



*BD/MSPS* 

# Guide d'utilisation des tableaux croisés

Le présent guide contient une description complète de la Fonction Tableaux croisés précisée par l'utilisateur du MSPS ainsi qu'une série d'exemples. Il constitue ainsi tant un guide complet qu'un guide de référence final. On trouvera dans le document BD/MSPS - Guide d'utilisation une description sommaire de la fonction Tableaux croisés.



Statistics  
Canada

Statistique  
Canada

Canada



## Table des matières

Introduction.....	2
Terminologie.....	2
Variables totalisées .....	3
Unité d'analyse .....	4
Variables de classe.....	4
Étiquetage et format.....	5
La spécification du jeu de tables.....	5
Sélection et pondération des enregistrements .....	6
Sélection des enregistrements .....	6
Pondération des enregistrements.....	7
La spécification du jeu de tables.....	8
Paramètre du fichier de commande.....	8
Le contenant des demandes de tables .....	8
Syntaxe -- Contenu .....	9
Syntaxe -- Apparence.....	9
Chaînes non valides .....	9
Paramètres connexes.....	10
La demande de table .....	12
Structure générale .....	12
Les niveaux de totalisation et leur ordre .....	13
Unités d'analyse.....	13
HH (Ménage).....	13
EF (Famille économique).....	13
CF (Famille de recensement).....	13
NF (Famille nucléaire) .....	14
IN (Individu ou personne) .....	14
Variables de totalisation.....	14
Le niveau de la variable de totalisation d'une demande de table .....	14
Variables de totalisation multiples dans une demande .....	15
Sources des variables de totalisation.....	15
Variables de totalisation spéciales .....	17
Qualificatifs des variables de totalisation .....	18
Contenu de cellules et étiquetage pour les expressions et variables de totalisation.....	22
Variables de classe.....	24
Nature des variables de classe.....	24
Contribution à la dimensionnalité des tables .....	24
Nature des niveaux de classe .....	24
Sources des variables de classe.....	25
Création d'une catégorie sommaire "Tous".....	26
Étiquetage des variables de classe .....	26
Niveau d'analyse.....	26
Titres des tables.....	28
Exceptions et divers .....	29
Exemples pratiques .....	30

Agrégat d'une seule variable .....	30
Une demande de tables multiples avec une variable de classe .....	30
Demande de variables de totalisation multiples dans une seule table .....	31
Présentation des qualificatifs et des expressions totalisés .....	32
Deux niveaux de classe .....	33
Une variable de classe définie par l'utilisateur et une demande sur lignes multiples.....	34
Présentation de la normalisation et d'une troisième dimension.....	35
Trucs pour les tableaux croisés .....	39
Limites imposées de l'extérieur sur la fonction Tableaux croisés.....	39
Le choix d'une exécution du MSPS de la taille appropriée.....	40
Exportation de données vers d'autres progiciels .....	40
Utilisation des fichiers d'inclusion de paramètres .....	41
Production de demandes de totalisation efficaces et élégantes.....	41
Résumé.....	42
Annexe A Principaux messages d'erreur de la fonction Tableaux croisés.....	43
Annexe B Détails de la syntaxe .....	45

## Introduction

Le présent guide décrit la fonction Tableaux croisés du MSPS, en indiquant ce qu'elle est, ce qu'elle peut faire pour l'analyste, la façon dont elle a des interactions avec d'autres volets du MSPS et la façon dont on l'utilise pour produire les tables voulues.

La fonction Tableaux croisés est l'une des trois grandes voies que l'analyste peut emprunter pour obtenir des résultats d'analyse d'une exécution du MSPS. Les deux autres sont :

- l'ensemble des tables normalisées qui font partie intégrante du MSPS, ensemble qui est décrit dans le document *Guide d'utilisation*; et
- la capacité du MSPS d'exporter des données sélectionnées vers un fichier pour traitement subséquent sur PC-SAS, capacité qui est décrite ailleurs dans le même guide.

Il y a une autre voie, l'exportation des données vers un fichier ASCII, mais elle est utilisée principalement pour l'examen des données sur des individus ou des ménages en particulier; cette fonction de sortie texte est aussi décrite dans le document *Guide d'utilisation*. L'analyste utilisera la fonction Tableaux croisés lorsque :

- l'information désirée ne se trouve pas dans les tables normalisées; et
- la fonction Tableaux croisés fournit un moyen plus efficace que l'exportation des données pour analyse subséquente sur PC-SAS ou que la fonction de sortie texte.

Suffisamment petite et rapide pour être intégrée directement dans le MSPS, la fonction Tableaux croisés offre une puissance considérable sous une forme facile à utiliser. Par exemple, la fonction Tableaux croisés reconnaît l'importance de résultats bien étiquetés, en offrant un étiquetage automatique substantiel qui permet cependant à l'analyste de fournir ses propres étiquettes. Elle peut faire la totalisation d'une grande variété de variables directement de la base de données ou du modèle MSPS, ainsi que d'expressions utilisant ces variables, et de dresser le sommaire des résultats jusqu'à l'un ou l'autre des cinq niveaux d'unité d'analyse. La fonction a d'autres caractéristiques dont la capacité de générer plusieurs tables à partir d'une seule spécification de tables, de générer des tables multidimensionnelles et d'inclure plusieurs variables totalisées dans une table donnée.

Le présent guide a été rédigé d'abord comme manuel de référence, puis comme guide de formation. L'utilisateur expérimenté y trouvera des choses rapidement en utilisant la table des matières, tandis que l'utilisateur néophyte peut lire les chapitres dans l'ordre pour obtenir de l'information de base sur tous les aspects de la fonction Tableaux croisés. Les nombreux exemples peuvent servir ces deux clientèles. L'utilisateur qui connaît le principe des tableaux croisés et qui préfère apprendre par l'exemple peut aller directement à la section *La spécification du jeu de tables*, puis apprendre ce qu'il lui manque plus tard en consultant les sections pertinentes du Guide.

## Terminologie

La présente section donne la terminologie utilisée dans le Guide. Pour ce faire, elle utilise une table "typique" qui illustre plusieurs des plus importants éléments automatiques et déterminés par l'utilisateur. Les sections suivantes du chapitre définissent ensuite plus en détail chacun des éléments. La section *Terminologie* vise principalement à fournir des notions de base avant que

des définitions détaillées ne soient données. Essentiellement, elle donne un aperçu du fonctionnement des tableaux croisés dans le MSPS.

L'explication s'articule autour d'une illustration annotée d'un tableau croisé typique du MSPS, la figure 1, et de la spécification de table qui a servi pour la produire. Comme cette présentation vise principalement la terminologie, l'illustration ne tente pas de démontrer toutes les capacités de la fonction Tableaux croisés. En outre, le lecteur devrait savoir que, tant ici qu'ailleurs dans la documentation, les tables utilisées dans les illustrations sont basées sur un sous-échantillon particulier de la BDSPS. Toute tentative de reprise de l'exemple avec le MSPS peut produire des tables qui ne correspondent pas exactement à celles qui sont illustrées.

Figure 1 :

Table 1U: Unit Count (000) for Census Families by Province and Census family type

Province	Census family type						All
	With	With	With	With	Other, 1	Other, 1	
	Kids, 1	Kids, 2+	Elderly, 1	Elderly, 2+	Adult	2+ Adult	
	Adult	Adult	Adult	Adult			
Newfoundland	2.4	80.3	18.1	24.3	33.2	17.8	176.2
P.E.I.	2.9	12.5	8.2	4.1	4.8	6.9	39.4
Nova Scotia	7.2	119.9	36.7	44.2	46.5	63.6	318.2
New Brunswick	11.8	87.6	30.3	36.1	37.7	53.3	256.9
Quebec	138.1	854.0	206.5	253.7	566.8	580.6	2599.6
Ontario	262.2	941.8	463.4	448.2	613.0	883.8	3612.5
Manitoba	7.9	130.6	48.2	36.9	82.1	93.8	399.5
Saskatchewan	40.7	135.8	20.5	46.0	75.1	94.5	412.6
Alberta	59.8	311.8	86.5	60.0	306.8	192.0	1016.9
B.C.	59.1	385.1	129.3	94.9	318.3	234.9	1221.6
All	592.3	3059.5	1047.8	1048.3	2084.3	2221.2	10053.4

Cette table a été produite par la spécification suivante :

```
XTSPEC
CF: {units}
    * hdpov+
    * cftype+;
```

ou, sur une seule ligne :

```
XTSPEC
CF: {units} * hdpov+ * cftype+;
```

## VARIABLES TOTALISÉES

Essentiellement un tableau croisé devrait faire une totalisation, c'est-à-dire compter ou additionner quelque chose et présenter les résultats dans une table. Ce "quelque chose" est désigné "variable totalisée". En termes généraux, une variable totalisée peut être toute variable d'analyse disponible dans le BD/MSPS. (Les variables d'analyse contiennent des données qui

peuvent être “additionnées”).

Les variables totalisées peuvent provenir :

- de variables (de base de données) de BDSPPS,
- de variables (de modèle) du MSPS,
- de variables ou d’expressions définies par l’utilisateur, ou
- d’expressions définies “à la volée” dans la spécification de table.

La fonction Tableaux croisés permet qu’il y ait plusieurs variables totalisées dans une même table. Elle permet la totalisation de variables de type “unité” (p. ex., famille de recensement), de variables en dollars (p. ex. les allocations familiales), ainsi que les variables définies par l’analyste.

Dans la figure 1, la variable totalisée est le nombre d’“unités”. Nous le savons à la lecture du titre que la fonction Tableaux croisés a automatiquement donné à la table le titre “Table 1U: Unit Count (000) for Census Families by Province and Census family type” (Table 1U : Compte d’unité (000) pour les familles de recensement, par province et type de famille de recensement). Quand elle a composé automatiquement ce titre, la fonction Tableaux croisés a incorporé ce qu’elle savait de l’“unité d’analyse” décrite ci-dessous.

## **UNITÉ D’ANALYSE**

L’analyste sait que, pour certains genres de tables, l’unité (niveau) d’analyse inhérente est critique. Si nous désirons compter le nombre de familles au Canada, le résultat différera selon que l’on parlera de familles économiques, de familles de recensement ou de familles nucléaires. De même, il est important que la définition de l’unité d’analyse soit précise pour la totalisation du revenu moyen par unité. La fonction Tableaux croisés permet à l’analyste de présenter des tables à cinq niveaux d’analyse distincts. Ménages, familles économiques, familles de recensement, familles nucléaires et individus.

À la figure 1, l’analyste a choisi la famille de recensement comme unité d’analyse. Comme on l’a décrit ci-dessus, la fonction Tableaux croisés indique ce choix dans le titre qu’il compose pour la table; c’est-à-dire que le titre indique explicitement que la table fournit un “compte d’unité (000) pour les familles de recensement”.

## **VARIABLES DE CLASSE**

La fonction Tableaux croisés permet à l’analyste d’afficher certaines variables totalisées par classe ou catégorie définie à partir d’autres variables, et par combinaison de ces catégories. Du point de vue tant de la mécanique qui produit une telle table, que de la capacité de l’analyste de la comprendre, le nombre de ces catégories devrait en général être petit. Dans le MSPS, ces variables sont appelées “de classe” puisqu’elles classent tout enregistrement dans exactement une catégorie ou classe. Le MSPS veille à l’obtention de ce caractère distinct en exigeant que toutes les variables de classe soient des nombres entiers. Les variables seront soit définies comme variables entières dans la base de données, soit créées comme variables entières dans leur définition par l’utilisation de la fonction Fraction dans la fonction variables utilisateurs UVAR ou variable définie par l’utilisateur en mode “boîte de verre”.

Dans la figure 1, l’analyste a totalisé les nombres des familles de recensement dans deux

variables de classe. Les catégories de colonnes correspondent aux divers types de familles (p. ex., unités de personnes âgées, familles de deux adultes avec des enfants, etc.), avec les définitions de catégories appliquées de façon que chaque famille de recensement tombe dans exactement une catégorie de l'ensemble. Les catégories de rang correspondent à la province de résidence, avec toute famille de recensement donnée résidant dans exactement une province au moment même où l'Enquête sur la dynamique du travail et du revenu recueillait l'information. Ces variables de classe sont toutes deux disponibles de la BDSPS.

La plupart des variables de classe sont des variables comprises à la section *Caractéristiques des individus et des familles*, et des variables drapeau d'autres sections du document *Guide des variables*.

## ÉTIQUETAGE ET FORMAT

Les tables ne sont pas d'une grande utilité si le lecteur ne peut pas facilement les comprendre. Par conséquent, la fonction Tableaux croisés fait en sorte que le contenu soit clair et qu'elle s'affiche dans le format approprié. Cette documentation de la fonction Tableaux croisés va au-delà de l'étiquetage détaillé des pages que le MSPS fournit pour tous ses produits de sortie.

- L'en-tête d'une table indique ce qui est totalisé ainsi que les classes qui sont utilisées.
- Des étiquettes de rangs et de colonnes précisent la combinaison pertinente des variables de classe de la table.
- Pour les tables à trois dimensions et plus, des renseignements "de segmentation" supplémentaires sont aussi fournis dans l'information d'en-tête de chaque segment de la table.

À la figure 1, la catégorie de rang de la totalisation, la province de résidence, se trouve dans la colonne de gauche, clairement étiquetée par des chaînes de texte identifiant la province de résidence. La fonction Tableaux croisés a récupéré ces chaînes de la partie "Description de la variable" de la base de données et les a appliquées automatiquement. De même, la fonction Tableaux croisés a récupéré et appliqué les en-têtes de colonnes qui indiquent les types de familles.

Chose moins évidente, la fonction Tableaux croisés a choisi des conventions de mise en forme appropriées. Par exemple, elle a choisi comme échelle "milliers de familles" pour les entrées de cellule, en indiquant ce choix par l'élément "(000)" du titre de la table. De même, elle a choisi d'utiliser un seul chiffre après la virgule. L'analyste aurait pu modifier l'un ou l'autre de ces choix si d'autres formes lui avaient mieux convenu.

## LA SPÉCIFICATION DU JEU DE TABLES

Le moyen utilisé pour spécifier les tables qui sont désirées est la "spécification du jeu de tables" ou la spécification de tableaux croisés, et cela se fait par le paramètre XTSPEC dans le fichier de paramètres de commande. Le paramètre permet à l'analyste de spécifier le jeu de tableaux croisés désiré.

À la suite de la figure 1, on trouve la spécification du jeu de tables qui a produit la table illustrée dans la figure. Le MSPS n'inclut pas lui-même la spécification dans les tables qu'il produit. Cette spécification particulière demande seulement la table qui est illustrée; en général, cependant, un seul paramètre XTSPEC peut lancer la production d'une grande variété de tables



définies par l'utilisateur.

Le paramètre XTSPEC utilisé ici comprend la chaîne de texte :

```
XTSPEC
CF: {units}
    * hdpov+
    * cftype+;
```

- La demande peut être formulée sur en deux lignes, mais il est suggéré d'utiliser une demande qui se poursuit sur plusieurs lignes pour faciliter la lecture et la modification de la demande.
- Le symbole CF: utilisé dans la demande indique que la table utilise la famille de recensement comme unité d'analyse sous-jacente.
- L'identité de la variable totalisée est transmise par l'identificateur "units" qui est donné entre accolades; c'est-à-dire que les unités de famille de recensement sont totalisées.
- Les variables désirées pour les rangs et les colonnes sont définies par les identificateurs hdpov et cftype que le lecteur connaît car ils se trouvent dans la liste des variables de la BDSPS (voir le document *Guide des variables*.).
- Le caractère + qui suit les noms de ces variables de classe indique qu'une catégorie "ALL" (toutes) doit être créée pour chacune d'elles.
- Un point-virgule est nécessaire à la fin de la demande de table.

Ce bref exemple donne la présentation la plus générale possible des capacités de la fonction Tableaux croisés. Les sections suivantes reprennent les détails nécessaires, indiquant les données qui doivent être comprise dans les tables, la nature de la spécification du jeu de tables et les demandes de tables individuelles qu'elle contient.

## Sélection et pondération des enregistrements

### SÉLECTION DES ENREGISTREMENTS

Toute étude significative du processus de production de tables doit commencer par une spécification précise des données qui doivent être traitées dans la production des tables. Pour commencer, l'utilisateur spécifie, dans le fichier de paramètres de commande, l'identité de la base de données à utiliser pour l'exécution et, par conséquent, pour la production des tables pendant l'exécution. Par exemple, l'analyste pourrait spécifier la BDSPS au complet ou un sous-échantillon de 5 % de toute la base de données.

En outre, toujours par le fichier de paramètres de commande, l'analyste peut spécifier que seulement certains enregistrements doivent être traités. Plus précisément, l'analyste peut utiliser les paramètres SELFLAG, SELUNIT et SELSPEC pour restreindre les données traitées par le MSPS (voir le document *Guide d'utilisation* pour obtenir une description complète de ces paramètres).

Cette capacité de sélection améliore fortement l'efficacité du calcul (et, par conséquent, accélère l'exécution) et produit des tables plus concentrées quand seulement certains enregistrements sont pertinents à une analyse (lorsque, p. ex., seulement les familles nucléaires comptant des enfants présentent de l'intérêt). Lorsque l'analyste utilise les paramètres de sélection pour exclure certains enregistrements, il doit accepter la responsabilité de l'interprétation des tables

ainsi produites seulement dans le contexte des données sélectionnées pour l'exécution.

L'analyste devrait aussi noter que les critères de sélection d'enregistrements sont liés au choix de l'unité d'analyse. L'analyste détermine si la spécification de sélection est interprétée comme comptant tous les individus d'une unité qui respectent le critère de sélection, ou seulement les individus d'une unité plus vaste qui respectent le critère de sélection. Pour une étude plus détaillée, l'analyste devrait consulter les sections du document *Guide d'utilisation* qui portent sur les paramètres de sélection et la notion de cumul entre les unités d'analyse.

## **PONDÉRATION DES ENREGISTREMENTS**

De même, parce que la BDSPPS est un ensemble de données pondérées, la production de tables doit tenir compte de la façon dont le poids des enregistrements individuels influe sur les résultats. L'Enquête sur la dynamique du travail et du revenu, dont le cadre d'échantillonnage est à la base du BD/MSPS, est un échantillon aléatoire stratifié. Cela signifie que ce ne sont pas tous les ménages du Canada qui ont une possibilité égale de participer à l'enquête. Par conséquent, lié à chaque enregistrement de la BDSPPS, il y a un "poids d'enregistrement", essentiellement l'inverse de la probabilité que le ménage soit choisi pour l'enquête, ou le nombre de ménages canadiens que l'enregistrement représente.

Dans la pratique, les poids des enregistrements de la BD/MSPS sont stockés dans un fichier distinct de celui de l'information sur les caractéristiques des ménages et des individus. Par conséquent, le même fichier de caractéristiques d'unité peut être utilisé de concert avec plusieurs fichiers de poids (p. ex., les fichiers de poids correspondant à diverses hypothèses de croissance démographique). Le paramètre de commande INPWGT contient le nom du fichier de poids particulier qui doit être utilisé pour une exécution. Il est possible d'en apprendre plus sur ces poids d'enregistrement, leur interprétation et leur calcul en consultant les documents *Guide d'utilisation* et *Guide de création de bases de données*.

Dans la plupart des totalisations de la fonction Tableaux croisés, ainsi que dans le menu du MSPS des tables encodées normalisées, le processus de production de tables fait la pondération des variables totalisées avec le poids d'enregistrement approprié. Cette opération vise à ce que la table soit ainsi "portée aux dimensions" des totaux canadiens globaux (pour les genres d'enregistrements respectant les critères de sélection) qui sont du plus grand intérêt pour l'analyste. Ce ne sont que ces totaux pondérés qui sont comparables à des sources de données exogènes, avec, par exemple, une estimation du total des revenus d'emploi payés au Canada.

Il y a cependant deux exceptions à cette règle générale de production de tables dont les entrées sont pondérées avec les poids d'enregistrement appropriés.

- Une exécution particulière du MSPS pourrait ne pas traiter tous les enregistrements de la BDSPPS respectant les critères de sélection. Ceci pourrait se produire parce que a) l'analyste utilisait intentionnellement un sous-échantillon de la BDSPPS, par exemple l'échantillon de 5 %, et qu'il désirait seulement obtenir des résultats approximatifs, ou b) l'analyste a interrompu l'exécution avant que le traitement de l'échantillon ne soit terminé.

Dans ces deux cas, le MSPS produit une série de "normalisations" appropriées sur les enregistrements qu'il a traités afin de les "porter aux dimensions" d'un total canadien (soumis

aux critères de sélection d'enregistrements). Par conséquent, si l'analyste a, par exemple, interrompu une exécution au tiers du traitement de l'échantillon de 5 % de la BDSPS, les tables devraient alors être pondérées avec des poids équivalant à 60 fois ceux qui sont réellement liés aux enregistrements. Ici, le facteur de 60 est calculé comme l'inverse de  $(0,05) \cdot (1/3)$ .

- La seconde exception, décrite plus en détail ci-dessous après l'étude du concept des variables de totalisation, survient pour certaines variables qui ont été spécifiquement prévues de façon à ne pas être pondérées. Ces variables de totalisation "spéciales" sont utilisées lorsque l'analyste désire compter les enregistrements eux-mêmes plutôt qu'une variable de totalisation plus typique comme les familles, les gains ou les taxes et impôts.

Les variables spéciales (spsdrecs, scfrecs and fxvrecs) sont utiles pour des applications comme l'évaluation du nombre d'enregistrements de la BDSPS sur lesquels diverses moyennes reposent, par exemple comme moyen d'évaluation de la fiabilité de la mesure. Par conséquent, lorsque la fonction Tableaux croisés crée des totalisations de ces variables spéciales, elle n'applique pas les poids d'enregistrement. Cependant, l'interprétation des totalisations touchant des variables spéciales dépend quand même du choix que fait l'analyste quand au niveau de cumul pour les tables.

## La spécification du jeu de tables

La présente section explique l'utilisation des spécifications de jeu de tables dans le MSPS. On y étudie la condition de la spécification comme paramètre de fichier de commande (XTSPEC), la mesure dans laquelle la spécification contient une ou plusieurs demandes de tables individuelles, la syntaxe de la spécification du jeu de tables et d'autres paramètres de commande qui influent sur le fonctionnement de la spécification du jeu de tables.

### PARAMÈTRE DU FICHIER DE COMMANDE

Le paramètre de spécification du jeu de tables est l'un des nombreux paramètres d'un fichier de commande du MSPS. (Pour obtenir une description complète des fichiers de commande du MSPS, qui ont pour caractéristique l'extension de nom de fichier .cpr, consulter le document *Guide d'utilisation*.) La spécification du jeu de tables est donnée au MSPS par le paramètre XTSPEC. Ce paramètre XTSPEC prend la forme d'une chaîne de texte, souvent une chaîne plutôt longue. Parce que la chaîne peut être longue, le MSPS permet qu'elle contienne des retours de chariot. Par conséquent, le paramètre peut occuper plusieurs lignes du fichier de paramètres de commande.

Habituellement, l'utilisateur crée ou modifie le paramètre XTSPEC avec un éditeur de texte (comme BRIEF) lorsqu'il crée ou modifie le fichier de paramètres de commande approprié. Bien que le paramètre XTSPEC puisse être entré "à la volée" au début de l'exécution du MSPS, ceci ne se ferait habituellement que par des analystes expérimentés et pour l'entrée de spécifications qui ne sont pas particulièrement longues ou compliquées. L'analyste peut aussi utiliser la fonction "à la volée" du MSPS pour corriger les erreurs de syntaxe dans un paramètre XTSPEC existant.

### LE CONTENANT DES DEMANDES DE TABLES

Deux points nous intéressent ici. En premier lieu, XTSPEC est un moyen unique que l'analyste peut utiliser pour demander que le MSPS produise une totalisation personnalisée pendant une exécution. Le MSPS offre une variété de tables normalisées et des fonctions d'exportation de données, mais seul le paramètre XTSPEC permet à l'analyste d'exécuter des tableaux croisés personnalisés à l'intérieur du MSPS.

En second lieu, la spécification du jeu de tables est principalement un "contenant" pour les demandes de tables individuelles de l'analyste. Ces demandes de tables, le sujet de la section suivante, sont des "ensembles d'information" qui demandent au MSPS de produire les tables individuelles désirées. En effet, la seule information contenue dans le paramètre XTSPEC est une série de demandes visant ces tables, c'est-à-dire que XTSPEC ne contient rien sauf ces demandes de tables. Le nombre de demandes pourrait être de zéro, c'est-à-dire que XTSPEC lui-même peut être vide si l'analyste ne désire pas utiliser la fonction Tableaux croisés; cependant, notre étude se poursuivra comme si XTSPEC contenait au moins une demande de table.

## **SYNTAXE -- CONTENU**

Dans un fichier de paramètres de commande du MSPS, une spécification de jeu de tables vide comprend le mot clé XTSPEC suivi d'un retour de chariot (après lequel le paramètre suivant du fichier de commande commence). Une spécification de jeu de table qui n'est pas vide contient le mot-clé XTSPEC suivi d'au moins un espace ou un caractère de tabulation suivi d'un point virgule. Le caractère point-virgule (";") sert à séparer les demandes de tables individuelles. Le retour de chariot final termine la définition du paramètre XTSPEC.

La définition et la structure interne des éléments demande de table individuelle est clairement importante et sera expliquée plus en détail à la section suivante. L'annexe B du présent guide donne une définition structurée. Au niveau de description plus général où nous nous trouvons, qui traite de la structure générale de XTSPEC, il suffit de considérer les demandes de table simplement comme des "paquets", dont chacun fournit toute l'information nécessaire au MSPS pour produire "une" table.

## **SYNTAXE -- APPARENCE**

Un des avantages de la possibilité de répartir XTSPEC sur plusieurs lignes réside dans la capacité d'entrer des spécifications de totalisation plus longues. Un autre avantage est la possibilité d'améliorer la lisibilité des demandes de table individuelles en plaçant leurs éléments sur des lignes distinctes en les entourant de "blancs" appropriés. Cette capacité est expliquée plus en détail dans l'étude sur les demandes de table.

## **CHAÎNES NON VALIDES**

Parce que la spécification du jeu de tables, alliée aux demandes de table individuelles qui en font partie, peut être passablement complexe, même un utilisateur expérimenté peut faire une erreur à sa formulation. S'il s'agit d'une erreur de contenu (c'est-à-dire que l'analyste a entré une spécification légitime, même si elle n'a pas trop de signification ou n'est pas ce qui était prévu), la fonction Tableaux croisés, bien sûr, produira simplement les tables demandées. L'analyste devra alors étudier les tables obtenues afin de déterminer s'il s'agissait d'une erreur de spécification.

Cependant, si l'analyste avait fait une erreur de syntaxe, de façon que la fonction Tableaux croisés ne pouvait pas "comprendre" la spécification, les choses seraient différentes. Le MSPS découvre ce type d'erreur de syntaxe en lisant le fichier de paramètres de commande lorsqu'il tente d'interpréter la signification de la chaîne XTSPEC. La présence de ce genre d'erreur est immédiatement évidente à l'analyste parce que le MSPS produit un message d'erreur approprié, refusant de poursuivre l'exécution jusqu'à ce qu'un paramètre acceptable ait été entré.

En général, les messages d'erreur ont été conçus de façon à donner beaucoup d'information sur la nature du problème détecté. Beaucoup d'entre eux ont été rédigés de façon que, lorsqu'ils sont invoqués, ils fassent une référence précise au contenu du paramètre spécifique de l'analyste; de cette façon, l'analyste sait immédiatement la partie du paramètre qui cause des problèmes. L'annexe A du présent guide donne une liste des principaux messages d'erreur qui peuvent s'afficher. Les entrées individuelles indiquent aussi ce que l'analyste doit faire pour corriger la situation. Le message qui s'affiche entre guillemets est aussi utile à l'interprétation et à la correction des erreurs de XTSPEC; il décrit la syntaxe correcte à laquelle le MSPS s'attend.

S'il s'agit d'une erreur simple et que la nature de la correction est évidente, l'analyste peut corriger le paramètre XTSPEC "à la volée" en utilisant les fonctions d'édition de paramètre intégrée dans le MSPS. Dans une demande faite sur plusieurs lignes, l'utilisation de CTRL-X permet de passer à la ligne suivante. On trouvera de plus amples renseignements sur l'éditeur de paramètre dans le document *Guide d'utilisation*. Pour les erreurs plus compliquées, l'analyste peut choisir d'abandonner l'exécution et d'employer un éditeur de texte plus puissant pour modifier la spécification du paramètre XTSPEC.

## **PARAMÈTRES CONNEXES**

Bien que le paramètre XTSPEC soit le plus important, et le plus complexe, des paramètres du fichier de commande qui agissent sur la fonction Tableaux croisés et déterminent ses sorties, ce n'est pas le seul. Trois autres paramètres de commande influent sur la façon dont XTSPEC fonctionne.

### **XTFLAG**

Le paramètre XTFLAG est un drapeau, ou un commutateur, qui indique au MSPS s'il faut ou non produire les tables spécifiées par XTSPEC. Si XTFLAG est "désactivé", c'est-à-dire qu'il a la valeur zéro, alors le MSPS ne produit pas les tables définies dans le paramètre XTSPEC.

On peut songer à utiliser XTFLAG dans des cas qui ressemblent à ce qui suit. Un analyste a déjà développé un jeu exhaustif de tables d'analyse, et il a entré le paramètre XTSPEC dans le fichier de paramètres de commande. Maintenant, l'analyste désire faire plusieurs exécutions du modèle pour examiner diverses options de politique, cherchant à trouver une configuration de paramètres de politique qui permettront d'atteindre un certain objectif politique n'exigeant, par exemple, aucune dépense supplémentaire du fédéral. Cependant, les tables de distribution détaillées ne sont pas utiles à l'analyste tant qu'il n'a pas découvert la configuration de paramètres politiques désirée. Par conséquent, pour accélérer l'exécution du modèle, l'analyste laisse XTSPEC inchangé, mais "en coupe l'exécution" pendant les épreuves de sensibilité; pour ce faire, il donne à XTFLAG la valeur zéro pendant ces exécutions. Quand la configuration désirée est connue, l'analyste fait une exécution supplémentaire après avoir activé XTFLAG; il obtient alors les

tables désirées.

## **XTDBLFLAG**

Les variables du MSPS sont stockées et manipulées avec des fonctions mathématiques à faible précision, ce qui donne une précision d'environ 6,5 chiffres, une précision plus que suffisante pour le calcul des impôts et des transferts. Par défaut, la fonction Tableaux croisés utilise aussi les mathématiques simple précision pour le calcul des tables. Un nouveau paramètre, XTDBLFLAG, fera plutôt que la fonction Tableaux croisés utilisera les fonctions mathématiques double précision pour le calcul des tables. Cela fait en sorte que le MSPS utilise plus de mémoire et s'exécute plus lentement, mais cela donne des résultats d'une précision accrue pour certains genres de totalisation.

## **XTLINES**

Le paramètre XTLINES permet à l'utilisateur de déterminer l'aspect du produit de sortie de la fonction Tableaux croisés en ce qui a trait à l'insertion de segments de tables multiples sur une même page. Du point de vue des nombres, le paramètre est interprété comme le nombre de lignes sur une page logique. En général, une valeur plus petite cause l'insertion d'un nouveau saut de page (et l'impression d'une en-tête de page) pour chaque segment de table, tandis qu'une valeur plus grosse amènera la fonction Tableaux croisés à imprimer plusieurs segments de tables avant de provoquer un saut de page et l'impression d'une nouvelle en-tête de page. Si la valeur de XTLINES permet à la fonction Tableaux croisés de placer plusieurs segments de table sur une seule page logique, le logiciel laisse trois lignes vides entre les segments adjacents, pour les séparer.

Le paramètre XTLINES doit être présent dans le fichier de paramètres de commande. L'analyste devrait veiller à ce que XTLINES ait une valeur se situant entre 0 et 32767. Une valeur typique est 66, cette valeur correspondant aux 66 lignes qu'il peut y avoir sur une page de 11 po avec l'espacement standard de six lignes par pouce. L'analyste pourrait choisir d'utiliser une valeur très grosse pour XTLINES afin de supprimer les sauts de page qui rendraient encore plus difficile l'exportation du fichier de sortie vers un progiciel de chiffrier. Si un analyste désire exporter ces tables vers un progiciel de chiffrier, cependant, il devrait songer à utiliser la fonction d'importation qui a été conçue spécifiquement pour cette tâche. On trouvera plus d'information de cette fonction dans le document *Guide d'utilisation des outils*.

## **XTCOLS**

Le paramètre XTCOLS permet à l'utilisateur de déterminer dans une certaine mesure l'aspect de la sortie de la fonction Tableaux croisés en modifiant le compromis qu'il y a entre la présence de lignes multiples dans les en-têtes de colonne et une largeur plus grande des colonnes.

Le paramètre XTCOLS doit être présent dans le fichier de paramètres de commande. L'analyste devrait veiller à ce que XTCOLS ait une valeur se situant entre 80 et 32767. L'interprétation générale de la valeur du paramètre est celle du nombre maximum désiré de colonnes d'impression dans le fichier de sortie. Une valeur typique de XTCOLS est 132, qui correspond au nombre maximal de colonnes d'impression disponibles sur de nombreuses imprimantes.

Même si la fonction Tableaux croisés ne répartira jamais horizontalement une table large en segments multiples, il tentera, si possible, de réduire la largeur des colonnes individuelles de façon que le tableau entier se situe à l'intérieur de la valeur de XTCOLS. Étant donné les contraintes imposées par les nombres qui doivent être imprimés dans la table et les questions statistiques entourant la mise en place d'étiquettes de colonnes sur des lignes multiples, la fonction Tableaux croisés ne réussira pas toujours à conserver la table à l'intérieur de la largeur donnée par XTCOLS; par conséquent, la fonction Tableaux croisés traite la valeur de XTCOLS plus comme un objectif que comme une contrainte absolue.

L'algorithme utilisé par la fonction Tableaux croisés pour choisir les largeurs de colonne dépend de facteurs suffisamment nombreux pour qu'aucune description simple ne soit possible. En général, l'utilisateur devrait entrer une valeur élevée si la largeur de l'imprimante ne pose pas de problèmes et qu'il désire éviter que les étiquettes de colonnes soient sur des lignes multiples. Si, par contre, la largeur de la sortie est un problème et qu'il est acceptable que les étiquettes de colonne soient sur des lignes multiples, une valeur plus petite convient.

Du point de vue général du paramètre XTSPEC dans son ensemble, notre description est maintenant complète : XTSPEC est un contenant où se trouvent les commandes de table individuelles qui font tout le vrai travail. Les limites entre les demandes de table qui constituent l'ensemble sont indiquées par un point-virgule. Dans la section suivante, nous portons notre attention sur les demandes individuelles dont nous décrivons le contenu et la structure interne à un niveau de détail qui convient pour un analyste qui communique ses besoins de totalisation.

## La demande de table

La présente section décrit l'utilisation que fait le MSPS des demandes de table. Rappelons le paramètre XTSPEC de l'exemple donné en introduction :

```
XTSPEC
CF: {units}
    * hdprow+
    * cftype+;
```

La demande de table peut aussi être rédigée sur une seule ligne :

```
CF: {units} * hdprow+ * cftype+;
```

Tous les éléments critiques de la demande de table étaient présents : une unité d'analyse (famille de recensement), une variable totalisée (units), et des variables de classe (hdprow et cftype). Dans la présente section, nous étudierons plus à fond les demandes de table, dont un certain nombre de caractéristiques (utilisation d'expressions, normalisations, étiquetage, etc.) qui n'ont pas été expliquées ci-dessus.

## STRUCTURE GÉNÉRALE

La structure générale d'une demande de table est la suivante : elle commence par une spécification (facultative) de l'unité d'analyse qui est suivie d'un ou de plusieurs niveaux, dont exactement un niveau doit être une spécification des variables de totalisation de la table. Le terme "variable" totalisé est utilisé ici pour les fins de simplicité; cependant, le MSPS reconnaît explicitement soit une variable explicite, soit une expression. L'astérisque sert à séparer les niveaux de la demande, un peu comme le point-virgule sert à séparer les demandes entre elles.

## LES NIVEAUX DE TOTALISATION ET LEUR ORDRE

À quelques exceptions passablement mineures (et quelque peu évidentes) indiquées plus loin, les niveaux, de gauche à droite, d'une demande de table sont considérées comme en ordre "descendant". Les niveaux qui sont donnés en premier dans la demande de table sont "plus élevés" tandis que ceux qui sont donnés par la suite sont "plus bas". Pour qu'un niveau soit "plus bas", dans ce contexte, il faut que le cycle de la table à travers ses catégories soit plus fréquent. Par conséquent, le niveau "le plus bas" est le dernier à droite, ou le dernier niveau de la demande. Ce niveau le plus bas détermine les catégories de colonne de la table, les catégories qui reviennent dans le cycle une fois à chaque rang. Le niveau qui précède le niveau le plus bas est l'avant-dernier; il détermine les catégories de rangs, dont le cycle se déroule une fois pendant chaque segment de la table. Le niveau immédiatement supérieur détermine les catégories de segments, dont le cycle survient seulement entre les segments, et ainsi de suite. **La fonction Tableaux croisés permet jusqu'à six niveaux combinant moins de 260 variables de classe** dans une demande de table, ce qui devrait être plus que suffisant si l'on considère la densité de données réduites qui résulte inévitablement de l'imposition de catégories de plus en plus étroites de données.

## UNITÉS D'ANALYSE

Le MSPS et la fonction Tableaux croisés acceptent cinq types différents d'unités d'analyse. Ces cinq types forment une hiérarchie, une unité plus élevée contenant nécessairement une ou plusieurs occurrences de tous les niveaux inférieurs d'unité d'analyse. Voici les cinq unités d'analyse acceptées, de haut en bas :

### HH (Ménage)

Le ménage (p. ex., appartement, maison en rangée, logement unifamilial, etc.) est l'unité de base de l'Enquête sur la dynamique du travail et du revenu, l'unité sur laquelle la BDSPS est essentiellement basée. La définition du ménage est déterminée par l'emplacement du logement et ne tient pas compte des interrelations de ses membres au-delà du fait qu'ils vivent dans le même logement.

### EF (Famille économique)

Une famille économique, qui peut être identique au ménage, comprend une personne, ou plusieurs personnes apparentées par le sang, le mariage (y compris l'union de fait) ou l'adoption, vivant ensemble comme unité économique à l'intérieur d'un même ménage. Bien qu'un simple ménage puisse contenir des familles économiques multiples, chaque ménage comprend nécessairement exactement une famille économique principale.

### CF (Famille de recensement)

Une famille de recensement, dans le BD/MSPS, qui peut être identique à une famille économique, comprend une personne, le conjoint de la personne, le cas échéant (y compris le conjoint de fait) et l'un ou l'autre de leurs enfants jamais marié qui font partie de la famille économique. Étant donné qu'il est axé sur l'aspect politique, le BD/MSPS traite une personne célibataire comme si elle était une véritable "famille" de recensement; par conséquent, un analyste peut faire une demande de table pour les familles de recensement sans devoir faire une demande parallèle distincte pour faire la totalisation des individus. Bien qu'une famille



économique puisse comprendre plusieurs familles de recensement, chaque famille économique comprend nécessairement exactement une famille de recensement principale.

### **NF (Famille nucléaire)**

Une famille nucléaire, dans le BD/MSPS, est très semblable à la famille de recensement, sauf que la famille nucléaire est définie de façon à exclure tout enfant jamais marié de 18 ans et plus. Conformément à la pratique qui s'installe progressivement dans le domaine de l'impôt et qui traite les individus de 18 ans comme des adultes, tout enfant de ce genre dans le ménage est considéré comme constituant sa propre famille nucléaire. Bien qu'une famille de recensement puisse comprendre plusieurs familles nucléaires, chaque famille de recensement comprend nécessairement exactement une famille nucléaire principale; le chef de la famille de recensement est nécessairement le chef de la famille nucléaire principale.

### **IN (Individu ou personne)**

Le concept de l'individu ou de la personne est évident. Le BD/MSPS contient de l'information substantielle pour chaque personne de 15 ans ou plus. L'existence d'enfants plus jeunes est reconnue par la BD/MSPS du fait qu'elle contient l'information la plus fondamentale à leur sujet (p. ex., l'âge, le sexe et la situation scolaire) dans leur propre enregistrement individuel. Cependant, il y a de toute évidence des limites à l'information que contiennent ces enregistrements. Ainsi, la variable de revenu d'emploi indépendant sera présente, mais aura invariablement la valeur zéro. Chaque famille de recensement ou famille nucléaire comprend une personne désignée chef de famille. Dans la BDSPPS qui est basée sur l'EDTR, quand il y a deux parents dans la famille nucléaire ou la famille de recensement. L'homme est arbitrairement désigné chef de famille.

Le premier élément d'une demande de table, commençant habituellement à la première colonne d'une ligne, est l'unité d'analyse. Cette spécification comprend le code de deux caractères de l'unité désirée (HH, EF, CF, NF, ou IN) qui est suivi du caractère deux-points qui sépare la spécification d'unité du reste de la demande de table. L'utilisateur doit entrer le code de deux caractères en majuscules. Comme on l'a vu ci-dessus, la spécification d'unité est facultative. Si l'utilisateur ne l'indique pas (et ne met pas non plus les deux-points), alors la fonction Tableaux croisés utilise l'individu, "IN:", comme spécification par défaut.

## **VARIABLES DE TOTALISATION**

Pour qu'une demande de table soit significative, elle doit indiquer ce qui doit être totalisé, c'est-à-dire compté ou additionné, dans la table. Par conséquent, (exactement) un des niveaux de la demande de table doit indiquer la variable ou les variables de totalisation.

### **Le niveau de la variable de totalisation d'une demande de table**

Dans la fonction Tableaux croisés, une demande de table amorce la totalisation d'une ou de plusieurs variables de totalisation par rapport aux catégories définies par d'autres variables, des variables de classe. Par conséquent, une demande de table doit comprendre exactement un niveau désignant ces variables de totalisation. L'analyste distingue ce niveau unique en l'insérant entre accolades, c'est-à-dire les caractères { et }. Tel qu'il est décrit ci-dessous, le niveau de la variable de totalisation est "spécial", étant donné qu'il est différent tant dans la forme que dans le contenu des autres niveaux qui désignent les variables de classe.

En général, le contenu d'un niveau de totalisation, c'est-à-dire ce qu'il y a entre les accolades, consiste des expressions ou variables "appropriées". Dans cette application, "appropriées" a trait à des variables d'analyse de base de données et de modèle ou à des expressions qui découlent de variables de modèle et de base de données et qui mènent à des valeurs d'analyse. Les sources possibles de ces valeurs sont données ci-dessous. Chacune des expressions et (ou) variables de totalisation au niveau de totalisation est par la suite soumis à des "qualificatifs" explicites; ces qualificatifs sont décrits ci-dessous.

### **Variables de totalisation multiples dans une demande**

La fonction Tableaux croisés permet à l'analyste d'inclure des variables totalisées multiples dans une seule demande de table. Du point de vue de la syntaxe, l'inclusion de variables multiples est simple. L'analyste met les expressions ou variables désirées à l'intérieur du niveau de totalisation et sépare les expressions ou variables de l'élément par des virgules. Par conséquent, pour totaliser, par exemple, le crédit d'impôt pour enfants ainsi que les prestations de la SV dans la même table, l'analyste utiliserait une expression de totalisation semblable à ce qui suit :

```
{ imctc, imigis };
```

ou

```
{ imctc,  
  imigis };
```

Lorsque l'analyste utilise des expressions ou variables de totalisation multiples à un niveau de totalisation, la liste des expressions ou variables pertinentes agit beaucoup comme les catégories d'un niveau de classification. Par exemple, si un niveau de totalisation à variables multiples est le plus bas niveau de la demande, les expressions et variables de cet élément servent à définir les colonnes de la table. S'il s'agit du niveau précédant le niveau le plus bas, les expressions et variables de totalisation multiples définissent les catégories de rang de la table, et ainsi de suite.

### **Sources des variables de totalisation**

Seuls certains genres de variables et d'expressions sont permis pour les variables de totalisation dans la fonction Tableaux croisés. Heureusement, cet ensemble comprend la plupart de celles qui intéressent l'analyste. Les quatre sources reconnues de "variables" de totalisation sont les suivantes :

### **Variables d'analyse de la BDSPS (base de données) :**

L'analyste peut spécifier, comme variables à totaliser, les variables d'analyse disponibles de la BDSPS. Le mécanisme est alors évident : il suffit à l'analyste d'utiliser le nom de la variable dans la demande de table. Par conséquent, en utilisant, par exemple, la variable d'analyse idiemp comme variable totalisée, on créerait une table qui contiendrait les cellules correspondant au revenu d'emploi (payé). Certaines des ces variables seront définies au niveau de l'individu (les variables id), tandis que d'autres seront définies au niveau du ménage (les variables hd et fx). D'autres variables "calculées" existeront dans la base de données à des niveaux intermédiaires, p. ex., les variables de famille de recensement cf ou de famille nucléaire nf.

On trouvera plus d'information sur les variables de base de données dans le document [Guide des variables](#). Il est évident que ce ne sont pas toutes ces variables qui conviendront comme variable

de totalisation dans une demande de table. Par exemple, la variable `hdwghh` (qui fournit le poids du ménage) serait un choix illogique comme variable de totalisation (puisqu'elle serait pondérée par elle-même pendant le processus de totalisation). La question importante des variables "appropriées" et de leur traitement dans la production d'une table à plusieurs niveaux d'unité d'analyse est traitée plus en détail un peu plus loin dans le Guide. En général, cependant, la fonction Tableaux croisés permet à l'analyste de totaliser toute variable de base de données d'analyse que l'analyste est susceptible de considérer comme raisonnable dans le cadre d'une analyse de politique.

### **Variables d'analyse du MSPS (modélisées) :**

C'est l'essence même du MSPS, tant en mode "boîte noire" qu'en mode "boîte de verre", de calculer les variables qui intéressent l'analyste. Habituellement, mais non exclusivement, ces variables calculées sont les impôts, les taxes et les présentations pertinentes pour les unités. Tout comme il le ferait avec les variables d'analyse disponibles directement de la base de données, l'analyste peut totaliser les variables d'analyse modélisées appropriées simplement en utilisant leur nom dans la demande de table. Par exemple, en utilisant `imctc` comme variable de totalisation, on créerait une table qui accumulerait les montants des crédits d'impôt pour enfant payables.

Les variables modélisées disponibles en mode "boîte noire" du MSPS sont aussi expliquées dans le document *Guide des variables*. Comme c'est le cas avec les variables de base de données, ce ne sont pas toutes ces variables qui conviennent comme variables de totalisation dans une demande de table. En général, cependant, la fonction Tableaux croisés permet à l'analyste de totaliser toute variable d'analyse modélisée que l'analyste est susceptible de considérer comme raisonnable dans le cadre d'une analyse de politique.

### **Expressions d'analyse de demande de tables :**

L'autre capacité de la fonction Tableaux croisés de totaliser des expressions "arbitraires" fournies par l'analyste (exprimées à l'aide de variables de base de données, modélisées et (ou) `ex`) permet de simplifier l'expression et d'économiser beaucoup de temps. Cette capacité signifie parfois que l'analyste n'est pas obligé (a) de revenir dans le modèle, (b) de modifier le code de source du modèle pour créer la variable désirée, puis (c) de recompiler le modèle en rétablissant les liens, et (d) de réexécuter le modèle en demandant la table désirée, le tout dans l'ordre, pour totaliser quelque chose qui est facilement spécifié à partir de variables existantes. Le qualificatif "parfois" reflète bien la dépendance de cette expression sur le niveau de cumul; c'est-à-dire que les entrées de la table sont calculées comme des expressions d'agrégats plutôt que comme des agrégats d'expressions.

Pour certaines expressions, par exemple des expressions purement d'addition, la distinction n'est pas vraiment importante; cependant, pour d'autres formes, l'interprétation de la table dépendra essentiellement de cette distinction.

La fonction Tableaux croisés offre une souplesse considérable dans la définition de ces expressions "à la volée". Essentiellement, l'expression de totalisation peut être toute expression C "appropriée". Dans ce contexte, "appropriée" signifie des expressions (d'analyse)

dans lesquelles les opérateurs unaires ou binaires suivants sont soit mathématiques (+, -, \*, /), soit logiques (<, <=, ==, >=, >, !=, !, && et ||). Les opérateurs logiques, bien que permis par la syntaxe dans une expression de totalisation, sont rarement utilisés puisque les variables d’une expression de totalisation ont trait à des quantités agrégées.

L’analyste peut en outre utiliser les parenthèses pour faire des regroupements au besoin. Cependant, l’expression peut utiliser, dans ses arguments, seulement des variables d’analyse; cette restriction n’est cependant pas particulièrement fastidieuse puisque pratiquement toutes les variables que l’analyste peut désirer utiliser comme argument dans une expression sont “déjà” définies comme variables d’analyse.

La fonction Tableaux croisés créerait toute cellule donnée pour une expression en totalisant d’abord, pour tous les cas se retrouvant dans la cellule et au niveau d’unité d’analyse défini par la table, une valeur pour chacun des arguments symboliques de l’expression. Pour arriver à la valeur numérique qui se retrouvera dans les cellules, la fonction Tableaux croisés évalue alors l’expression en y insérant les diverses valeurs totalisées.

Donc, par exemple, une expression totalisée de la forme {imctc/immtot} produirait le rapport des crédits d’impôt pour enfants (pour les unités se trouvant dans cette cellule) au total des revenus pour ces unités. Elle ne produirait pas la somme des fractions imctc/immtot pour toutes ces unités, cela exigeant la création d’une variable utilisateur. Il faut noter que cette interprétation que la fonction Tableaux croisés fait de l’expression est précisément celle que l’analyste désire habituellement.

### **Variables de totalisation spéciales**

En plus des variables de totalisation “habituelles” décrites ci-dessus, la fonction Tableaux croisés permet à l’analyste de spécifier certaines variables de totalisation (ou expressions qui les comprennent) qui sont “spéciales” dans le sens qu’elles reçoivent un traitement inhabituel.

#### *Units et persons :*

Même si les variables units ne figurent pas comme variable de base de données ou variable modélisée, il est suffisamment probable que l’analyste désire totaliser les nombres (pondérés) des diverses unités d’analyse pour que units ait été incluse comme une variable permise dans toute expression ou variable de totalisation. Lorsqu’elle est utilisée, elle désigne l’unité d’analyse spécifiée au début de la demande de table. Par conséquent, si l’analyste totalise units et que le niveau d’unité d’analyse est la famille de recensement, le résultat sera interprété comme compte du nombre (pondéré) de familles de recensement.

De même, l’analyste désirera souvent totaliser les nombres de personnes (pour compter, p. ex., le nombre de personnes vivant dans les unités au-dessus des seuils de pauvreté de Statistique Canada) ou utiliser le nombre de personnes dans les expressions à totaliser (p. ex, pour évaluer le revenu par habitant dans une famille de recensement). Pour cette fin, la variable *persons* est disponible. Lorsqu’elle utilisée, elle désigne le nombre de personnes dans l’unité d’analyse spécifiée au début de la demande de table. Par conséquent, pour une spécification de sélection appropriée, la totalisation de l’expression {immdisp/persons} fera en sorte que les entrées dans les cellules soient correctement interprétées comme le “revenu disponible par habitant”.

## Les variables non soumises à la pondération :

Trois des variables de la BDSPS, *spsdrecs*, *scfrecs* et *fxvrecs*, font exception à la règle selon laquelle les variables totalisées sont pondérées pendant le processus de totalisation. Ces trois variables ont toutes été incluses dans la base de données à la seule fin de permettre à l'analyste d'évaluer le nombre de cas qui sous-tend une entrée de table donnée, particulièrement quand la densité des données pose un problème d'interprétation des résultats de la totalisation.

Quand il interprète des tables qui utilisent l'une ou l'autre de ces variables, l'analyste doit reconnaître que les entrées de cellules sont sensibles au niveau de cumul de la table. Par conséquent, il y a des enregistrements pertinents beaucoup plus nombreux quand l'unité de totalisation est l'individu que quand c'est la famille économique.

*spsdrecs* : Utilisée en l'absence du qualificatif "M", la variable de totalisation *spsdrecs* permet à l'analyste de totaliser le nombre d'enregistrements de la BDSPS sur lequel une cellule ou un total est basé. Ce compte est basé sur le nombre de ménages qu'il y a dans la BDSPS.

*scfrecs* : Utilisée sans le qualificatif "M", la variable de totalisation *scfrecs* permet à l'analyste de totaliser le nombre d'enregistrements EDTR originaux sur lequel une cellule ou un total est basé. Ce compte est basé sur le nombre d'enregistrements de ménages dans l'Enquête sur la dynamique du travail et du revenu qui constitue la base de la BDSPS. Le nombre le plus petit d'enregistrements de l'EDTR, par rapport aux enregistrements de la BDSPS, reflète la reprise, pendant le développement de la BDSPS, de certains des enregistrements de l'EDTR afin d'enrichir l'information pour certaines unités ayant des revenus plus élevés ou des prestations d'A.-C.

*fxvrecs* : Utilisée sans le qualificatif "M", la variable de totalisation *fxvrecs* permet à l'analyste de totaliser le nombre d'enregistrements EDM originaux sur lequel une cellule ou un total est basé. Ce compte est basé sur les enregistrements d'unité de dépenses de la base de données EDM qui fournit les variables de dépenses pour les enregistrements de la BDSPS. Il est clair que, du fait que la base de données EDM est beaucoup plus petite que l'EDTR, il se peut qu'on ait eu à utiliser un même enregistrement EDM afin de fournir les régimes de dépenses pour plusieurs enregistrements EDTR/BDSPS.

Le qualificatif "M" dont il est question dans les trois paragraphes ci-dessus demande que la fonction Tableaux croisés exécute une normalisation dans l'une des dimensions de classification de la table. Par conséquent, si le qualificatif "M" est présent, la table devrait donner non le nombre de cas dans les cellules, mais leur répartition en pourcentage. Les divers qualificatifs et les rôles qu'ils jouent dans l'adaptation des totalisations sont repris à la section suivante.

### Qualificatifs des variables de totalisation

La fonction Tableaux croisés permet à l'analyste, s'il le désire, de "qualifier" une variable de totalisation de plusieurs façons qui peuvent améliorer sa lisibilité ou son utilité dans la table qui sera produite. En utilisant ces qualificatifs, l'analyste peut, par exemple, obtenir des étiquettes plus informatives, forcer l'ajustement des entrées des cellules ou amorcer la normalisation. Ces genres d'amélioration valent seulement pour les variables de totalisation; on ne s'attend pas, par

exemple, à ajuster ou à normaliser une variable de classe.

Purement facultative : l'utilisation des qualificatifs est entièrement facultative. La fonction Tableaux croisés a des valeurs par défaut pour toutes les caractéristiques déterminées par les qualificatifs, certaines de ces valeurs par défaut étant conditionnelles aux caractéristiques de la table demandée (p. ex., la nature de la variable ou de l'expression qui est totalisée). Dans la plupart des cas, la fonction Tableaux croisés produit des tables dont l'aspect est très raisonnable, même si aucun qualificatif n'est spécifié. Cependant, l'analyste peut désirer utiliser des qualificatifs pour mettre en oeuvre certaines caractéristiques (étiquetage, ajustement, valeurs marginales, etc.) qui rendent la table plus utile.

Aspect : en lui-même, un qualificatif ressemble un peu à une instruction d'affectation. Il consiste en une lettre majuscule ( L , S , P , ou M ) qui indique le qualificatif qui est invoqué, et qui est suivi immédiatement (sans espaces) du caractère égal ( = ), immédiatement suivi d'une chaîne ou d'une valeur numérique qui doit être affectée au qualificatif. Ainsi, le qualificatif S=6 indique que la variable totalisée doit se retrouver dans la table en unités de un million.

Syntaxe/emplacement : l'analyste indique la présence des qualificatifs pour une variable totalisée en plaçant un deux-points après l'expression ou la variable totalisée. Les qualificatifs suivent alors les deux-points et précèdent (1) la virgule qui sépare la variable de totalisation des variables de totalisation supplémentaires qui suivent ou (2) l'accolade qui marque la fin de la partie niveau de totalisation de la demande de table.

S'il y a plusieurs qualificatifs, ils sont séparés les uns des autres et des autres parties de l'instruction par un "blanc" approprié (espaces, retours de chariot et tabulations). Par exemple, un niveau de totalisation de { units : S=6 P=1 } indique que la fonction Tableaux croisés doit compter les unités d'analyse pondérées et les données en millions, avec une décimale. Une entrée de 6.1 dans la table indiquerait alors la valeur 6 100 000.

Chacune des variables ou expressions de totalisation au niveau de totalisation d'une demande de table peut avoir son propre ensemble de qualificatifs.

Voici la description des quatre qualificatifs possibles ( L , S , P et M ).

1. L : Étiquette -- Le qualificatif L permet à l'analyste de préciser, pour la table de sortie, une étiquette textuelle pour la variable totalisée. Cette étiquette prend la forme d'une chaîne de texte délimitée par des guillemets. La chaîne de texte, utilisée pour l'étiquetage des tables, peut compter jusqu'à 40 caractères. Si la fonction Tableaux croisés utilise la chaîne dans l'en-tête d'une colonne, et que la chaîne est trop longue pour la largeur de la colonne, la fonction Tableaux croisés coupera la chaîne de façon intelligente, choisissant les points de coupure entre les mots, si cela est possible. Lorsque la chaîne est utilisée comme en-tête de rang, elle n'est jamais coupée; le corps de l'ensemble de la table est plutôt déplacé vers la droite pour laisser la place à l'en-tête.

Quand il est présent dans la demande, le qualificatif L demande à la fonction Tableaux croisés de remplacer l'étiquette par défaut qui aurait été obtenue s'il n'y avait pas eu de

qualificatif. Par défaut, la fonction Tableaux croisés utilise l'étiquette obtenue de la base de données ou du modèle, s'il y en a une. S'il n'y a pas d'étiquette par défaut, alors la fonction Tableaux croisés utilise le nom de la variable totalisée, ou la chaîne de texte qui définit l'expression totalisée. Comme on le voit ci-dessous, les qualificatifs S et M ont aussi une influence sur la chaîne de texte qui sera utilisée comme étiquette.

Par exemple, le niveau de totalisation :

```
{immdisp/persons: L="Rev. disp. par habitant"};
```

ferait en sorte que l'expression totalisée aurait pour étiquette la chaîne de texte "Rev. disp. par habitant" plutôt que par la chaîne par défaut, qui est beaucoup plus nébuleuse, "immdisp/persons".

2. S Facteur d'ajustement -- Le qualificatif S permet à l'analyste de spécifier un facteur d'ajustement pour la présentation de l'expression de la variable de totalisation. Ce facteur d'ajustement prend la forme d'un exposant entier; par exemple, S=3 indique que l'analyste désire voir les résultats présentés en milliers. L'analyste devrait éviter des facteurs d'ajustement qui se situent à l'extérieur de la gamme [-6, 9].

La valeur par défaut est 6, puisque la plupart des totalisations visent la somme de montants en dollars. Cependant, si l'analyste fait la totalisation *units* ou *persons* et que l'opération de division (/) n'est pas utilisée, alors la valeur par défaut devient 3. L'utilisation de l'opération de division ou la totalisation de l'une ou l'autre des trois variables spéciales (*spsdrecs*, *scfrecs* et *fxvrecs*) donne zéro à la valeur par défaut (il n'y a pas d'ajustement). En outre, l'utilisation du qualificatif "M" remplace le facteur d'ajustement par défaut, lui donnant la valeur -2 pour refléter la présentation typique des valeurs marginales en pourcentage. Une spécification explicite d'un qualificatif S remplacera bien sûr ces valeurs par défaut.

Le facteur d'ajustement touche les nombres qui se retrouvent dans les cellules de la table produite mais, en outre, il a naturellement une influence sur l'étiquette qui s'applique à l'expression ou à la variable de totalisation. En effet, étiquetage de la table doit indiquer quand un facteur d'ajustement a été appliqué et, si c'est le cas, quel facteur a été utilisé en particulier. Cet effet se produit, sur les étiquettes que le facteur d'ajustement ait été spécifié explicitement avec le qualificatif S ou qu'il ait été sélectionné comme valeur par défaut par la fonction Tableaux croisés.

En général, si un facteur non-zéro est appliqué, le texte de l'étiquette recevra un suffixe qui indiquera le facteur d'ajustement numérique utilisé en particulier, par exemple il y aura ajout de (000) pour un facteur d'ajustement de 3. Cependant, pour certaines valeurs utilisées couramment, d'autres suffixes sont subdivisés pour améliorer la lisibilité. Ces substitutions spécifiques sont "(%)" lorsque S=-2, "(M)" lorsque S=6 et "(B)" lorsque S=9.

Prenons un cas concret : le niveau de totalisation de {idinvnd:S=9} totaliserait des revenus en intérêts et autres revenus de placements et présenterait les résultats en milliards de dollars. On suppose que cette spécification serait utilisée seulement dans une table ayant un nombre relativement faible de cellules de façon que la plupart des entrées donneraient des valeurs ajustées supérieures à 1.0.

3. **P Décimales** -- Le qualificatif P permet à l'analyste de déterminer la présence du point de décimale dans les entrées de tables ainsi que le nombre de décimales après le point. P doit être égal à 0 (zéro) ou à quelque nombre entier positif. Une valeur zéro de P indique qu'il n'y a pas de point de décimale pour la variable totalisée; bien sûr, en l'absence du point de décimale, il n'y a pas de décimale. Une valeur positive de P indique le nombre de décimales après le point. L'analyste devrait restreindre la valeur de P à l'intérieur de la plage [0, 8].

Comme c'est le cas avec le qualificatif S, le qualificatif P a une "valeur par défaut intelligente". La "valeur par défaut" est 1, mais la fonction Tableaux croisés remplace cette valeur selon la nature de l'expression ou de la variable de totalisation. Spécifiquement, si l'analyste utilise l'opérateur de division et qu'il m'emploie pas *units* ou *persons* dans l'expression de totalisation, alors la fonction Tableaux croisés suppose que l'analyste crée un rapport et fixe la valeur de P à 4, en déterminant qu'il s'agit d'une valeur raisonnable pour la plupart des fractions. Cependant, (1) si l'analyste utilise l'opérateur de division et totalise avec *units* ou *persons*, ou (2) si l'analyste totalise les valeurs *spsdrecs*, *scfrecs* ou *fxvrecs*, la valeur par défaut de P est 0 (zéro). En outre, si l'analyste utilise le qualificatif M, la fonction Tableaux croisés donne 1 comme valeur par défaut de P. Une spécification explicite du qualificatif P remplace bien sûr ces valeurs par défaut.

4. **M Marge** -- Le qualificatif M permet à l'analyste de normaliser l'expression ou la variable totalisée par rapport à (exactement) une des variables de classe qu'il y a dans la demande de table. Cette normalisation convient lorsque l'analyste est intéressé dans les distributions en pourcentage ou dans les tailles relatives des variables; la production automatique des valeurs marginales par la MSPS est de loin plus commode qu'un mécanisme de rechange comme le calcul manuel.

La normalisation doit être spécifiée relativement à exactement une des variables de classe définissant un niveau de la demande de table. La spécification de cette variable de classe par l'analyste définit pour la fonction Tableaux croisés la dimension qui doit être visée par la normalisation. La fonction Tableaux croisés ne permet pas des normalisations multiples sur une seule variable de totalisation. Par conséquent, l'analyste pourrait générer, dans un tableau donné, soit des pourcentages de rang, soit des pourcentages de colonne, mais pas les deux. Cependant, pour poursuivre cet exemple jusqu'à la fin, on pourrait utiliser la demande de totalisation

```
CF: {units,  
     units:M=cftype,  
     units:M=hdprov}  
    * cftype+  
    * hdprov+;
```

pour générer trois segments de tables parallèles qui présentent collectivement les nombres d'unités de famille de recensement et les pourcentages de rang et de colonne pertinents.

De même, actuellement, il n'y a aucune façon simple de produire des "pourcentages de segment" pour une table bidimensionnelle, pour, par exemple, obtenir les contributions relatives de combinaisons de tailles de famille et de tailles de places de résidence par rapport à la population de familles pauvres. (Bien sûr, l'analyste pourrait générer une nouvelle variable de classe composite qui aurait une catégorie pour chaque combinaison, mais ce



mécanisme n'est pas particulièrement simple pour l'utilisateur.)

Une utilisation typique du qualificatif est la production de pourcentages de rang ou de colonne dans une table. Si l'analyste, par exemple, fait la totalisation de *mdctc* et emploie le qualificatif *M=cftype*, alors la table produite donnera les pourcentages des crédits d'impôt pour enfants reçus par les divers types de famille.

La mécanique de la normalisation est relativement simple et consiste principalement dans le calcul d'un rapport. Le numérateur du rapport est la valeur qu'il y aurait eu dans la cellule s'il n'y avait pas eu de normalisation. Le dénominateur est la valeur de l'entrée qu'il y aurait pour l'entrée "toutes les catégories de la variable de classe de normalisation réunies" telle qu'elle l'aurait été s'il n'y avait pas eu de normalisation (et s'il avait fallu la calculer). Par conséquent, si l'entrée "TOUTES" pour la variable de classe de normalisation est vraiment demandée dans une table (avec le suffixe + décrit ci-dessous), sa valeur normalisée devrait toujours être égale à 1.0.

Comme on l'a décrit ci-dessus, l'utilisation du qualificatif *M* a une influence sur les valeurs par défaut des autres caractéristiques de la table. Parce que l'analyste désire le plus souvent voir la normalisation sous forme de pourcentage, l'utilisation de *M* fixe la valeur par défaut du facteur *S* (ajustement) à -2. En outre, l'utilisation du qualificatif *M* établit une valeur par défaut de 1 pour le facteur *P* (nombre de décimales). Il faut se rappeler qu'un analyste qui n'accepte pas ces choix peut facilement les remplacer en donnant une valeur explicite des qualificatifs *S* et *P* à l'expression ou à la variable totalisée pertinente.

Le qualificatif *M* a en outre une influence sur l'étiquette utilisée pour la variable de totalisation parce que les personnes qui consulteront la table produite doivent comprendre qu'il y a eu normalisation et doivent savoir quelle variable de classe a été utilisée pour cette normalisation. La fonction Tableaux croisés comble ces besoins en plaçant le nom de la variable de normalisation entre crochets à la fin de la chaîne de texte qui étiquette la variable de totalisation (après tout suffixe ayant trait au facteur d'ajustement).

### **Contenu de cellules et étiquetage pour les expressions et variables de totalisation**

Maintenant que l'on a fini l'étude des qualificatifs, il est maintenant possible de s'attaquer à la question fondamentale de la relation qu'il y a entre la variable de totalisation pouvant être qualifiée et le contenu d'une cellule de la table produite. Il y a quatre cas distincts :

#### **a) Variables simples sans qualificatif *M* :**

Il s'agit du cas le plus courant. L'analyste désire totaliser une seule variable, plutôt qu'une expression qui peut toucher d'autres variables constantes. L'entrée de cellules sera simplement la somme (habituellement pondérée par les poids d'enregistrement) des valeurs de cette variable d'analyse dans tous les enregistrements respectant les critères d'inclusion dans la cellule. Par exemple, si *immicons* est la variable de totalisation, alors l'entrée de cellules correspondra au total du revenu consommable.

Si l'analyste n'a pas utilisé le qualificatif *L*, l'étiquette comprend habituellement l'étiquette de la variable, s'il y en a une. S'il n'y en a pas, alors l'étiquette par défaut est le nom de la variable.

Pour les facteurs d'ajustement autres que zéro, l'étiquette comprendra un suffixe qui reflète le facteur appliqué à l'entrée.

#### **b) Variables simples avec qualificatif M :**

Si l'analyste totalise une variable simple (plutôt qu'une expression), mais qu'il utilise le qualificatif M, l'entrée de cellule n'est plus la simple somme de la variable, mais elle devient la distribution en pourcentage dans la dimension de la variable de classement donnée par le qualificatif. Par exemple, si l'analyste utilise la variable *immicons*, mais qu'il donne le qualificatif M=hdprov, alors les entrées de cellules seront les pourcentages de contribution de chacune des provinces au revenu consommable.

Si l'analyste n'a pas utilisé le qualificatif L, l'étiquette comprend habituellement l'étiquette de la variable, s'il y en a une. S'il n'y en a pas, alors l'étiquette par défaut est le nom de la variable. Si le facteur d'ajustement n'est pas égal à zéro, alors l'étiquette comprend un suffixe qui reflète le facteur appliqué à l'entrée. Dans tous les cas, l'étiquette comprendra un suffixe comprenant entre crochets le nom de la variable de classe utilisée pour le qualificatif M.

#### **c) Expressions sans qualificatif M :**

Si l'analyste totalise une expression (par rapport à une variable de base de données ou à une variable modélisée), mais qu'il n'utilise pas de qualificatif M, la fonction Tableaux croisés accumule les sommes appropriées pour toutes les variables de l'expression, puis génère l'entrée de cellule en calculant les résultats de l'expression avec les sommes comme argument. Elle ne fait pas la somme des résultats de l'expression telle que calculée pour chacun des enregistrements individuels pondérés. Il faut noter que c'est exactement cela que l'analyste est susceptible de vouloir. Par exemple, si le revenu moyen est désiré, alors l'analyste désire diviser le total du revenu par le total des unités, exactement ce que fait la fonction Tableaux croisés. Par contre, l'analyste n'est fort probablement pas intéressé à voir la somme des rapports individuels pour les unités pertinentes.

Si l'analyste n'a pas utilisé le qualificatif L, l'étiquette comprend la chaîne de texte utilisée pour spécifier l'expression. L'utilisation d'un qualificatif L remplace bien sûr la chaîne de texte. Si le facteur d'ajustement correspondant à l'entrée est non-zéro, l'étiquette comprend aussi un suffixe qui reflète le facteur appliqué à l'entrée.

#### **d) Expressions avec qualificatif M :**

La situation la plus compliquée se produit lorsqu'un analyste totalise une expression, mais qu'il lui applique le qualificatif M. Dans ce cas, la fonction Tableaux croisés calcule essentiellement l'entrée de cellule en utilisant la méthode "expression de sommes", comme en c), mais, avant de l'insérer dans la table, il lui applique ensuite la méthode "normalisation par la variable de qualificatif", comme en b).

La capacité de la fonction Tableaux croisés de combiner la souplesse des expressions à une normalisation peut se révéler très puissante. Par exemple, supposons que l'expression de l'utilisateur totalise la partie des familles économiques dont le revenu tombe sous le seuil de

pauvreté. Si l'utilisateur invoque le qualificatif M pour tous les types de famille, alors la table produite donne les incidences relatives de ces situations de faible revenu pour le type de famille, normalisée en fonction de l'incidence globale.

Si l'analyste n'a pas utilisé un qualificatif L, l'étiquette comprend la chaîne de texte utilisée pour spécifier l'expression à totaliser. L'utilisation d'un qualificatif L explicite remplace cette valeur par défaut. En outre, si le facteur d'ajustement n'est pas zéro, alors l'étiquette comprend un suffixe qui reflète le facteur appliqué à l'entrée. Dans tous les cas, l'étiquette se termine par un suffixe formé du nom, entre crochets, de la variable de classe utilisée dans le qualificatif M.

## **VARIABLES DE CLASSE**

Tout comme le niveau de la variable de totalisation détermine le contenu de ce qui sera totalisé, le niveau de la variable de classe indique à la fonction Tableaux croisés la façon dont le contenu doit être ventilé, par exemple la forme de la table, quand au nombre de rangs, de colonnes et de segments. La présente section décrit la nature de ces niveaux de classe, leur source (par rapport aux variables de classe qui les comprennent) et la manière dont ils influent sur les tables produites.

### **Nature des variables de classe**

Les variables de classe décrivent à la fonction Tableaux croisés la situation de tout cas donné dans un ensemble de catégories exhaustives et mutuellement exclusives, par exemple province de résidence ou sexe d'un individu. Par conséquent, ces variables sont parfois appelées "variables de catégorie". Ces variables font en sorte que tout enregistrement donné se retrouve dans exactement une catégorie nominale. Puisque le nombre de catégories doit être relativement petit pour être utile dans une table, les valeurs de ces variables tendent à être des nombres entiers non négatifs et petits. Cette situation de nombres entiers nominaux des variables de classe contraste fortement avec le statut analyse requis pour les variables de totalisation. Par conséquent, la fonction Tableaux croisés impose certaines exigences sur les variables qu'elle accepte comme variables de classe aux fins de la construction des tables. Les sources permises de ces variables de classe sont décrites un peu plus loin dans la présente section.

### **Contribution à la dimensionnalité des tables**

En général, ce sont les variables de classe qui déterminent la forme des tables de sortie, chaque variable de classe ajoutant une dimension à la table. (Il faut se rappeler que l'utilisation d'expressions ou de variables totalisées multiples contribue à une autre dimension de la table.) Par conséquent, les catégories des variables de classe correspondent habituellement aux rangs, aux colonnes ou aux segments d'une table.

### **Nature des niveaux de classe**

Les niveaux de classe d'une demande de table sont facultatifs. Bien qu'une demande de table comprenne habituellement un ou plusieurs de ces niveaux, aucun de ces niveaux n'est nécessaire. Cependant, chaque niveau de classe qu'il y a doit être constitué d'une seule variable de classe.

Cette exigence du MSPS qu'il y ait exactement une variable de classe par niveau contraste avec certains progiciels de statistiques comme SAS, TPL ou SPSS dans lesquels l'analyste peut demander des tables parallèles dans une même demande de table (p. ex., des tables distinctes

donnant le revenu comme une fonction de la province, du type de famille, de la situation de pauvreté et de l'appartenance à la population active). Dans le MSPS, l'analyste devrait produire des tables parallèles par des demandes de table parallèles qui auraient la même variable totalisée (revenu), mais des variables de classe différentes (province, etc.).

### **Sources des variables de classe**

Tout comme le MSPS impose certaines restrictions sur les variables qui peuvent être utilisées comme variables de totalisation et qu'il les accepte seulement de certaines sources, il acceptera comme variables de classe seulement certaines variables de certaines sources. (Des restrictions similaires s'appliquent dans les fonction de totalisation d'autres progiciels comme SAS ou SPS.) Dans le MSPS, ces sources de variables de classe sont les suivantes :

### **Variables de classe de base de données**

Certaines variables de la BDSPS sont, de façon inhérente, des variables de classe, par exemple la province (variable *hdprov*) ou la variable codée pour la grosseur du lieu de résidence (variable *hdurb*). L'analyste peut utiliser l'une ou l'autre de ces variables en entrant simplement le nom de la variable comme un niveau dans la demande de table. Une liste des variables de classe de base de données est disponible comme élément de ""; on trouvera une description complète des variables elles-mêmes, ce qui comprend leur catégorie propre, dans le document *Guide des variables*.

### **Variables de classe modélisées**

D'autres variables de classe sont définies dans le MSPS; par exemple, le modèle sous-jacent crée des variables comme le statut nominal de déclarant imposable (variable *imfiler*) et la classification nominale des familles par catégorie relativement à l'administration de la SV (variable *imgistyp*). L'analyste peut utiliser l'une ou l'autre de ces variables en entrant simplement le nom de la variable comme un niveau de la demande de table. La liste des variables de classe modélisées disponible est disponible comme élément de ""; on trouvera une description complète des variables elles-mêmes, ainsi que de leur catégorie propre, dans le document *DB/MSPS - Guide des variables*.

### **Variables de classe définies par l'utilisateur**

Le BD/MSPS ne limite pas l'analyste au choix des variables de classe parmi celles qui sont disponibles dans la base de données ou le modèle. L'analyste peut obtenir les variables de ces sources, mais il peut aussi créer de nouvelles variables de classe personnalisées qui seront acceptées par la fonction Tableaux croisés. L'analyste crée ces nouvelles variables de classe définies par l'utilisateur, les variables *uvar*, dans le fichier de paramètres de commande (*.cpr*). En général, une variable définie par l'utilisateur est définie par rapport aux valeurs de certaines variables d'analyse ou variables entières "définies auparavant", que ce soit de variables de base de données ou des variables modélisées.

On trouvera une description de la capacité de UVAR dans le document *Guide d'utilisation*. Il suffit de noter ici que cette capacité donne à l'analyste la possibilité de spécifier la manière dont le continuum d'une variable existante peut être subdivisé (par une série finie de valeurs de points

de découpage numériques donnés explicitement) pour créer un ensemble de domaines sans chevauchement et exhaustifs ainsi que mutuellement exclusifs qui correspondent aux catégories de variables de classe.

### **Création d'une catégorie sommaire "Tous"**

L'analyste désire parfois demander une table qui inclut les totaux pour une ou plusieurs variables de classe. Par exemple, même si une table pourrait utiliser la province de résidence comme variable de classe, l'analyste pourrait aussi désirer voir les résultats correspondants pour l'ensemble du Canada. La fonction Tableaux croisés du MSPS permet à l'analyste de faire cette demande en entrant le caractère + immédiatement après la variable de classe avec la réorganisation des niveaux s'il le désire. Par conséquent, un niveau de classe `hdprov+` non seulement ventilerait la variable totalisée entre les provinces, mais produirait aussi une catégorie "Tous" (tout le Canada). Si jamais la variable de classe ne possède que deux catégories, la fonction Tableaux croisés utilise "Les deux" au lieu de "Tous".

La fonction de demande de la catégorie "Tous" est à la fois facultative et générale. L'utilisateur peut la demander pour un sous-ensemble de variables de classe d'une demande de table (y compris les cas extrêmes de "Aucun d'eux" ou "Tous"), et peut l'utiliser, qu'un qualificatif M soit utilisé ou non pour la variable totalisée.

### **Étiquetage des variables de classe**

Les variables de classe contribuent de deux façons aux étiquettes des tables produites par la fonction Tableaux croisés. En premier lieu, comme avec d'autres variables du BD/MSPS, les variables de classe ont des chaînes de texte qui indiquent leur nature générale, p. ex., "Province" pour la variable `hdprov`. Ces descriptions de variables se retrouvent dans les tables pour l'identification des variables de classe qui se retrouvent variables de commande pour la table. Pour les variables de classe de base de données et de modèle, ces étiquettes sont définies dans la base de données ou le modèle, respectivement. S'il n'y a pas de description de variable définie pour une variable de classe, alors la fonction Tableaux croisés prend comme étiquette par défaut le nom de la variable de classe.

L'emplacement de ces étiquettes de variables est tout ce qu'il y a de plus simple. Par exemple, l'étiquette de la variable de rang se trouvent dans le coin supérieur gauche de la table ou du segment de table, tandis que la description des variables de colonne est imprimée juste au-dessus de la table ou du segment de table. Lorsqu'une table s'étend compte plusieurs segments, alors l'étiquette d'une variable de classe qui est constante à l'intérieur d'un segment se retrouve comme élément de l'information qui étiquette le segment au complet.

En second lieu, les variables de classe ont des étiquettes qui correspondent aux valeurs individuelles, aux gammes de valeurs ou aux classes de variables. Ces "étiquettes de classe" constituent habituellement une grande partie de l'information d'étiquetage d'une table ou d'un segment de table. Pour les variables de modèle ou de base de données, les étiquettes de classe sont prédéfinies par l'utilisateur et sont habituellement textuelles, par exemple "Manitoba" comme étiquette de classe pour une valeur particulière de la variable `hdprov`.

### **Niveau d'analyse**

Chaque variable de classe est à l'origine définie à l'un des cinq niveaux de famille d'analyse du BD/MSPS : individu, famille nucléaire, famille de recensement, famille économique et ménage. Par exemple, *hdprov* (province de résidence) est définie au niveau du ménage, *cfnkids* (nombre d'enfants dans la famille de recensement) est définie au niveau de la famille de recensement et *idcfrh* (relation de l'individu avec le "chef" de la famille de recensement) est définie au niveau de l'individu. Tel qu'indiqué à la section 5.3, le premier élément d'une demande de table (IN: , NF: , CF: , EF: , HH:) indique le niveau d'analyse de la famille. La présente section décrit ce qui arrive lorsque le niveau d'analyse d'une demande de table diffère du niveau d'analyse de l'une des variables de classe de la demande.

Si le niveau d'analyse d'une variable de classe est "plus élevé" que le niveau d'analyse d'une demande de table, la valeur de la variable est bien définie et est celui de l'unité "plus élevée". Ainsi, la demande de table

```
CF: hdprov * {units}
```

spécifie le compte des familles de recensement par province de résidence. Même si *hdprov* est définie au niveau du ménage, il n'y a aucune ambiguïté, puisque la province de résidence d'une famille de recensement est la province de résidence du ménage qui la contient. Cependant, si l'on prend la demande de table suivante :

```
HH: cfnkids * {units}
```

le compte des ménages a été demandé, réparti par le nombre d'enfants dans les familles de recensement. Si un ménage contient seulement une famille de recensement, la signification de *cfnkids* est claire, mais que se passe-t-il si le ménage comprend deux familles de recensement, avec des nombres d'enfants différents? Dans ces cas, le MSPS désigne une des familles de recensement comme famille de recensement "de référence" dans le ménage et prend la valeur de *cfnkids* qui est celle de la famille de recensement de référence. La famille de recensement de référence du ménage est la première famille de recensement du ménage.

Pour définir en général ce que signifie une unité de famille "de référence", il faut comprendre la nature hiérarchique d'un ménage de la BDSPS. Un ménage de la BDSPS contient un ensemble ordonné de familles économiques dont chacune contient un ensemble ordonné de familles de recensement, dont chacune contient un ensemble ordonné de familles nucléaires, dont chacune contient un ensemble ordonné d'individus. Les individus, dans une famille nucléaire, sont ordonnés en commençant par la personne la plus âgée, suivie du conjoint de la personne, le cas échéant, puis suivent les enfants dans l'ordre croissant d'âge. Les familles nucléaires sont ordonnées dans les familles de recensement en commençant par le couple marié ou le parent seul, le cas échéant, qui est suivi des enfants "âgés" dans l'ordre croissant de l'âge. Les familles de recensement sont ordonnées par les familles économiques en commençant par la famille de recensement "principale". Les familles économiques sont ordonnées de façon similaire à l'intérieur des ménages.

L'unité de famille ou d'individu "de référence" est par conséquent définie comme étant la première unité de ce genre au plus haut niveau d'analyse. Ainsi, la famille nucléaire de référence dans une famille économique est la première famille nucléaire de la famille économique, selon l'ordre donné à l'alinéa précédent.

Si la sélection a été activée par les paramètres de commande SELFLAG et SELSPEC, la

définition de la famille ou de la personne de référence a besoin d'être clarifiée un peu plus : l'unité de famille ou l'individu "de référence" devient alors la première unité sélectionnée au plus haut niveau d'analyse. Dans l'exemple hypothétique suivant, si les paramètres de commande de sélection

```
SELFLAG      1
SELUNIT      0
SELSPEC      idiemp > 5000
```

sont donnés, seuls les individus ayant plus de 5 000 \$ de revenu d'emploi seront sélectionnés. La demande de table

```
CF: idsex * {units};
```

fera alors le compte des familles de recensement, par catégorie de sexe de la première personne de chaque famille de recensement ayant un revenu d'emploi de plus de 5 000 \$. S'il n'y a pas de personne ayant ce revenu dans une famille de recensement, la famille de recensement n'est pas comptée dans la totalisation.

## TITRES DES TABLES

Étant donné que les tables sont d'une utilité très restreinte si leur contenu ne peut pas être facilement compris, la fonction Tableaux croisés compose un titre informatif pour chaque demande de table. Ces titres de tables sont distincts des en-têtes de page que le MSPS place sur tous ces produits de sortie (c.-à.-d. en plus de l'information sur la version du BD/MSPS, la date d'exécution, l'identification de la version de base et de variante utilisées pour l'exécution, les critères de sélection, le cas échéant, à utiliser pour l'exécution, le sous-échantillon utilisé ainsi que l'algorithme de vieillissement utilisé en particulier). Il est plus facile de comprendre le titre des tables ou de segments si l'on sait qu'ils sont formés d'un ou de deux chaînes de texte qui dépendent du contenu de la table. De gauche à droite, le contenu des lignes d'un titre est le suivant :

- a) La première ligne du titre d'une table ou d'un segment commence par l'expression "Table" suivie d'un nombre entier donnant la position cardinale de la demande de table dans le paramètre XTSPEC, suivi par la lettre "U" (pour indiquer qu'il s'agit d'une table définie par l'utilisateur), puis de deux points. Par conséquent, un titre dont la première partie est "Table 3U" indique que la table qui suit est la troisième table définie par l'utilisateur traitée au cours de l'exécution. Si la table se poursuit sur plusieurs segments parce qu'elle a trois dimensions ou plus, alors la chaîne "(suite.)" se trouve juste avant les deux points, pour tous les segments autres que le premier.
- b) La première ligne du titre se poursuit en indiquant l'aspect "central" de la table, c'est-à-dire ce qui est totalisé.

Si l'analyste utilise un niveau de totalisation qui comprend seulement une variable, alors l'étiquette de la variable se trouve à la suite (ou le nom de la variable, s'il n'y a pas d'étiquette).

Si l'analyste utilise une expression simple, alors le texte de l'expression se retrouve dans le titre. Bien sûr, si l'analyste a utilisé le qualificatif L dans le niveau de totalisation, alors l'étiquette donnée explicitement remplace ces valeurs par défaut automatiques. Dans ces

deux cas aussi, l'identification du contenu comprendra un suffixe approprié, par exemple “(000)”, si le facteur d'ajustement n'est pas zéro. S'il y a un qualificatif M, alors un suffixe indique la variable de classe sur laquelle est basée la normalisation.

Enfin, si l'analyste utilise un niveau de totalisation qui comprend des expressions ou variables multiples, le titre comprend l'expression “quantités sélectionnées”. Ce mécanisme empêche le titre de devenir trop compliqué en tentant d'être complet. Bien sûr, l'information essentielle à l'identification (étiquettes ou noms, ou facteurs d'ajustement, ou identité des variables “marginales”) se retrouve alors ailleurs dans les en-têtes de table ou de segment où elle informera mieux le lecteur.

- c) La première ligne du titre se poursuit avec le mot “pour” suivi de l'unité d'analyse de la table. Par exemple, si la demande de table commençait par “CF : ”, alors cette partie du titre serait formée de la chaîne de texte “pour les familles de recensement”. Cette partie du titre est présente que l'unité d'analyse ait pris ou non la valeur par défaut dans la demande de table.
- d) La première ligne du titre se termine par l'identification de toutes les variables de classe qui déterminent les catégories de colonnes, de rangs, de segments, etc., de la table. Les variables de classe sont données dans l'ordre descendant, par niveau dans la demande de table, le niveau le plus élevé étant précédé de l'expression “par” et les niveaux subséquents de la chaîne “et”. L'identification elle-même prend la forme des étiquettes des variables de classe (avec le nom de la variable s'il n'y a pas eu d'étiquettes définies). Par conséquent, le titre d'une table ayant la province comme variable de rang et le type de famille comme variable de colonne pourrait comprendre la chaîne de texte “par province et type de famille” comme partie finale de la première ligne.
- e) La première ligne du titre de la table, formée du texte décrit de a) à d) ci-dessus se retrouve à tous les segments de la table. Si la table a plus qu'un segment, la seconde ligne indique les choses qui sont “conservées constantes” à l'intérieur du segment; si la table compte plusieurs segments, il y a nécessairement au moins un élément constant. Donc, si une table utilise, par exemple, le sexe de l'individu comme variable de classe et qu'un segment de table en particulier ne compte que des hommes, alors la chaîne de texte “Sexe = Masculin ” se retrouvera sur la seconde ligne du titre de segment. Cette chaîne de texte est composée de deux éléments (la description de la variable et l'étiquette de classe) séparés par un signe égal (“ = ”). S'il y a plusieurs variables “de commande” de ce genre pour un segment, alors les variables de classe se retrouvent toutes sur la seconde ligne du titre, dans l'ordre descendant de niveaux, et les chaînes de texte individuelles qui forment la ligne sont séparées par des virgules.

## EXCEPTIONS ET DIVERS

Maintenant que l'on a vu en détail les variables de classe et de totalisation ainsi que l'étiquetage, nous pouvons indiquer comment il est possible d'annuler l'effort des variables “à fréquence croissante de cycle dans les valeurs” de gauche à droite s'il y a une seule variable de totalisation. Dans la description de fonctionnement présentée ci-dessus, le niveau le plus à droite était traité comme variant à la fréquence la plus grande (catégories de colonne), le niveau suivant, à gauche, comme variant à la deuxième plus grande fréquence (catégories de rang), etc. En fait, c'est



précisément ce qui se passe si l'analyste emploie un niveau de totalisation qui comprend plusieurs expressions ou variables.

Cependant, pour des fins de simplicité et d'aspect de la table, la fonction Tableaux croisés déroge à la règle générale dans le cas où la demande de totalisation utilise une seule expression ou variable de totalisation. Dans ce cas en particulier, la fonction Tableaux croisés crée la table (et son titre) comme si l'analyste avait entré le niveau de totalisation de la demande comme le plus élevé (le premier, ou celui qui est le plus à gauche). L'ordre des niveaux restants, les variables de classe, est laissé inchangé. Cette dérogation optimise le nombre de variables de classe qui se retrouvent dans la table ou le segment de table. Elle empêche aussi qu'il y ait des tables à segments multiples, dont chaque segment comprendrait une seule colonne ou un seul rang.

## Exemples pratiques

Maintenant que l'on a terminé l'étude des éléments de la demande de totalisation de la fonction Tableaux croisés, nous porterons notre attention sur les demandes de totalisation prises dans leur ensemble ainsi que sur les genres de tables que l'on obtient. La présente section donne une série d'exemples gradués. Numérotés X1 à X9, les exemples comprennent des demandes de table et les tables qui sont obtenues. Les sept premiers exemples utilisent l'échantillon de 5 % de la BDSPS de 1988, avec une spécification de sélection qui n'exclut aucun enregistrement. Les exemples restants utilisent toute la base de données (100 %), mais appliquent un critère de sélection qui sélectionne seulement les personnes employées à temps plein travaillant pendant une année complète.

### AGRÉGAT D'UNE SEULE VARIABLE

Le premier exemple illustre la demande la plus simple possible. L'utilisateur désire savoir ce que l'échantillon de 5 % du BD/MSPS utilise comme population du Canada en 1988. Le paramètre XTSPEC est donné comme suit :

```
XTSPEC
{persons};
```

c'est-à-dire que l'analyste désigne une seule variable de totalisation, n'utilise aucune variable de classe et accepte la valeur par défaut de l'unité d'analyse de la fonction Tableaux croisés (individus). La table suivante est produite :

Table X1: Person Count (000) for Individuals

```
+-----+-----+
|Person Count (000) | 25212.8|
+-----+-----+
```

La table contient un seul nombre, bien étiqueté, la fonction Tableaux croisés fournissant le facteur d'ajustement par défaut de 1 000.

### UNE DEMANDE DE TABLES MULTIPLES AVEC UNE VARIABLE DE CLASSE

Pour les deuxième et troisième exemples de table, l'analyste désire savoir combien il y a de revenus de pension de régimes de retraite privés (variable *idipens*) dans la BDSPS, répartis par situation de l'individu vis-à-vis de l'activité (variable *idlfst*). Ici, le paramètre XTSPEC comprend deux demandes de totalisation distinctes :

```
XTSPEC
IN: idlfst
  *{idipens};
IN: idlfst
  *{persons};
```

Les tables X2 et X3 donnent les résultats suivants.

Table X2: Pension income (M) for Individuals by Labour force status

Labour force status	Pension income (M)
N/A	0.0
Employed	1270.1
Unemployed	55.1
Not in LF	5169.8

Table X3: Person Count (000) for Individuals by Labour force status

Labour force status	Person Count (000)
N/A	5536.8
Employed	10996.6
Unemployed	1465.6
Not in LF	7213.8

Comme on pourrait s'y attendre, le gros des revenus de pension de régimes privés, quelque 5 milliards de dollars, est associé à des personnes qui ne font pas partie de la population active, probablement parce qu'elles sont retraitées. Certains revenus de pension se retrouvent aussi chez des personnes qui ont encore un emploi, ou qui n'ont pas d'emploi, mais le montant moyen de ces autres prestataires sont de beaucoup inférieurs à ceux qui reviennent à la classe "Not in LF".

### DEMANDE DE VARIABLES DE TOTALISATION MULTIPLES DANS UNE SEULE TABLE

Bien que l'analyste désire souvent inclure des demandes de totalisation multiples dans un même XTSPEC, avec des variables de totalisation différentes se retrouvant dans des tables différentes, il peut aussi vouloir que les variables de totalisation multiples se retrouvent dans une même table. Dans cet exemple, l'analyste désire examiner, par province, les montants globaux de revenus selon quatre définitions de revenu différentes. Il désire aussi connaître les montants globaux de ces revenus pour tout le Canada. Dans cet exemple, la demande de totalisation se lit comme suit :

```
XTSPEC CF:hdprov+
      {immtot,
        imitax,
        immdisp,
        immicons};
```

Ici l'analyste a choisi de spécifier la famille de recensement comme unité d'analyse; il faut noter, cependant, que, comme la spécification de sélection comprend tous les enregistrements et que les

variables totalisées sont des mesures du revenu, l'unité d'analyse ne sera pas pertinente dans la table produite. Voici la table obtenue --

Table X4: Selected Quantities for Census Families by Province

Province	Total income (M)	Taxable income (M)	Disposable income (M)	Consumable income (M)
Newfoundland	4146.9	2067.0	3544.5	3150.7
P.E.I.	940.7	466.5	803.8	719.6
Nova Scotia	8440.7	4653.8	7025.8	6347.9
New Brunswick	6668.6	3491.6	5541.4	4951.5
Quebec	79654.1	44294.7	62291.6	56486.1
Ontario	117893.7	70393.8	96575.5	87004.8
Manitoba	11645.7	6561.4	9586.1	8605.6
Saskatchewan	11387.1	6063.1	9591.2	8740.6
Alberta	31204.8	19142.6	25509.8	24027.2
B.C.	38351.3	22733.6	31243.9	28608.8
All	310333.5	179868.0	251713.7	228642.8

Le revenu total (colonne 1) est, bien sûr, beaucoup plus grand que la partie qui est imposable aux fins de l'impôt sur le revenu (colonne 2) à cause des déductions et exemptions qui sont disponibles. Le revenu disponible (colonne 3) tombe alors entre ces deux premières définitions parce que l'impôt sur le revenu a été soustrait. Enfin, le revenu consommable (colonne 4) est quelque peu inférieur au revenu disponible parce que les diverses taxes à la consommation ont été soustraites aussi.

## PRÉSENTATION DES QUALIFICATIFS ET DES EXPRESSIONS TOTALISÉS

Dans cet exemple, l'analyste désire aussi voir des variables multiples dans la même table, plus précisément le revenu disponible, les familles de recensement et le revenu disponible moyen par famille de recensement. Cependant, ce qui distingue cette demande de la demande précédente, la dernière variable est disponible ni de la base de données, ni du modèle. Par conséquent, l'analyste entre une expression, plutôt qu'une variable, comme troisième élément à totaliser. Cet exemple est aussi le premier à utiliser les qualificatifs pour les variables dépendantes. Ici, l'analyste a choisi de faire ressortir le revenu disponible, en milliards de dollars (S=9) et de fournir une étiquette ("Avg Disp Inc") pour le revenu disponible moyen par unité. L'analyste fixe aussi le niveau de totalisation avant le niveau de classe de façon que les variables totalisées serviront d'en-tête de rang tandis que les catégories de province se retrouvent comme en-tête de colonne. Le paramètre XTSPEC devient alors --

```
XTSPEC
CF:{immdisp:S=9,
    units,
    immdisp/units:L="Avg Disp Inc"}
*hdprov+;
```

L'analyste utilise *units* comme variable de totalisation et, comme élément de l'expression de totalisation, il désigne CF, ou la famille de recensement, unité d'analyse à utiliser pour la table. La table suivante est produite --

Table X5: Selected Quantities for Census Families by Province

Quantity	NFLD	PEI	NS	NB	QUE	ONT	MAN	SASK	ALTA	BC	ALL
Disp.	3.5	0.8	7.0	5.5	62.3	96.6	9.6	9.6	25.5	31.2	251.7
Income (B)											
Unit Count	176.2	39.4	318.2	256.9	2599.6	3612.5	399.5	412.6	1016.9	1221.6	10053.4
(000)											
Avg Disp	20114	20420	22082	21568	23962	26734	23996	23247	25085	25577	25038
Inc											

Comme on s’y attendait, le revenu disponible moyen de la famille de recensement est plutôt bas dans les Maritimes tandis que l’Ontario et l’Alberta ont les salaires moyens les plus élevés.

## DEUX NIVEAUX DE CLASSE

Jusqu’à maintenant, les exemples n’illustraient pas la vraie fonction de tableaux croisés parce qu’ils avaient au plus une variable de classe. Dans l’exemple qui suit, l’analyste désire examiner les catégories de professions (variable *idocc*) déterminant le sexe des individus (variable *idsex*). Dans la demande de totalisation donnée ici, deux autres choses sont nouvelles.

En premier lieu, la demande utilise le ménage (HH) comme unité d’analyse et, en second lieu, l’analyste désire compter les enregistrements EDTR (non pondérés) de la base de données plutôt que de donner le total des personnes ou des montants, en dollars.

La demande de cette table est formulée comme suit :

```
XTSPEC
HH: idocc+
  * idsex
  * {scfrecs};
```

Il faut se rappeler que, s’il y a seulement une variable de totalisation au niveau de totalisation, la fonction Tableaux croisés traite la demande comme si le niveau de totalisation était le premier niveau de la demande. Voici la table obtenue :

Table X6: SLID Records for Households by Occupation and Sex

Sex		
Occupation	Male	Female
Never worked	3	71
Managerial	143	33
Professional	104	47
Teaching	23	15
Clerical	53	58
Sales	103	15
Services	124	83
Agricultural	117	4
Mining, Processing	104	3
Fabrication, Assembly	105	7
Construction	113	0
Transport, Handling	106	5
Last worked > 5 years	214	181
All	1312	520

Les parties de la table sont exactement comme on aurait pu s’y attendre après avoir consulté les

analyses historiques dans les choix d'emploi. Les femmes sont relativement concentrées dans des classes des domaines de l'enseignement, du travail de bureau, des services, des gens qui n'ont jamais travaillé et des gens qui ont travaillé, la dernière fois, il y a plus de cinq ans. Les hommes sont surreprésentés dans les professions comme la direction, les mines, la fabrication, la construction et le transport.

La fonction de cumul de la fonction Tableaux croisés explique la grosseur relative des deux colonnes, c'est-à-dire la raison pour laquelle il y a plus de deux fois plus d'enregistrements pour des hommes que pour des femmes. Parce que l'analyste a demandé une totalisation pour les ménages, même si la profession est habituellement considérée comme étant une caractéristique individuelle, le MSPS a fait le cumul des caractéristiques de profession déclarées pour donner une seule profession pour chaque ménage, spécifiquement, la profession de l'homme (pour les couples) dans la principale famille économique du ménage. Par conséquent, les hommes, qui tendent à avoir des revenus plus élevés, ont une place disproportionnée dans les comptes d'enregistrements qui constituent les entrées de la table.

Une seconde caractéristique à noter est la mesure dans laquelle, même s'il y a seulement deux variables de classe, les nombres d'enregistrements qui entrent dans certaines cellules de la table sont très bas ou même nuls. L'analyste doit régulièrement vérifier que les entrées de la table sont basées sur des enregistrements suffisamment nombreux pour permettre des conclusions significatives. La capacité du MSPS de totaliser des nombres d'enregistrements est un outil très utile dans la poursuite de ce but.

## **UNE VARIABLE DE CLASSE DÉFINIE PAR L'UTILISATEUR ET UNE DEMANDE SUR LIGNES MULTIPLES**

Dans cet exemple, l'analyste désire savoir dans quelle mesure les dépenses liées au logement varient comme proportion du revenu pour diverses combinaisons de niveaux de revenus et de logements. Encore une fois, la variable de totalisation désirée n'est pas disponible directement de la base de données ou du modèle et, par conséquent, l'analyste doit la composer. La composition, dans ce cas, exige l'ajout d'un certain nombre de dépenses liées au logement tirées de la partie de la base de données dont la source est l'EDM et la division de cette somme par la mesure du revenu total des familles. Une autre chose nouvelle est l'utilisation d'une variable de classe définie par l'utilisateur, `cl1`, pour fournir les catégories de revenu. Pour cette fin, l'analyste a inclus, ailleurs dans la demande UVAR, les instructions suivantes :

```
classinc = split(immtot, 10000,20000,30000,40000,50000,60000,70000);  
levels(classinc) = "MIN-10","10-20","20-30","30-40","40-50","50-60","60-  
70","70-MAX";  
label(classinc) = "Total income Group";
```

La demande répartie sur plusieurs lignes crée huit catégories de revenus basées sur `immtot` définie dans la définition de `levels()`. La définition de `levels()` désigne `classinc` variable de classe avec une longueur maximale de 60 caractères. Les étiquettes définies par l'utilisateur pour les classes correspondent aux limites suivantes : Minimum à 10000, 10001 à 20000, 20001 à 30000, 30001 à 40000, 40001 à 50000, 50001 à 60000, 60001 à 70000, 70001 à maximum. S'il n'y avait pas d'étiquette attribuée, l'utilisateur devrait utiliser `nlevels()` pour attribuer le bon nombre de classifications à la variable `classinc` et la sortie identifierait les classes comme 0 à 7.

L'étiquette "Total Income Group" est utilisée pour donner un type à la variable dans la table produite.

Il faut noter que la variable *immtot* influe sur la table de deux façons distinctes : 1) la classification des variables par groupe de revenus et 2) la définition de l'expression de totalisation. La demande de totalisation devient alors ce qui suit :

```
XTSPEC
HH: {(fxio7+fxio8+fxio9+fxio10+fxio11+fxio12)/immtot:
      L="Housing Exp Prop" S=% P=3 }
    * classinc+
    * hdtenu+;
```

Dans la demande, l'analyste désire aussi fournir une étiquette pour la variable totalisée, ainsi que déterminer tant le facteur d'ajustement que le nombre de décimales qu'il y aura dans la table. L'analyste peut toujours utiliser autant de lignes qu'il en faut pour que la demande se lise facilement. La table suivante est produite :

Table X7: Housing Exp Prop (%) for Households by Total Income Group and Tenure

Total income Group	Tenure			All
	Rented	Owned with Mtge.	Owned w/o Mtge.	
Min-10	37.025	8.951	19.065	30.457
10-20	25.711	6.938	10.656	18.352
20-30	20.248	5.995	6.523	11.904
30-40	17.661	5.199	4.803	9.186
40-50	13.273	4.337	4.284	6.072
50-60	14.459	4.128	2.722	4.500
60-70	9.631	4.070	2.631	6.541
70-Max	9.387	3.653	2.041	4.219
All	17.829	4.498	4.583	8.494

Le choix qu'a fait l'analyste quant au facteur d'ajustement signifie que les entrées de la table correspondent à des pourcentages du revenu total. Comme on s'y attendait, la proportion du revenu total dépensé pour cette combinaison particulière de dépenses de logement diminue constamment à mesure que l'on franchit les catégories de revenu. Comme on s'y attendait aussi, les propriétaires dépensent une fraction de leur revenu moins grande pour les catégories de dépenses de logement que ne le font les locataires. La table laisse aussi entrevoir une inversion dans la proportion propriétaire avec ou sans hypothèque à mesure que l'on monte dans la répartition du revenu.

## PRÉSENTATION DE LA NORMALISATION ET D'UNE TROISIÈME DIMENSION

Dans les deux derniers exemples, l'analyste désire comparer les gains relatifs des hommes et des femmes, en les normalisant à un montant moyen de gains annuels pour les personnes travaillant au moins 50 semaines par année (*idlyww* ≥ 50) et à temps plein plutôt qu'à temps partiel (*idlyfp* = 1), et, ainsi, il commence en établissant le paramètre de spécification de sélection pour l'analyse :

```
SELFLAG 1
```

```
SELUNIT      0
SELSPEC      idlyww>=50 && idlyfp==1
```

L'analyste se demande aussi si la situation matrimoniale (variable *idmarst*) et l'âge (variable *idage*) semblent contribuer à la tendance observée et, par conséquent, les inclut comme variables de classe. En outre, l'analyste veut à tout prix éviter de tirer des conclusions erronées de l'analyse et, par conséquent, il décide d'utiliser la BDSPE au complet (échantillon 100 %) plutôt que l'échantillon de 5 %. Dans le même sens, l'analyste décide, quelque peu arbitrairement, que, dans le calcul des valeurs normalisées, il ne serait pas conseillé d'utiliser toute valeur des gains annuels moyens qui ne serait pas basée sur au moins une centaine de cas de la base de données.

Parce que l'analyste désire utiliser un nombre relativement petit de catégories d'âge, il faudrait utiliser la fonction UVAR du MSPS. Voici les lignes à utiliser --

```
agegrp = split(idage, 35,45,55,65);
levels(agegrp) = "MIN-35", "36-45", "46-55", "56-65", "66-MAX";
label(agegrp) = "Age Group";
```

Donc, l'analyste désire générer deux tables. La première de ces tables donnera les gains annuels normalisés par personne, tel qu'on le désire, tandis que la seconde donnera les nombres d'enregistrements de base de données sur lesquels les valeurs normalisées sont basées (de façon que des valeurs qui pourraient ne pas être fiables puissent être supprimées, ou du moins pondérées). Le paramètre XTSPEC de ces deux tables se lit comme suit :

```
XTSPEC
IN: {(idiemp+idisefm+idisenf)/persons:
      L="Per Capita Normalized" M=idsex S=% P=2}
      * idsex+
      * agegrp+
      * idmarst+;
IN: {scfrecs: L="Instances (records)" P=0 S=0}
      * idsex+
      * agegrp+
      * idmarst+;
```

Dans la première demande de totalisation, l'analyste additionne le revenu des salariés ainsi que les revenus de travailleurs autonomes tant agricoles que non agricoles et divise la somme par persons pour en arriver à des gains moyens par personne. Les qualificatifs fournissent une étiquette "Per Capita Normalized" pour la quantité obtenue, impose la normalisation dans la dimension sexe ( $M=idsex$ ), et fixe le facteur d'ajustement ( $S=\%$ ) ainsi que le nombre de décimales ( $P=2$ ). L'ordre des trois dimensions de classe fait en sorte que le sexe varie seulement d'un segment à l'autre, tandis que les segments de la table des individus donnent l'âge comme catégorie de rang et la situation matrimoniale comme catégorie de colonne. Voici donc la table qui intéresse le plus l'analyste.

La seconde demande de table est très semblable, sauf que l'analyste compte les enregistrements de l'EDTR pour les individus. Si, tant pour les hommes que pour les femmes, un segment de table contient un nombre d'enregistrements inférieur à cent, alors les valeurs normalisées pour les emplacements correspondants des segments de la première demande seront considérés comme pouvant être suspects. Les nombreux segments de table que la fonction Tableaux croisés produit pour combler ces deux demandes se retrouvent à la page suivante.

Table X8: Per Capita Normalized (%) [idsex] for Individuals by Sex and Age

Group and Marital status  
Sex = Male

Marital status					
Age Group	Married or CLU	Single (never married)	Widow(er)	Divorced	All
Min-35	115.07	106.14	119.44	115.26	113.28
36-45	115.32	108.27	122.58	120.79	115.96
46-55	114.34	104.52	139.26	113.91	115.34
56-65	110.25	101.32	137.48	122.35	112.86
66-Max	101.78	103.64	85.02	97.71	105.97
All	114.50	105.57	132.32	117.32	114.99

Table X8 (cont.): Per Capita Normalized (%) [idsex] for Individuals by Sex and Age Group and Marital status

Sex = Female

Marital status					
Age Group	Married or CLU	Single (never married)	Widow(er)	Divorced	All
Min-35	75.33	92.14	76.13	88.86	80.76
36-45	66.11	91.23	94.10	83.95	69.92
46-55	62.57	96.07	83.05	87.28	67.64
56-65	63.84	99.23	81.20	85.79	70.09
66-Max	82.78	97.86	106.39	104.93	77.29
All	68.62	93.35	84.62	86.50	73.75

Table X9: Instances (records) for Individuals by Sex and Age Group and Marital status

Sex = Male

Marital status					
Age Group	Married or CLU	Single (never married)	Widow(er)	Divorced	All
Min-35	5056	2158	16	201	7431
36-45	4166	251	14	222	4653
46-55	2867	135	41	148	3191
56-65	1769	96	56	63	1984
66-Max	249	22	12	8	291
All	14107	2662	139	642	17550

Table X9 (cont.): Instances (records) for Individuals by Sex and Age Group and Marital status

Sex = Female

Marital status



Age Group	Married or CLU	Single (never married)	Widow(er)	Divorced	All
Min-35	2878	1629	20	285	4812
36-45	1868	237	55	286	2446
46-55	1135	192	117	175	1619
56-65	505	167	140	98	910
66-Max	26	58	26	6	116
All	6412	2283	358	850	9903

Si l'analyste prend au sérieux l'exigence de cent cas, alors presque toutes les entrées de la troisième colonne ainsi que du quatrième et du cinquième rang sont suspects (toutes ces cellules à l'exception de la première colonne et du premier rang). Si l'analyste choisit de supprimer ces cellules (la fonction Tableaux croisés ne fait pas ce travail elle-même), alors les tables de la première demande auraient l'air de ce qui suit :

Table X8a: Per Capita Normalized (%) [idsex] for Individuals by Sex and Age Group and Marital status

Sex = Male

Marital status					
Age Group	Married or CLU	Single (never married)	Widow(er)	Divorced	All
Min-35	115.07	106.14		115.26	113.28
36-45	115.32	108.27		120.79	115.96
46-55	114.34	104.52		113.91	115.34
56-65	110.25				112.86
66-Max					105.97
All	114.50	105.57	132.32	117.32	114.99

Table X8a (cont.): Per Capita Normalized (%) [idsex] for Individuals by Sex and Age Group and Marital status

Sex = Female

Marital status					
Age Group	Married or CLU	Single (never married)	Widow(er)	Divorced	All
Min-35	75.33	92.14		88.86	80.76
36-45	66.11	91.23		83.95	69.92
46-55	62.57	96.07		87.28	67.64
56-65	63.84				70.09
66-Max					77.29
All	68.62	93.35	84.62	86.50	73.75

En général, les hommes semblent avoir des gains annuels plus élevés que la moyenne pour toutes

les personnes. Par exemple, les hommes mariés ont des gains annuels en gros 14 % plus élevés que la moyenne de toutes les personnes mariées et, étant donné les nombres relatifs d'hommes et de femmes occupant un poste complet, les femmes ont des revenus annuels moyens qui sont seulement 68 % environ de cette même valeur "de toutes les personnes mariées". La situation matrimoniale semble avoir une certaine valeur pour expliquer le niveau de disparité dans l'égalité des revenus (étant donné que l'analyste n'a pas pu contrôler les autres variables comme la profession et le mélange d'âge de la population active), mais l'âge ne semble pas faire beaucoup de différence.

## Trucs pour les tableaux croisés

Les sections ci-dessus portaient principalement sur les aspects plus techniques du mode de fonctionnement de la fonction Tableaux croisés. Sur cette base, nous verrons plusieurs points de manière dont l'analyste utilise la fonction Tableaux croisés pour exécuter des analyses typiques; les paragraphes qui suivent seront considérablement moins techniques, portant plus sur les méthodes.

### **LIMITES IMPOSÉES DE L'EXTÉRIEUR SUR LA FONCTION TABLEAUX CROISÉS**

Comme on l'a vu ci-dessus, le but visé de la fonction Tableaux croisés était la création d'une fonction qui comblerait les besoins de totalisation habituels dans le BD/MSPS. Par conséquent, la fonction Tableaux croisés permet à l'analyste de produire, de façon rapide et commode, diverses tables adaptées sans qu'il n'ait à exporter les données pertinentes, puis à importer les données exportées dans un autre progiciel qu'il utilisera pour traiter encore ces données. Cependant, étant donné l'exigence qu'elle doive fonctionner à l'intérieur du BD/MSPS sans imposer de grands délais à l'exécution, la fonction Tableaux croisés ne peut pas être aussi complète qu'un progiciel autonome utilisé strictement pour les analyses statistiques. (Par exemple, la fonction Tableaux croisés n'a pas la capacité d'imbriquer les variables de classe à l'intérieur les unes des autres dans une dimension d'une table.)

Certaines des limites sont imposées principalement du fait des machines sur lesquelles le MSPS a été conçu pour fonctionner ainsi que de leur système d'exploitation. Par exemple, l'architecture des ordinateurs signifie que le MSPS (qui comprend la fonction Tableaux croisés et tout algorithme "boîte de verre" fourni par l'analyste) ainsi que toutes les valeurs intermédiaires pour les tables et les variables requises par l'analyste doivent, ensemble, entrer dans une certaine quantité de mémoire. Bien que le MSPS et la fonction Tableaux croisés aient été conçus de façon à économiser cette mémoire autant que cela est possible, il se peut que l'analyste atteigne les limites en demandant trop de variables et de tables. D'autres limites "mécaniques", par exemple un maximum de 16 000 cellules dans une demande de totalisation, sont imposées du point de vue technique, mais ne poseront pas de contraintes particulières dans les analyses habituelles.

Parce que les limites peuvent être différentes d'un ordinateur à l'autre, par exemple, la version du système d'exploitation utilisé ou la présence de logiciels résidant en mémoire vive, il est impossible de fournir des valeurs limites précises ou même des algorithmes simples pour calculer ces valeurs. En général, l'analyste devrait savoir que ces genres de limites existent et que, en outre, les besoins de mémoire inhérents aux demandes de tableaux croisés peuvent les déclencher. S'il a conscience de ces limites, l'analyste pourra plus facilement les reconnaître

lorsqu'elles surviendront et, au besoin, il pourra trouver d'autres méthodes pour atteindre les mêmes buts d'analyse.

## **LE CHOIX D'UNE EXÉCUTION DU MSPS DE LA TAILLE APPROPRIÉE**

Une des principales caractéristiques du MSPS est qu'il tourne sur un micro-ordinateur. Pour l'analyste qui n'a pas accès à un ordinateur central et à des progiciels qui étaient auparavant nécessaires pour des analyses de microsimulation significative, cela signifie que ces analyses sont maintenant du domaine du possible; il n'est plus nécessaire de les faire exécuter par un sous-traitant. Pour l'analyste qui avait auparavant recours aux calculs sur ordinateur central, le MSPS sabre maintenant tant dans les coûts d'exploitation d'ordinateur que dans le temps d'exécution. Autre chose, moins évidente, la disponibilité du MSPS et de la fonction Tableaux croisés peut aussi changer la manière dont l'analyste s'y prend pour procéder aux analyses.

Le gros avantage est que l'analyste n'a plus à affronter des coûts marginaux associés aux analyses de simulation. Il peut explorer des orientations suggérées par des exécutions antérieures. Il peut approfondir les recherches sur les causes de résultats curieux ou inattendus. Il peut explorer d'autres perspectives de divers phénomènes. Par dessus tout, il peut le faire dans la même limite de temps et de ressources financières qui étaient auparavant associées à des analyses moins poussées.

L'analyste qui avait l'habitude d'exécuter ses propres simulations sur un ordinateur central sera probablement celui qui sera le plus touché. Un environnement d'ordinateur central a un certain nombre d'avantages liés à l'exécution d'analyses plus volumineuses et plus approfondies. Par exemple, il peut y avoir des problèmes de délai associés à la présentation des exécutions, à l'obtention du matériel comme les dérouleurs de bande ou à l'impression des résultats; l'analyste réagissait en incluant dans les exécutions plus de résultats qui "auraient pu" être nécessaires et qui auraient interrompu le travail s'ils n'avaient pas été disponibles. Encore là, il peut y avoir des économies de temps et d'argent associées à la production de tables dans un nombre minimal d'exécutions pour couper le nombre de passages dans la base de données pertinente. Étant donné que le MSPS et la fonction Tableaux croisés sont disponibles, les raisons qui prévalaient pour ces techniques de travail ne valent plus. En outre, les limites imposées par les micro-ordinateurs sur le MSPS peuvent signifier que certains styles de traitement "tout-à-l'évier" qui prévalaient autrefois ne sont plus possibles. Nous nous attendons à ce que les analystes s'accoutument graduellement à un style d'analyse plus interactif. Ils trouveront plus efficace de faire plus d'exécutions avec moins de sorties par exécution. Ils devraient être de plus en plus capables d'explorer de nouvelles avenues d'analyse en se basant sur les résultats des exécutions précédentes.

## **EXPORTATION DE DONNÉES VERS D'AUTRES PROGICIELS**

Si utile et si puissante que la fonction Tableaux croisés puisse être, elle n'est pas, comme on l'a déjà dit, conçue pour concurrencer des progiciels d'analyse statistique dotés de toutes les caractéristiques nécessaires. Par conséquent, l'analyste ne devrait pas se sentir obligé d'exécuter toutes ces totalisations avec la fonction Tableaux croisés. Une approche pragmatique est nécessaire. Les choses qui peuvent être faites de façon plus efficaces dans le MSPS, avec la fonction Tableaux croisés, devraient s'y retrouver. Cependant, si quelque chose est déraisonnablement difficile à faire avec la fonction Tableaux croisés et le MSPS, alors l'analyste

devrait songer sérieusement à le faire ailleurs. Les versions plus anciennes du présent chapitre ont déjà étudié la capacité du MSPS d'exporter des données en ASCII et en SAS, tirant tous les avantages de la puissante capacité de cumul du MSPS dans la production des données exportées. Il est aussi possible de créer de nouvelles tables à partir de tables du MSPS en utilisant la fonction `import` de concert avec les progiciels de chiffriers comme Lotus 1-2-3 ou Symphony. Pour obtenir plus d'information sur ce sujet, consulter le document *Guide d'utilisation des outils*.

## **UTILISATION DES FICHIERS D'INCLUSION DE PARAMÈTRES**

Il faut se rappeler que le MSPS fournit une fonction de d'édition de paramètres de façon qu'un analyste peut lire un fichier de paramètres, ce qui comprend les fichiers de paramètres de commande, puis modifier certaines valeurs de paramètre. En outre, l'analyste peut utiliser la fonction `read` du MSPS pour remplacer des jeux complets de paramètres en tirant les valeurs de remplacement de "pièces" stockées dans des fichiers sur disque.

Parce que XTSPEC est (avec les paramètres connexes comme XTFLAG, XTCOLS et XTLINES) un paramètre dans le fichier de paramètres de commande, la fonction `read` s'applique aussi à lui. Par conséquent, un analyste qui a régulièrement besoin d'un ensemble de certaines tables personnalisées qui sont communes d'une exécution à l'autre du MSPS peut stocker le paramètre XTSPEC pour générer ces tables dans un fichier sur disque. Par la suite, lorsqu'il a besoin de cet ensemble de tables, l'analyste peut, de façon rapide et commode, faire la demande en vue de les utiliser grâce à la fonction `read` du MSPS (de concert avec tout autre paramètre de commande pertinent).

## **PRODUCTION DE DEMANDES DE TOTALISATION EFFICACES ET ÉLÉGANTES**

Le MSPS fournit "des valeurs par défaut intelligentes" pour un certain nombre d'articles qui ont une influence sur la lisibilité des tables (titres, étiquettes de variables et de classe, segments, etc.). Cependant, la description que fait l'analyste des tables désirées a aussi d'importantes répercussions sur l'aspect, et le contenant, de ces tables. Les éléments suivants indiquent certains de ces choix les plus importants :

- a) Profiter du qualificatif `L` pour les variables de totalisation, particulièrement lorsque l'on totalise une expression. Il faut profiter des qualificatifs `S` et `P` pour appliquer les ajustements appropriés et déterminer le nombre de décimales dans les entrées des tables.
- b) Tenir compte des limites de l'appareil de sortie (p. ex. imprimante) quand on choisit l'identité et l'ordre des variables de classe d'une table. Par exemple, il faut éviter de produire une table plus large que ce que peut produire son imprimante.
- c) Tenir compte de la forme que prendront les tables demandées et de l'utilité de cette forme pour les fins prévues. Par exemple, trois segments de table de six colonnes et 12 rangs chacun, tous sur une même page, peuvent être de beaucoup plus utiles que 12 segments de table de trois colonnes et de six rangs, chaque segment se retrouvant sur une page distincte. Utiliser les paramètres XTCOLS et XTLINES et un ordre réfléchi des niveaux pour produire une table ayant un meilleur aspect.
- d) Se rappeler que la fonction Tableaux croisés permet à l'analyste d'inclure des variables de totalisation multiples dans une même table ou un même segment de table et de les placer dans

les mêmes dimensions de rang ou de colonne. Parfois, cela peut être beaucoup plus efficace que d'avoir des variables de totalisation multiples réparties entre plusieurs tables ou plusieurs segments de table.

- e) La normalisation comme technique d'analyse : l'application des résultats de simulations dans les politiques exige souvent que l'on étudie des résultats comparatifs, par exemple, les taux de pauvreté relatifs. La capacité du MSPS d'exécuter diverses normalisations (avec le qualificatif M pour les expressions et les variables totalisées) peut être d'un grand secours pour l'analyste. La capacité de normalisation de la fonction Tableaux croisés trouve régulièrement son utilité dans la production des pourcentages de rang et de colonne désirés.

## Résumé

La présente section résume les points-clés entourant la fonction Tableaux croisés.

- a) La fonction Tableaux croisés fournit un mécanisme commode à l'analyste pour la production de tableaux croisés adaptés sans qu'il n'ait à payer le prix de l'exportation des résultats intermédiaires en vue d'un traitement dans un autre progiciel.
- b) XTSPEC, un élément du fichier de paramètres de commande, est un outil majeur qui permet à l'analyste de spécifier les tables désirées. XTSPEC est principalement un ensemble de demandes de totalisation où les demandes individuelles sont séparées par des points-virgules.
- c) Une demande de totalisation comporte plusieurs niveaux; à l'exception des tables contenant une seule variable (ou expression) totalisée, chacun de ces niveaux correspond à une dimension de la table produite. Le niveau d'une demande de totalisation sont séparés par des astérisques. De droite à gauche, les niveaux successifs correspondent d'abord aux colonnes, puis aux rangs, puis aux segments, etc.
- d) Un niveau indique exactement ce qui est totalisé, qu'il s'agisse de variables ou d'expressions. Ce niveau est appelé le niveau de totalisation et il se distingue du fait qu'il est inséré entre accolades (les caractères { et }).
- e) Un niveau de totalisation peut contenir des expressions ou des variables totalisées multiples. Lorsque des éléments multiples sont présents, ils sont séparés par des virgules.
- f) Le niveau de totalisation peut aussi comprendre divers qualificatifs qui déterminent les étiquettes, l'ajustement, les décimales dans les résultats imprimés ainsi que la normalisation. La portée du qualificatif est limitée à la variable de totalisation qu'il suit. La présence de qualificatifs pour une expression/variable de totalisation donnée est indiquée par le caractère deux-points, des qualificatifs multiples étant séparés par des espaces.
- g) Les niveaux autres que les niveaux de totalisation sont des niveaux de classe. Une demande de table peut avoir de zéro à six niveaux de classe. Chaque niveau de classe comprend une variable de classe unique.
- h) Dans la formulation d'une totalisation, l'analyste peut employer tant des variables de base de données que des variables de modèle et il peut aussi totaliser des expressions d'analyse.

- i) La capacité du MSPS et de la fonction Tableaux croisés de traiter des niveaux d'unités d'analyse multiples et de permettre à l'analyste de mélanger ces niveaux à l'intérieur d'analyses et à l'intérieur de tables uniques, comporte certains risques et implique certaines responsabilités. Bien que la conception de la fonction Tableaux croisés fasse en sorte que la plupart des cas soient traités naturellement, l'analyste doit connaître les problèmes de cumul qui influent sur les variables de totalisation et de classe et d'en comprendre les répercussions.
- j) Le MSPS fournit, à part la fonction Tableaux croisés, diverses capacités qui permettent de choisir des sous-ensembles de l'ensemble de la base de données. Il incombe à l'analyste, quand il interprète la table, de connaître les critères de sélection qui ont servi dans ces tables.

## Annexe A Principaux messages d'erreur de la fonction Tableaux croisés

La présente annexe donne les principaux messages d'erreur du MSPS associés à la fonction Tableaux croisés. Ce résumé complète la documentation de la fonction Tableaux croisés destinée à l'analyste utilisant le MSPS.

Quand cela est possible, la fonction Tableaux croisés comprend, dans son message d'erreur, l'identification de l'information qui cause l'erreur. Dans les messages d'erreur et leur interprétation qui sont donnés ci-dessous, les chaînes en italique représentent les chaînes de texte critiques extraites de la demande de l'utilisateur. En outre, la notation "XTSPEC # *n*" devrait être interprétée comme la "nième" demande de table faite dans XTSPEC; par conséquent, "XTSPEC # 2", par exemple, a trait à la deuxième table de demande qu'il y a à l'intérieur de XTSPEC.

`error(100): out of memory`

La fonction Tableaux croisés manque de mémoire (mémoire vive) pour stocker les résultats de sortie demandés. (Il s'agit d'un message qui s'affiche en cours d'exécution, par rapport à la plupart des autres messages de la fonction Tableaux croisés, qui s'affichent pendant que le MSPS est en train d'interpréter le fichier de paramètres de commande.) L'analyste doit répartir la demande en éléments plus petits qui peuvent être traités dans des exécutions multiples du MSPS.

`error(957): error in expression expr`

La fonction Tableaux croisés est incapable de traiter l'expression *expr*, probablement du fait d'une erreur de syntaxe. L'analyste doit reformuler l'expression de façon qu'elle soit compréhensible.

`fatal error(959): too many classification variables in XTSPEC # n`

La fonction Tableaux croisés dispose d'un volume de stockage interne limité pour le traitement des variables de classe et la demande de table indiquée a dépassé cette limite. L'analyste doit réduire le nombre de variables de classe utilisées, ce qu'il pourrait faire en répartissant les demandes de tables dans des exécutions multiples.

`fatal error(960): too many expressions in XTSPEC # n`

La fonction Tableaux croisés dispose d'un volume de stockage interne limité pour le traitement des expressions définies par l'utilisateur et la demande de table indiquée a dépassé cette limite. L'analyste doit réduire le nombre d'expressions utilisées, ce qu'il pourrait faire en répartissant les demandes de table dans des exécutions multiples.

`fatal error 961): too many analysis variables in XTSPEC # n`

La fonction Tableaux croisés a un volume de stockage interne limité pour le stockage de l'information associé aux variables d'analyse (totalisation) et la demande de table indiquée a dépassé cette limite. L'analyste doit réduire le nombre de variables d'analyse utilisées, ce qu'il pourrait faire en répartissant les demandes de table dans des exécutions multiples, et (ou) en analysant des variables totalisées moins nombreuses à l'intérieur d'une demande de table.

`fatal error (962): var not a valid analysis variable in XTSPEC # n`

La fonction Tableaux croisés interprète la chaîne (indiquée ici comme VAR) comme une expression ou une variable de totalisation voulue, mais celle-ci ne respecte pas les exigences de quelque chose qui peut être totalisé. Il se peut qu'une variable entière ait été utilisée comme une variable ou une expression de totalisation, même si le MSPS exige que ces variables ou expressions touchent seulement des variables d'analyse. L'analyste devrait consulter la documentation pour connaître ce que la fonction Tableaux croisés peut totaliser.

`fatal error (963): var not a valid class variable in XTSPEC # n`

La fonction Tableaux croisés permet l'utilisation de seulement certains genres de variables comme variables de classe dans les demandes de table. L'analyste doit veiller à ce que la variable désirée soit définie comme variable entière dans la base de données ou le modèle, ou à ce qu'il s'agisse d'une variable "CL" dont les étiquettes ont été correctement définies (p. ex., comme égal en nombre aux catégories permises pour la variable).

`fatal error (964): no analysis expression in XTSPEC # n`

La fonction Tableaux croisés exige que chaque demande de table contienne exactement un niveau indiquant les expressions ou variables de totalisation. Ce niveau est identifié par le fait qu'il se retrouve entre accolades ( { et } ). La fonction Tableaux croisés a trouvé qu'il n'y a pas dans la demande de table indiquée un niveau comme celui-là. L'analyste doit éliminer la demande de table ou la modifier de façon à inclure le niveau de totalisation approprié. Si l'analyste désire simplement un compte dans chaque cellule, il doit utiliser la variable units.

`fatal error (965): bad margin dimension in XTSPEC # n`

La fonction Tableaux croisés est à traiter une demande de table qui contient un qualificatif M (valeur marginale) qui demande une normalisation; cependant, la variable de classe qui se trouve avec le qualificatif n'est pas une des variables de classe de la table, ce qui rend impossible la normalisation. L'analyste doit veiller à ce que la variable de normalisation soit une des variables de classe.

`fatal error (966): too many variable crossings in XTSPEC # n`

La fonction Tableaux croisés a découvert que la demande de table indiquée compte trop d'imbrications. Elle permet seulement six niveaux de classe. L'analyste doit éliminer au moins un des niveaux de variable de classe de la demande de table.

`error (967): invalid expression label in XTSPEC # n`

La fonction Tableaux croisés a découvert, pendant le traitement d'un qualificatif L pour une expression de totalisation dans la demande de table indiquée, que le qualificatif ne respecte pas la syntaxe d'une étiquette valide. L'analyste devrait modifier le qualificatif L en veillant à ce qu'il ait la forme des deux caractères L= suivi immédiatement d'une "courte" chaîne délimitée par des guillemets et suivie de l'un ou l'autre (1) de l'accolade qui termine le niveau de totalisation,

(2) d'une virgule qui indique la fin des qualificatifs pour l'expression de totalisation, ou (3) d'un ou plusieurs caractères qui préparent le qualificatif suivant.

```
error (971): invalid constant const encountered
```

La fonction Tableaux croisés, dans l'évaluation d'une expression de totalisation, a trouvé une chaîne, *const*, que, "selon elle", devrait être une constante, mais qu'elle ne peut interpréter. L'analyste doit convertir la chaîne en quelque chose que la fonction Tableaux croisés peut comprendre. Cette erreur peut aussi indiquer une erreur de syntaxe de l'expression.

```
error (972): expecting operator but encountered string
```

La fonction Tableaux croisés, dans l'évaluation d'une expression de totalisation, a trouvé une chaîne de texte *string* alors que la syntaxe d'une expression exigerait un opérateur mathématique/logique. L'analyste doit modifier l'expression de façon qu'elle respecte les exigences de la syntaxe.

```
error (975): class variable value error in XTSPEC # n
```

La fonction Tableaux croisés, en tentant de faire le cumul d'une table pour la demande de table indiquée, a trouvé une valeur inappropriée/inattendue pour une des variables de classe de la demande. Si cela s'est produit pour une variable de base de donnée ou une variable modélisée "en boîte noire", alors l'analyste devrait avertir l'équipe de développement du MSPS de façon qu'elle puisse corriger l'erreur. Si la cause est l'une des variables de l'analyste du modèle "en boîte de verre" ou d'une des variables "CL", alors l'analyste a la responsabilité de veiller à ce que les variables de classe prennent seulement des valeurs légitimes et que les étiquettes définies pour la variable de classe existent dans une correspondance 1 à 1 avec les valeurs valides de la variable.

```
error (976): unknown maximum for variable var in XTSPEC # n
```

La fonction Tableaux croisés, en tentant d'utiliser la variable *var* dans la demande de table indiquée, probablement dans le contexte d'utilisation d'une variable de classe ou d'une variable entière définie par l'analyste, a noté qu'il n'y a pas un ensemble approprié d'étiquettes pour les valeurs de la variable (dont la cardinalité définit le nombre de catégories considérées comme légitimes pour la variable). L'analyste doit veiller à ce qu'un ensemble approprié d'"étiquettes de valeur" soit défini avant l'exécution du MSPS. Cette erreur devrait se produire seulement dans les applications en mode "boîte de verre", et sera habituellement le résultat d'une définition incorrecte ou incomplète d'une nouvelle variable entière.

## Annexe B Détails de la syntaxe

Les analystes qui utilisent le MSPS et qui possèdent de bonnes connaissances en informatique reconnaîtront le paramètre XTSPEC comme une chaîne qui doit être formulée et interprétée par le MSPS avant que celui-ci ne puisse produire les tables désirées. La présente annexe donne les éléments de cette syntaxe, c'est-à-dire la syntaxe structurée de XTSPEC. Dans les règles de production suivantes, les accolades regroupent les éléments répétés et sont suivies de l'un des symboles `\*', `?', ou `+'. `\*' signifie 0 ou plus, '+' signifie 1 ou plus et '?' signifie 0 ou 1 de l'élément précédent. Les symboles entre guillemets indiquent les valeurs littérales.

```
Xtspec ::= table_request { ';' table_request }*
```



```

table_request      ::=  { level_of_analysis }? level_list
level_of_analysis  ::=  IN: | NF: | CF: | EF: | HH:
level_list         ::=  level_spec { '*' level_spec }*
level_spec         ::=  class_level | tabulation_level
class_level        ::=  class_var { '+' }?
Tabulation_level   ::=  '{' tab_expr_list '}'
tab_expr_list      ::=  tab_expr { ',' tab_expr }*
tab_expr           ::=  expr { ':' qualifier_list }?
Qualifier_list     ::=  { qualifier }+
qualifier          ::=  L="label"
                       | S=integer
                       | P=integer
                       | M=class_var
expr               ::=  '(' expr ')'
                       | unary_op expr
                       | expr bin_op expr
                       | analysis_var
                       | number
unary_op           :=   '-'
binary_op          :=   '+' | '-' | '*' | '/'

```