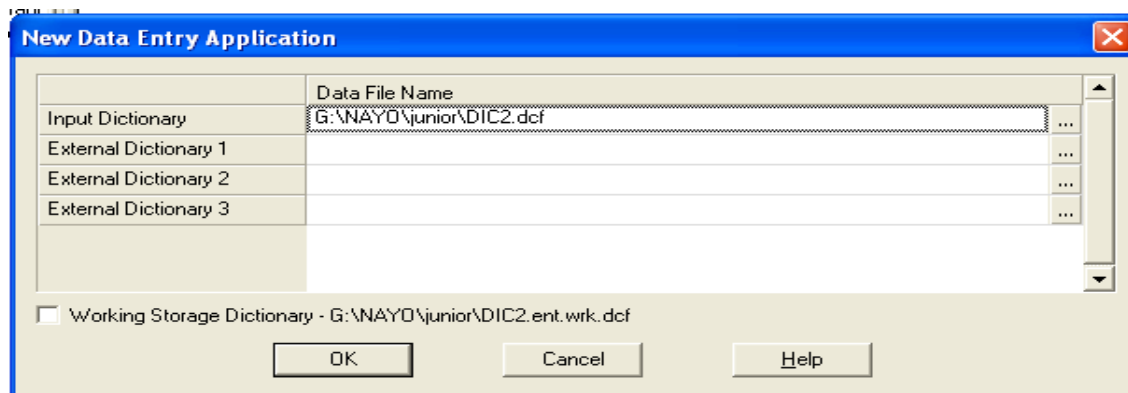
un clic droit. Dans le menu contextuel qui s'affiche, cliquer sur «Generate Value Set», suivez les instructions de la boîte de dialogue qui s'affichera et le tour est joué.


Le dictionnaire est fin prêt et la génération du masque de saisie est enfin possible.

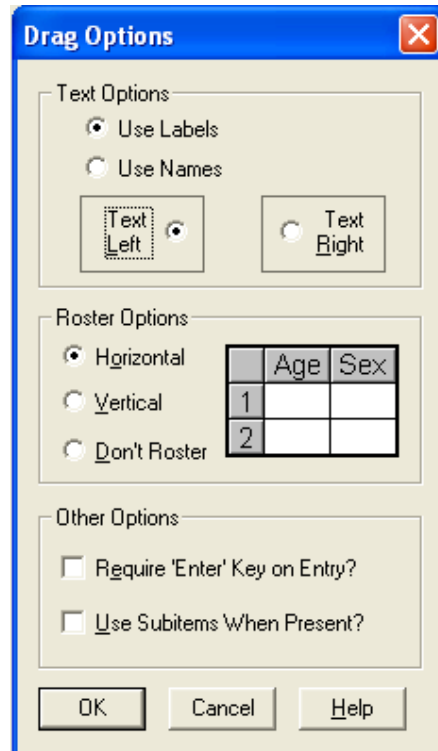
II.2. Génération du masque de saisie

Le masque de saisie se génère, à partir du dictionnaire des variables. Une fois le dictionnaire achevé, faire : **File/New** . Ensuite au niveau de la boîte de dialogue qui s'affiche, choisir comme type d'objet «**Data Entry Application**», qui permet de créer un fichier d'extension «ent».

Puis cliquer sur OK. A cette étape, il faut choisir le dossier où l'on aimerait enregistrer le fichier, préciser le nom du fichier et cliquer enfin sur « Create/Créer ». Cette dernière action conduit à l'affichage de la boîte de dialogue ci-dessous qui permet de choisir le dictionnaire correspondant.



Il est important de choisir le dictionnaire adéquat. Si le chemin d'accès indiqué en face de «**Input Dictionnary**» n'est pas le bon, il vous revient de trouver le bon dictionnaire en cliquant sur le bouton  en face du chemin d'accès. Une fois le dictionnaire adéquat choisi, cliquer sur OK. Un message s'affiche vous demandant si vous désirez générer le masque de saisie à partir du dictionnaire sélectionné, cliquer sur «Oui». L'étape suivante est l'affichage de la boîte de dialogue suivante :



Cliquer enfin sur Ok et le masque de saisie est généré.

Text Options	Permet de choisir le nom (Label ou Name) de la variable qui figurera dans le masque de saisie et l'option Text left et Text Right aide à disposer le Label/Name par rapport au champ de saisie (à droite ou à gauche)
Roster Options	L'option conduit à une présentation très conviviale des records sous forme de tableau lorsque ceux-ci ont plus d'une occurrence. Selon la disposition recherchée, choisissez horizontale, verticale ou pas de Roster du tout.
Require « Enter » Key on Entry	L'option ci-avant, si elle est sélectionnée, impose que la saisie de chaque variable soit validée avant de passer à un autre champ.
Use Subitems When present	Lorsqu'elle est cochée, cette option permet d'affiche dans le masque de saisie les Subitems . l' item pourra être déduit comme étant la concaténation de ses Subitems.

L'expérience montre que très souvent, le premier masque de saisie généré est loin d'être acceptable. « Pour la convivialité du travail et dans l'esprit de limiter les erreurs dues à la saisie, il convient de faire une mise en forme du masque de saisie (selon le besoin afin qu'il ressemble parfaitement au questionnaire administré) et d'introduire des contrôles de cohérence en respectant les règles de la programmation sous CSPro. »

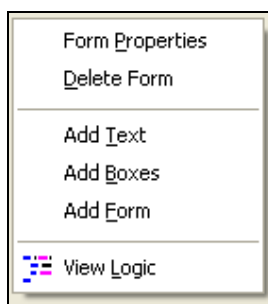
Quelques options de mise en forme


Champ-miroir :

Des champs miroir peuvent être introduits. Un champ-miroir est une variable qui est présente dans un écran de saisie sans qu'on puisse la modifier. Les identifiants sont très souvent placés comme champs-miroirs. Pour créer un champ miroir, on fait glisser la variable concernée du dictionnaire des données vers l'emplacement souhaité à l'écran de saisie.

Texte et encadré

Des textes peuvent être ajoutés dans l'écran de saisie afin de rendre le masque de saisie très convivial. L'on peut aussi tracer des traits ou dessiner des encadrés. Un clic droit sur l'écran de saisie donne accès au menu contextuel ci-dessous. **Add Text** pour ajouter un texte et **Add Boxes** pour tracer un trait ou dessiner un encadré.






<i>Form Properties</i>	Permet de modifier le Label, le nom et la couleur du Formulaire
<i>Delete Form</i>	Permet de supprimer le Formulaire
<i>Add Text</i>	Aide à ajouter des textes dans le Formulaire
<i>Add Boxes</i>	Aide à ajouter des encadrés, des traits
<i>Add Form</i>	Permet d'ajouter des Formulaires
<i>View Logic ou le bouton</i> 	affiche le module de programmation des procédures de contrôle à la saisie.

II.3. La spécification des procédures de contrôle

Une fois la mise en forme du masque de saisie achevée, il convient de spécifier les différents contrôles de saisie.

La saisie contrôlée permet de minimiser les erreurs de saisie. Les vérifications automatiques prennent en compte *les contrôles de validité de code, les amplitudes des valeurs, les cohérences avec d'autres variables du même questionnaire*. La vitesse de la saisie est

inversement proportionnelle au nombre de contrôles. Afin de ne pas ralentir (ou même de bloquer) la saisie, leur nombre est souvent volontairement limité.

Sous CSPRO, le bouton ci-après  dans la barre d'outils permet d'accéder à l'environnement de travail relatif à la rédaction des procédures de contrôle. Ces deux icônes   permettent de basculer tour à tour de l'écran de saisie à la zone de rédaction des procédures de contrôle. Le langage de programmation sous CSPRO est assez proche du Pascal. Chaque objet (record, roster, item, etc.) sur lequel on désire écrire une procédure de contrôle doit avoir un **preproc** et/ou un **postproc**. Le preproc permet de rédiger les contrôles que l'on souhaite effectuer avant la saisie de toute information concernant l'objet. Le preproc est exécuté avant l'apparition du curseur dans le champ. Dans certains cas, le preproc peut être remplacé par « onfocus ». Le « onfocus » permet d'exécuter une instruction dès que le champ de la variable est actif. Le postproc quant à lui permet de spécifier les actions à exécuter après la saisie dans le champ.

Toute procédure commence par le mot clé PROC suivi du nom de l'objet sur lequel elle s'applique. Exemple :

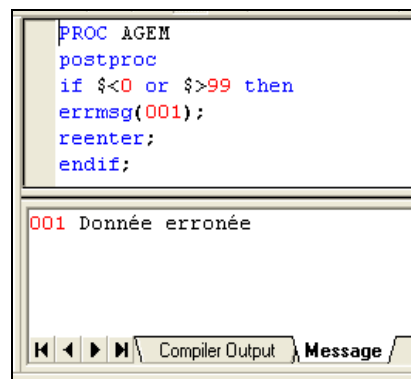
```
PROC AGE
  Preproc
  Instructions A
  Postproc
  Instructions B
```

C'est à travers quelques exemples que l'on va présenter les notions de programmation sous CSPRO.

Exemple 1 : Voici une procédure qui affiche le message « donnée erronée » et impose une nouvelle saisie si la valeur saisie n'est pas dans plage prédéfinie (l'âge du répondant doit être un nombre positif inférieur ou égal à 99).

```
PROC AGEM
  postproc
  if $<0 or $>99 then
    errmsg(001) ;
    reenter ;
  endif ;
```

Dans la fenêtre message, taper : **001 Donnée erronée**




Le caractère \$ fait référence à l'information qui a été saisie. La fonction **reenter** maintient le curseur dans le champ en cours lorsque la donnée saisie est hors de la plage. On remarque par ailleurs que chaque instruction se termine par un « ; ».

Exemple 2 : traitons d'un cas où on fait un saut à la question DIPLOME si l'âge est supérieur ou égal à 12.

```
PROC AGEM
  postproc
  if $>=12 then
    skip to DIPLOME;
  endif ;
```

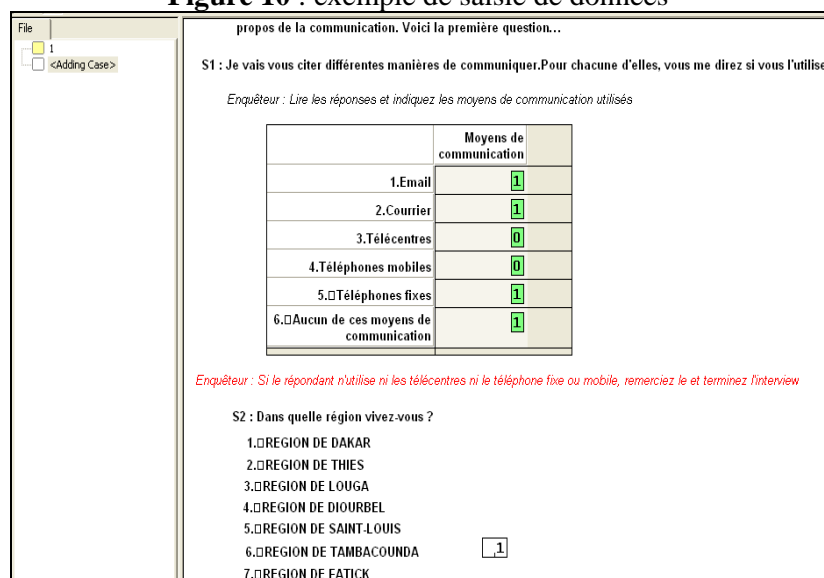
Pour approfondir le langage de programmation sous CSPro, confère l'aide sur le logiciel.

II.4. Création de l'application de saisie

Les contrôles de cohérence étant spécifiés, le masque est prêt pour la saisie. Toutefois, la saisie se fait dans un autre environnement. En effet, le masque de saisie n'est pas le support de saisie, il constitue juste une interface permettant de saisir les données. Pour générer le fichier de données, il suffit de cliquer sur le bouton ci-après  (RUN) dans la barre d'outils et de spécifier ensuite son nom. Ce nom doit porter obligatoirement l'extension « **dat** » (exemple : donnée.dat).

CSpro ne permet pas un accès direct aux données saisies. Ces dernières sont enregistrées dans le fichier de données de format dat. A la fin de la saisie, les données peuvent être transférées vers SAS, STATA et SPSS...

Figure 10 : exemple de saisie de données



propos de la communication. Voici la première question...

S1 : Je vais vous citer différentes manières de communiquer. Pour chacune d'elles, vous me direz si vous l'utilisez

Enquêteur : Lire les réponses et indiquez les moyens de communication utilisés

	Moyens de communication
1.Email	1
2.Courrier	1
3.Télécentres	0
4.Téléphones mobiles	0
5.☐Téléphones fixes	1
6.☐Aucun de ces moyens de communication	1

Enquêteur : Si le répondant n'utilise ni les télécentres ni le téléphone fixe ou mobile, remerciez-le et terminez l'interview

S2 : Dans quelle région vivez-vous ?

1.☐REGION DE DAKAR
2.☐REGION DE THIES
3.☐REGION DE LOUGA
4.☐REGION DE DIOURBEL
5.☐REGION DE SAINT-LOUIS
6.☐REGION DE TAMBACOUNDA
7.☐REGION DE FATICK

1

II.5. Quelques options

II.5.1 Création de liste de choix

CPro donne la possibilité de créer aussi des listes de choix. Considérons la variable Situation matrimoniale (SITMAT) dont les différentes modalités sont : Célibataire (CEL) Marié (MAR) Non concerné (NC).

Pour créer une liste de choix pour cette variable, dans la zone de rédaction des contrôles de cohérences, il suffit d'écrire la ligne de programme ci-dessous avec la fonction « **accept** »

```
PROC SITMAT
preproc
y = 0;
do until y in 1:3
  y = accept("SITMAT", "CEL", "MAR", "NC");
enddo;
$ = y;
noinput;
```

à la place de accepte, la fonction « set attributes » peut être aussi utilisée

« set attributes (SITMAT, SEXE, EDUCATION) assisted on; »

II.5.2 Génération de numéro séquentiel et utilisation de certaines fonctions

Pour que les champs des variables numériques génèrent de manière automatique des numéros séquentiels, faites un clic droit sur le champ concerné et sélectionnez « **Propriété/Sequential** ».

Il est souvent nécessaire, dans les procédures de contrôle, d'utiliser les fonctions présentées ci-dessous

TOTOCC	Elle fait référence au nombre d'occurrences utilisés dans un tableau ou un formulaire
SOCCURS	
NOCCUR	S'il est exécuté dans un postproc, elle fait référence au nombre d'occurrences d'un tableau, d'un formulaire ou d'une section
CUROCC	Elle fait référence à l'occurrence en cour
COUNT	Compte le nombre d'occurrence effectivement renseignées.

La saisie de données terminée, on peut maintenant procéder à leur apurement. Cependant, si la saisie a été effectuée sur différente machine ou dans différents fichiers de données, il convient de les fusionner d'abord.

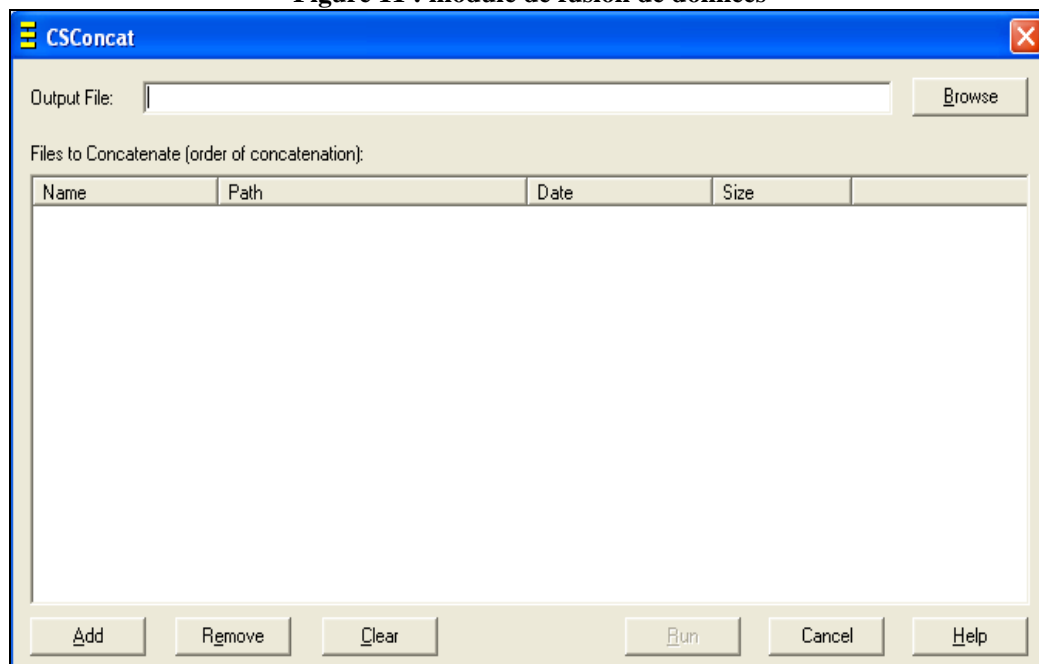
III.Fusion des fichiers et redressement des données

Si la saisie a été effectuée sur différentes machines, il convient de fusionner les différents fichiers afin de disposer d'un seul fichier.

III.1. Fusion de fichier

La fusion sous Csprou est possible grâce à la l'application **CSCConcat**. Pour accéder à cette dernière, dans la barre de menu, faites «*Tools/Concatenate Data*». Cette action, donne accès à la boîte de dialogue ci-dessous.

Figure 11 : module de fusion de données



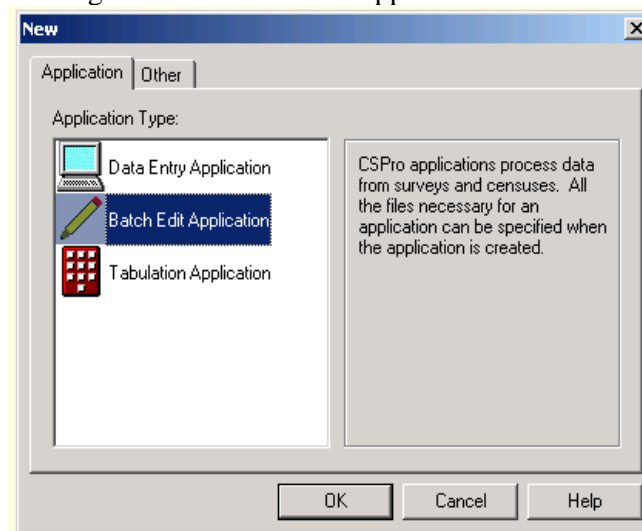
Add	Permet d'ajouter les fichiers de données que l'on désire fusionner
Remove	Permet de supprimer un fichier de données ajouté
Clear	Supprime tout les fichiers sélectionnés de l'environnement de travail
Run	Exécute la fusion des fichiers sélectionnés (par ajout d'observations)
Output file	Avant de lancer la fusion, il convient de spécifier le chemin d'accès et le nom du fichier de sortie

La fusion des fichiers se fait l'un à la suite de l'autre dans l'ordre de sélection. Le fichier de données unique étant disponible, on procédera à cette étape à l'apurement des données.

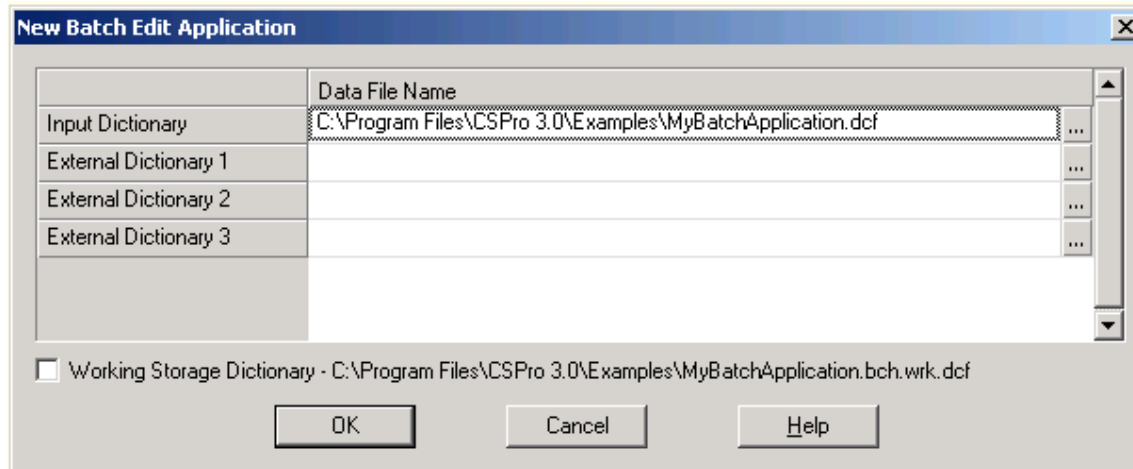
III.2. Apurement des données saisies par l'application Batch

La saisie contrôlée n'élimine pas la totalité des erreurs. En effet, certaines erreurs peuvent être introduites par l'agent de saisie par inadvertance et d'autres peuvent provenir du terrain et non détectées avant la saisie. Il convient donc d'apurer les données. Le module de contrôle et d'apurement des données est créé sur la base du dictionnaire des variables. Une fois la saisie des données achevée, faire : **File/New**. Ensuite, au niveau de la boîte de dialogue qui s'affiche, choisir comme type d'objet «**Batch Edit Application** ». Puis cliquer sur OK.

Figure 12 : création de l'application BATCH



A cette étape, il faut choisir le dossier où l'on aimerait enregistrer le fichier BATCH, préciser son nom et cliquer enfin sur « **Create/Créer** ». Cette dernière action conduit à l'affichage de la boîte de dialogue ci-dessous qui permet de choisir le dictionnaire correspondant.



Cliquer ensuite sur OK et un environnement de travail semblable à celui de la rédaction de contrôle de cohérence apparaît. Avant la correction d'une erreur, il convient de savoir quel type d'erreurs on est susceptible de commettre. Une fois l'application (Batch edit Application) créée, il faut maintenant rédiger les procédures d'apurement et d'imputation de manière analogue à la rédaction des contrôles de cohérence.

III.2.1 Correction d'une information erronée

Ceci est un bref exemple. Rappelons toutefois que certains utilisateurs ne conseillent pas de corriger (faire l'apurement des données sous CSPro)

On suppose que toute personne ayant moins de 12 ans est célibataire. Par conséquent on cherche à voir si un individu dont l'âge est inférieur à 12 ans révolus n'a pas été déclaré célibataire. Si l'on rencontre de tels cas, un message d'erreur doit être affiché pour le signaler (*errmsg*), l'enregistrement concerné est repéré par son identifiant (*write*) (*afin de retourner au questionnaire si possible*) et la situation matrimoniale de l'individu est corrigée (*impute*) (en espérant que l'âge enregistré est correct). Dans cet exemple, la modalité célibataire correspond au code 3.

```
PROC SITMAT
preproc
if AGEM<12 then
  if $<>3 then
    errmsg("situation matrimoniale %d erronée",$);
    write("ID %d",DIC_ID);
    impute($,3);
  endif;
endif;
```

Compile Successful

III.2.2 Redressements et corrections des données

CSPro offre aussi la possibilité de redresser ou de corriger les valeurs manquantes ou incohérentes par la méthode « hot-deck ». Elle consiste à affecter à une valeur manquante la modalité pour la même variable du dernier individu rencontré présentant des caractéristiques communes.

Après l'exécution de l'application BATCH, Un fichier texte est généré et donne un rapport des différentes erreurs rencontrées suites aux contrôles définies et celui des données éventuellement imputées.

Précisons que les redressements et corrections exécutés ne s'appliquent pas sur les données du fichier dans lequel elles ont été enregistrées. En effet, le fichier de données saisies reste intact. Les données corrigées sont enregistrées dans le *outfile* généré après l'exécution de l'application BATCH.

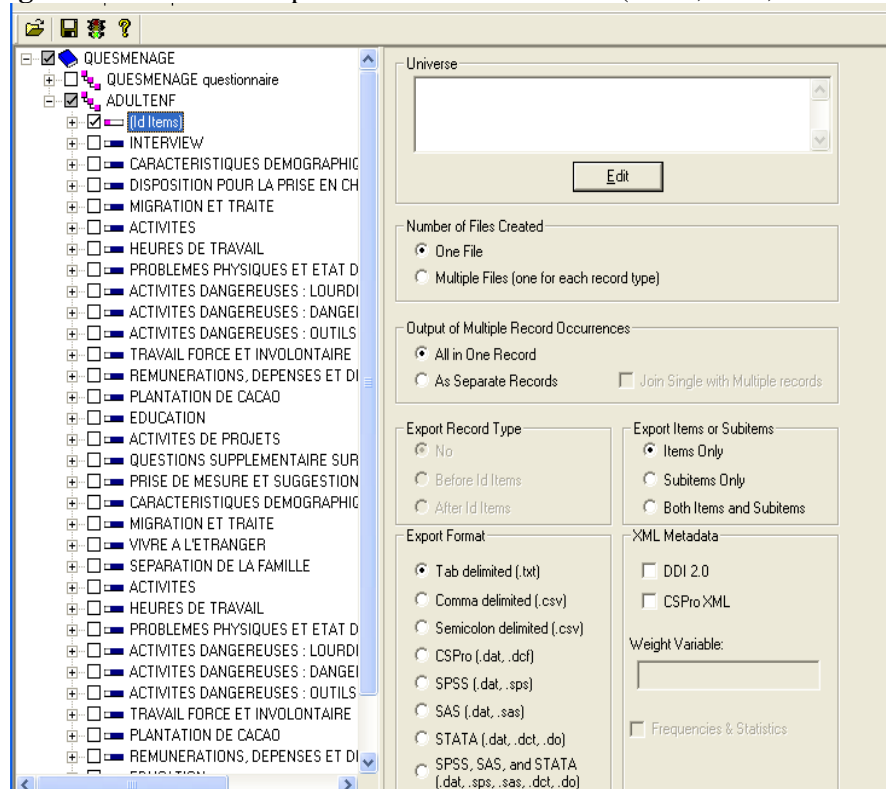
Une fois l'apurement terminé, les données peuvent être exportées vers le logiciel de votre choix (SAS, SPSS, STATA). N'oubliez pas d'exporter le *outfile* qui constitue le fichier de données apurées.


III.3. Exportation de fichier de données

Pour transférer les données, dans la barre d'outil, faites « **Tool /Export data** ». Une boîte de dialogue s'affiche, spécifiez le nom du dictionnaire et cliquez ensuite sur **Ouvrir**.

Cette dernière action donne accès à la boîte de dialogue ci-dessous. Sélectionnez toutes les variables qui feront l'objet de l'analyse en cliquant dans le carré à gauche du dictionnaire. Vous pouvez aussi choisir quelques unes d'entre elles en cliquant sur chacun des carrés qui se trouvent à gauche des items. Dans l'option «**Export Format**», choisissez le logiciel vers lequel les données seront exportées.

Figure 13 : module d'exportation des données vers (SPSS, SAS, STATA...)



Une fois les variables choisies, un clic sur  donne accès à la une boîte de dialogue qui permet de sélectionner le fichier de données apurées (dans lequel les données ont été stockées). Suivez les instructions et les données seront exportées aisément.

III.4. Les options d'exportation de fichier de données

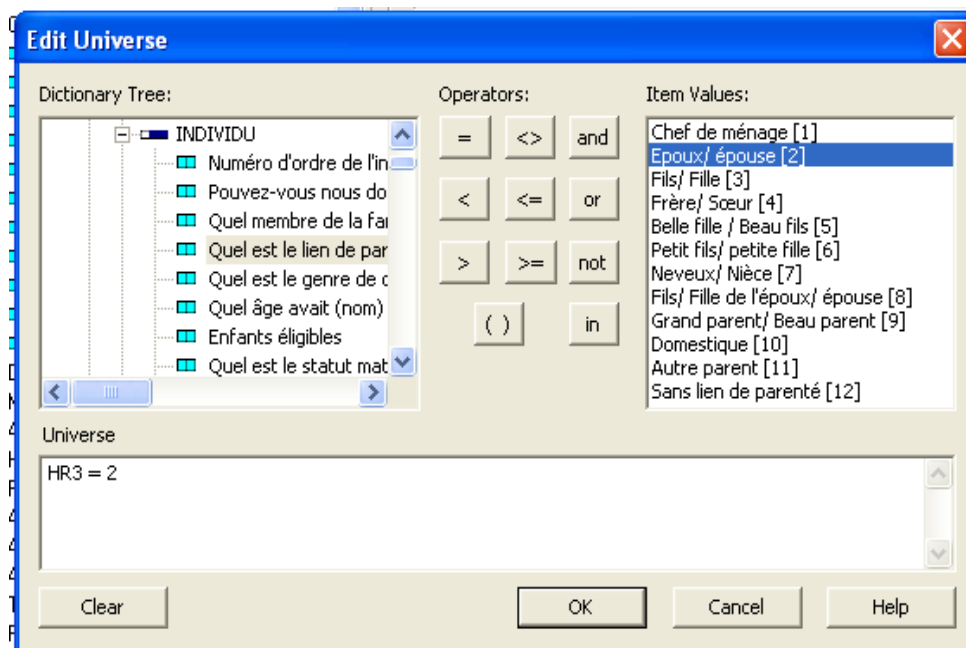
Universe

Il s'agit de préciser dans ce champ ce que l'on appelle (en mathématique) « l'univers » ou bien « population » pour les statisticiens.

Par exemple, si l'on désire avoir des données sur la fécondité, la population concernée ou l'univers est l'ensemble des femmes, résidentes, ayant un âge compris en 12 et 49 ans. Ainsi des critères de sélection seront précisés sur ces variables :

Sexe = femme et *Statut de résidence = résident* et $12 \leq \text{age} \leq 49 \text{ ans}$

Si des difficultés sont éprouvées pour définir l'univers, l'on peut faire appel à l'assistant par un clic sur le bouton « Edit » en dessous de l'option « Universe ». Cette action donne accès à la boîte de dialogue ci-dessous :



Dans la case « Dictionary Tree » sont présentées les différentes composantes du dictionnaire (records, Items). Un double clic sur l'item ou la variable correspondante permet d'afficher la variable dans le champ « Universe ». Si des « value set » sont définies pour la variable, alors celles-ci sont affichées dans le champ « Item Value ». Pour définir l'univers, il faudra cliquer sur l'opérateur logique correspondant et choisir la valeur de la modalité que l'on désire conservée ou exclure selon le besoin. Un clic sur « OK » permet de valider l'opération.

Number of files Created :

- « **One file** » permet de créer un seul fichier de données contenant tous les records du questionnaire pour un même enregistrement.
- « **Multiple file** » permet de créer autant de fichiers de données que de records déclarés dans le dictionnaire ou sélectionnés.

Output of multiple Record Occurrence

Cette option concerne les questionnaires contenant des Rosters. Les Rosters sont similaires à des tableaux de variables. Ces variables sont généralement créées dans des records dont l'option « Max » est supérieur à 1. Dans ce cas, la même information peut être saisie à plusieurs reprises. Par exemple, dans un ménage, on recueille de l'information sur chaque membre appartenant au ménage. Pour rendre cette opération de saisie répétée possible, le record relatif aux informations individuelles doit être défini comme « multi occurrence » en saisissant dans l'option « max » du record un chiffre relatif au nombre maximal d'occurrences qu'il est possible de renseigner.

- « ***All in One file*** » cette option permet d'avoir les différentes occurrences saisies sur une même ligne de données. Considérons l'exemple ci-dessus (des ménages et des individus du ménage). Si cette option est sélectionnée, dans les données exportées, on aura autant de lignes d'enregistrement que de ménages. Ainsi les informations individuelles se trouveront sur la même ligne que celles relatives au ménage correspondant.
- « ***As Separate Records*** » cette option est opposée à la précédente. En effet, dans ce cas, les données exportées auront autant de lignes qu'il y a au total d'individus : le nombre total de ligne de données sera égal au nombre total d'individus saisis.
- « ***Join Single with Multiple record*** » Cette option est activée lorsque l'option « As Separate Records » est sélectionnée. Dans ce cas, les informations communes aux individus sont répétées pour chacun. Par exemple, les informations relatives à l'habitat ou aux caractéristiques du ménage sont communes aux individus. Lorsque cette option est choisie et que les records relatifs aux caractéristiques communes sont sélectionnés, pour tous les individus, ces caractéristiques communes sont renseignées dans le fichier exporté. Au cas contraire, ces caractéristiques ne seront renseignées que pour les premiers membres de chaque ménage.

Export Items or Subitems

- « ***Items Only*** » cette option permet d'exporter uniquement les variables déclarées comme items ;
- « ***Subitems Only*** » celle-là permet d'exporter uniquement les variables déclarées comme subitems ;

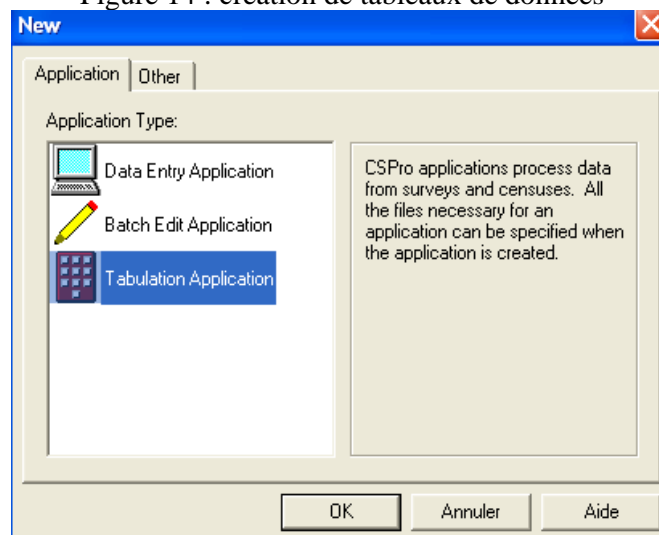
- « Both Items and Subitems » cette dernière option permet d'exporter autant les items que les subitems. Cependant, pour les items qui ont des Subitems, seuls les subitems seront exportés.

IV. Création de tableaux sous CSPro

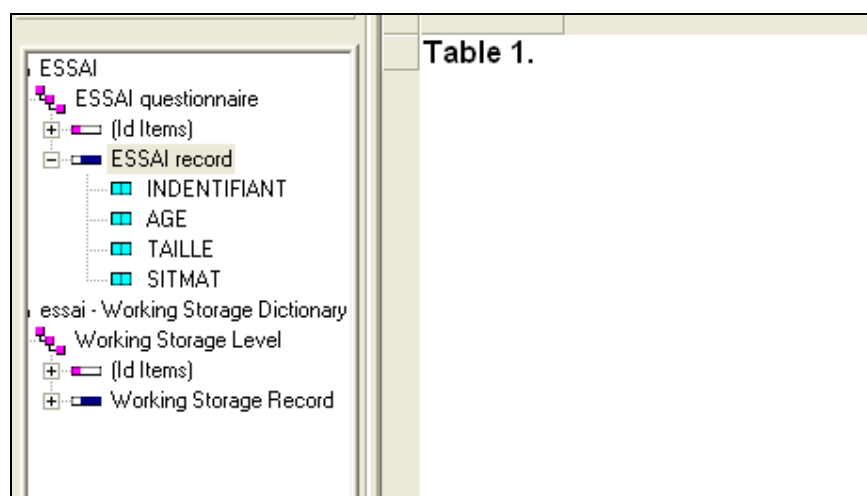
Une analyse descriptive peut être faite sous CSPro (tableaux, tableaux croisés dynamiques, génération de statistique : moyenne, variance...)

La production de tableau sous CSPro se fait dans un environnement de travail conçu à cet effet. Par conséquent il convient de créer une nouvelle application qui donne accès à cet environnement. Dans la barre d'outils, cliquez sur **File/New** et choisir comme type d'objet **Tabulate Application**.

Figure 14 : création de tableaux de données

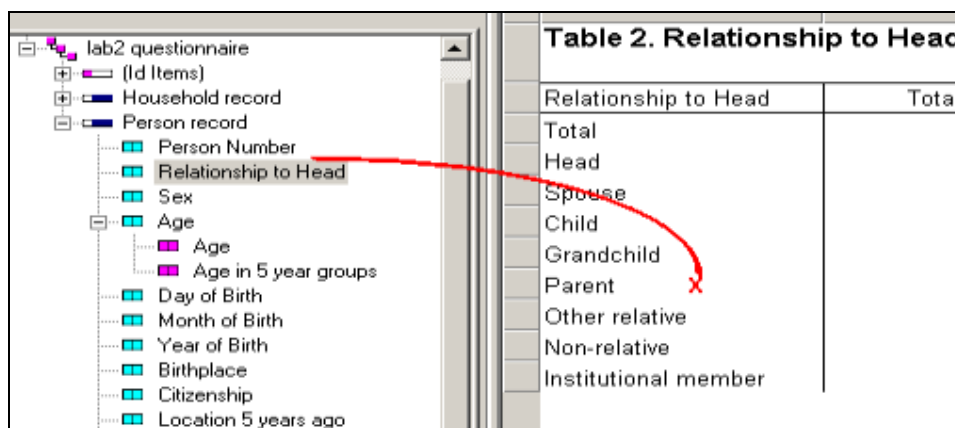


Après avoir cliqué sur ok, spécifiez le nom de la table et cliquez ensuite sur «**Create/créer**». Cette dernière action affiche une boîte de dialogue qui vous permet de choisir le dictionnaire de variable correspondant aux données dont on désire établir des statistiques. Cliquez enfin sur **Ok** et l'application sera créée (*l'environnement de travail ci-dessous s'affiche*).



Création de tableaux

Cette partie sera brièvement présentée. Pour avoir un tableau de répartition selon un variable donnée, il suffit de cliquer sur la variable concernée et de faire un *glisser/déposer* comme décrit ci-dessous.



Pour supprimer une variable du tableau, il suffit de procéder par de la même manière dans le sens inverse.

Prenons l'exemple de la variable situation matrimoniale. Le glisser déposer sur cette variable donne le tableau ci-dessous

Table 1. SITMAT	
SITMAT	Total
Total	
Célibataire	
Marié(e)	
NC	

Le tableau peut être disposé en colonne ou en ligne. La répartition des répondants (par effectif) selon la variable est obtenue par un clic sur RUN dans la barre d'outil (une boîte de dialogue vous permet de choisir le fichier de données correspondant) et vous obtenez votre tableau de répartition

Table 1. SITMAT	
SITMAT	Total
Total	17
Célibataire	5
Marié(e)	2
NC	10

Selon vos besoins, vous pouvez créer des tableaux de répartition avec plus d'une variable (comme le montre le tableau ci-dessous).

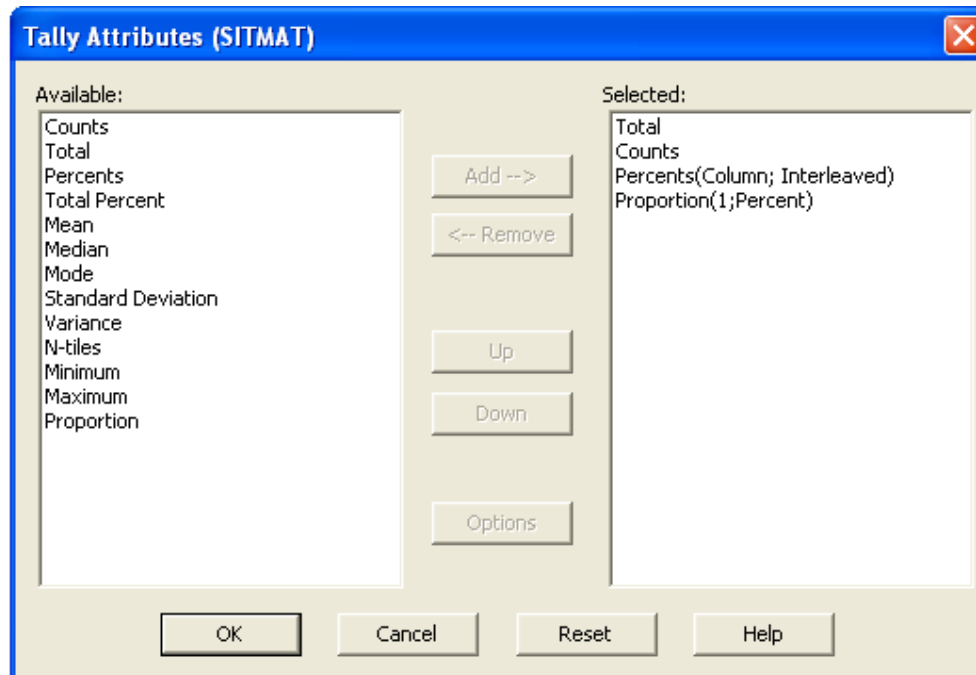
Table 1. Marital status by Sex and Literacy									
Marital status	Sex								
	Total			Male			Female		
	Literacy			Literacy			Literacy		
	Total	Literate	Illiterate	Total	Literate	Illiterate	Total	Literate	Illiterate
Total									
Married									
Divorced									
Separated									
Widowed									
Never Married									

Pour le tableau ci-dessus, il s'agit de la répartition des répondants par *situation matrimoniale* selon le *sexe* et la *situation d'alphabétisation*. Pour avoir un tel tableau il suffit de cliquer, glisser et déposer les variables concernées selon la disposition recherchée.

On peut aussi avoir des tableaux du genre

Table 1. Marital status by Sex, Literacy									
Marital status	Sex			Literacy					
	Total	Male	Female	Total	Literate	Illiterate	Not Reported	Not Applicable	
Total									
Married									
Divorced									
Separated									
Widowed									
Never Married									

La répartition en pourcentage des répondants peut être faite. Les statistiques telles que la variance, la moyenne... peuvent être obtenues. Pour introduire des répartitions en pour cent, faites un clic droit sur le libellé de la variable et dans le menu contextuel choisir «Tally Attributes». Cette option donne accès à une boîte de dialogue ci-dessous. Selon le type de la variable, choisir les options désirées.



Les options enfin choisies, un clic sur «RUN » dans la barre d’outils vous donne accès à une boîte de dialogue qui vous permettra de choisir le fichier de données et votre tableau est prêt.

Juste une partie insignifiant des capacités de CSPro en général et en analyse en particulier a été présentée dans ce document. Vos efforts de recherche vous permettront d’approfondir vos connaissances en la matière.

ANNEXES

- Arithmetic Operators

Operation	Symbol
Addition	+
Soustraction	-
Multiplication	*
Division	/
Modulo	%
Exponentiel	^

- Relational Operators

Operation	Symbol
Egal à	=
Différent de	<>
Inférieur à	<
Inférieur ou égal à	<=
Supérieur à	>
Supérieur ou égal à	>=
In range	in

- Logical Operators

Operation	Symbol	Keyword
Negation	!	not
Conjunction	&	and
Disjunction		or
Si et seulement si	<=>	

Lorsque plus d'un opérateur existe dans une expression, l'ordre dans lequel les opérateurs sont évalués est déterminé par leur priorité. Les opérateurs ci-dessus sont utilisés dans des expressions logiques pour examiner si un item (ou une variable) est dans un ensemble de valeurs ou de gammes. L'item ou la variable peut être numérique ou alphanumérique. Une plage de valeurs est séparée par des deux points, par exemple 1:5. Des éléments d'une liste de valeurs ou de gammes sont séparés par des virgules. Par exemple 1, 3:5 et 7.

Example 1:

if RELATIONSHIP in 1:5 then

is the same as

if RELATIONSHIP >= 1 and RELATIONSHIP <= 5 then

Example 2:

If WORK in 1, 3, 5 then

Is the same as

If WORK = 1 or WORK = 3 or WORK = 5 then

Example 3:

If X in 1:4, missing, notappl then

is the same as

If (X >= 1 and X <= 4) or X = missing or X = notappl then

Example 4:

If NAME in "A":"MZZ" then

Is the same as

If NAME >= "A" and NAME <= "MZZ" then