

# CSPro はじめに

## 補足

平成 16 年 11 月



独立行政法人

統計センター 研究センター

# 目次

補足について .....	1
ステップ1: サンプルデータについて .....	1
ステップ2: データ辞書の作成 .....	3
ステップ3: データの集計 .....	11

## 補足について

この補足では、CSPro を使って既存の個票データからクロス集計を行う手順を課題形式で説明します。

CSPro で処理する個票データは、次の条件が必要です。

- 固定長形式で収録されていること
- 少なくとも1つの ID 項目が収録されていること

課題では、最初のステップで、使用するサンプルデータの内容を説明します。次のステップでは、サンプルデータの構造を記述するためのデータ辞書を作成します。最後のステップでは、地域別のクロス集計を実行します。

### ステップ1: サンプルデータについて

この説明では、「CSPro はじめに」のチュートリアル<sup>1</sup>のサンプル調査票に基づく入力済みのサンプルデータ<sup>1</sup>を使います。チュートリアルで作成したデータは、個人の項目を含むレコードと世帯の項目を含むレコードの2種類が同一ファイルに記録されていますが、ここでは、世帯内の最大10人の各項目と世帯についての各項目を同一のレコードに記録した、1世帯1レコードの形式を用います。さらに、世帯の所在地に関する地域コードを付加しています<sup>2</sup>。サンプルデータに収録された項目は、次の表のとおりです。

項目名	左端位置	桁数	繰り返し	説明 <sup>3</sup>
一連番号 (ID 項目)	1	4		0001～
世帯員について	5		10	
世帯員 1	世帯員番号	5	2	01～10
	年齢	7	2	△0～99
	性別	9	1	1: 男性、2: 女性
	配偶関係	10	1	1: 既婚、2: 未婚
世帯員	世帯員番号	11	2	01～10
	年齢	13	2	△0～99
	性別	15	1	1: 男性、2: 女性

<sup>1</sup> CSProに付属しているexampleの“Popstan Census 2000.dat”を用いて作成した架空のデータです。

<sup>2</sup> 架空のデータにランダムに地域コードを割り当てています。

<sup>3</sup> △は半角の空白を示します。

2	配偶関係	16	1	1:既婚、2:未婚
:				
世帯員 10	世帯員番号	59	2	01~10
	年齢	61	2	△0~99
	性別	63	1	1:男性、2:女性
	配偶関係	64	1	1:既婚、2:未婚
世帯について		65		
	居住形態	65	1	サンプル調査票参照
	屋根の種類	66	1	サンプル調査票参照
空白		67	1	
世帯の所在地		68		
	都道府県番号	68	2	01~47
	市区町村番号 <sup>4</sup>	70	3	000
	町丁字等コード <sup>5</sup>	73	4	0000

サンプルデータは、統計センターの [ホームページ](#) からダウンロードし、フォルダ“c:\¥CSProTutorial¥example”に、ファイル名“example1.dat”で保存してあるものとします。同様に、ステップ3で用いる地域名称ファイル“japanpref.anm”をあわせてダウンロードし、同じフォルダに保存してください。

まず、テキストビューアーを使って、このサンプルデータを確認します。

- メインメニューの **Tools/Text Viewer** を選択して、テキストビューアーを起動します。
- [ファイルを開く]ダイアログで、サンプルデータのあるフォルダ“c:\¥CSProTutorial¥example”を指定し、ファイル名“example1.dat”を入力します。

メニューの[View]で[Ruler]を選択すると、次のような画面になります。

<sup>4</sup> ステップ3で取り上げる集計表の例では使用しません。

<sup>5</sup> ステップ3で取り上げる集計表の例では使用しません。

```
.....5...10...15...20...25...30...35...40...45...50...55...60...65...70...75...80...85...90...95...100...
1|0001013011023221|11|070000000
2|000201472202221203201204|722051622|22|100000000
3|0003018022022612032122041412|12|400000000
4|000401192202|72203|412|12|130000000
5|000501281102272103|81204|422|11|130000000
6|000601481102452103151204141205111206101207|92208|71209|922|11|340000000
7|000701362202142203121204|82205|82206|422|12|250000000
8|0008013411|21|280000000
9|0009012422|11|380000000
10|001001451102412103191204172205142206102207|412|12|400000000
11|001101371102302103|022|22|270000000
12|001201331102362103192204181205112206|91207|52208|222091812|22|130000000
13|0013012912025722|12|080000000
14|001401271102242103|812|11|110000000
15|0015015111024321031512041012|11|180000000
16|0016015511025321031812041012|11|260000000
17|001701531102342103|41204|41205|022065022|12|270000000
18|001801712202332103152204141205132206112207|52208|322|12|110000000
19|001901312202|212033712041322|22|430000000
20|002001411102392103202204|21205191206132207112208|31209|31210|11211|080000000
21|002101341102332103152204|61205|51206|31207|022|11|030000000
22|002201592202212203|412|12|380000000
23|002301351102372103132204112205|82206|72207|32208|012|11|220000000
24|0024014822|11|280000000
25|002501401102412103201204152205131206111207|92208|62209|41210|31233|110000000
26|002601381102322103131204111205|92206|71207|51208|122092212|22|080000000
27|0027016311024621|11|450000000
28|002801421102352103181204161205141206|92207121208|222|22|130000000
29|002901441102332103|51204|022|21|130000000
30|003001301102272103|51204|22205|012|11|270000000
31|003101311102252103|82204|62205|41206|212|11|270000000
32|003201521102542103261204251205231206221207201208191209162210151222|130000000
33|0033014611023421031412041212|11|400000000
```


1行目は2人の世帯、2行目は5人の世帯、3行目は4人の世帯のデータです。また、各行の最初の4桁が一連番号を表し、67桁目が空白で、68桁目から9桁分が世帯の所在地のコードになっていることなどが確認できます。

次のステップへ続きます。

## ステップ2: データ辞書の作成

このステップでは、前のステップで確認したサンプルデータに対応するデータ辞書を作成し、項目を定義していきます。

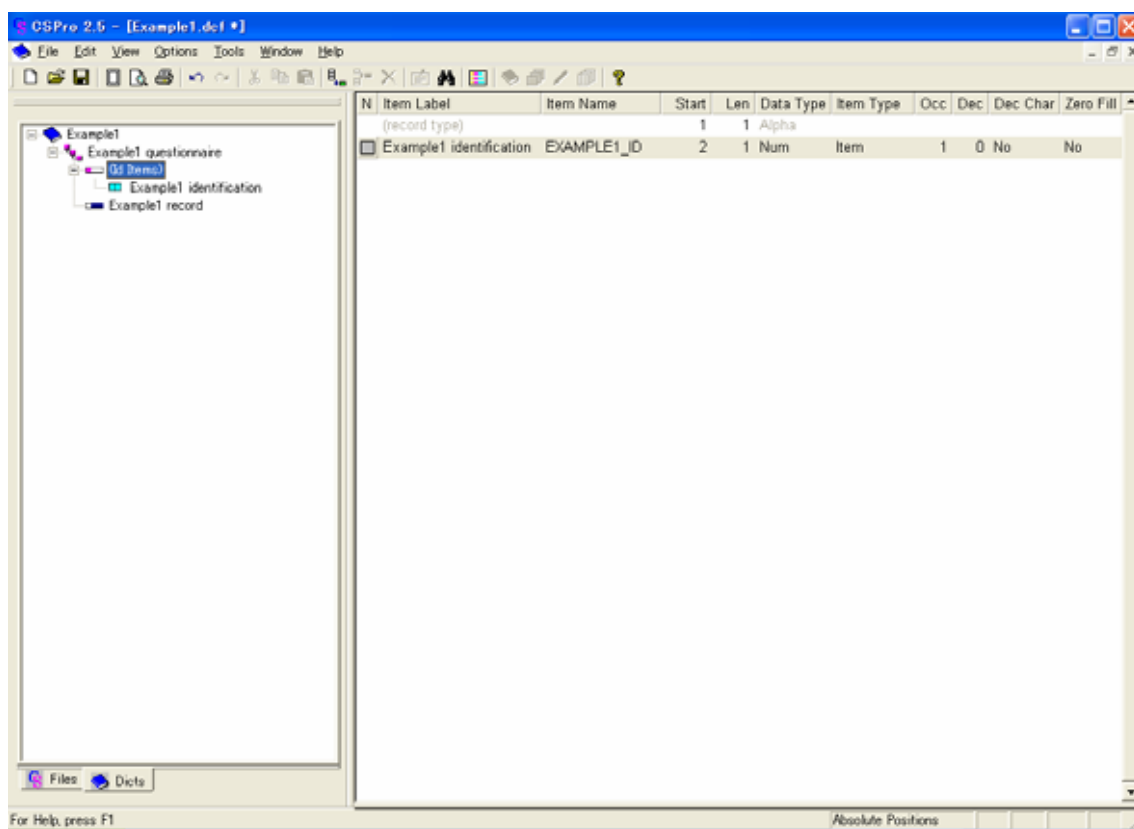
最初に、ウィザードを用いて、新しいデータ辞書を作成します。

- メインメニューの **File/New** を選択するか、ツールバーの  ボタンをクリックします。
- [Data Dictionary File] をクリックします。
- 辞書ファイル名を入力します。例として、ここでは、“Example1” と入力します。

- 辞書を作成するフォルダ名を入力します。例として、ここでは、サンプルデータと同じフォルダ“c:\¥CSPProTutorial¥example”を使用します。
- [次へ]をクリックします。
- チェッカーフラグが表示されます。[完了]をクリックすると、“Example1.dcf”の画面が表示されます。

次に、CSPro が自動的に作成したデータ辞書のスケルトンを変更して、サンプルデータの項目の情報を追加していきます。追加の前に、メインメニューの[File]で[Relative Positions]のチェックをはずし、「Absolute Positioning」モードにします。「Absolute Positioning」モードでは、空白の項目をスキップすることや、ID 項目を任意の位置に指定することができます。

最初に、ID 項目の変更を行います。画面の左側の[(Id Items)]をクリックし、その隣の[+]をクリックしてツリーを展開します。CSPro がすでに1つのID項目「Example1 identification」を作成していることが分かります。

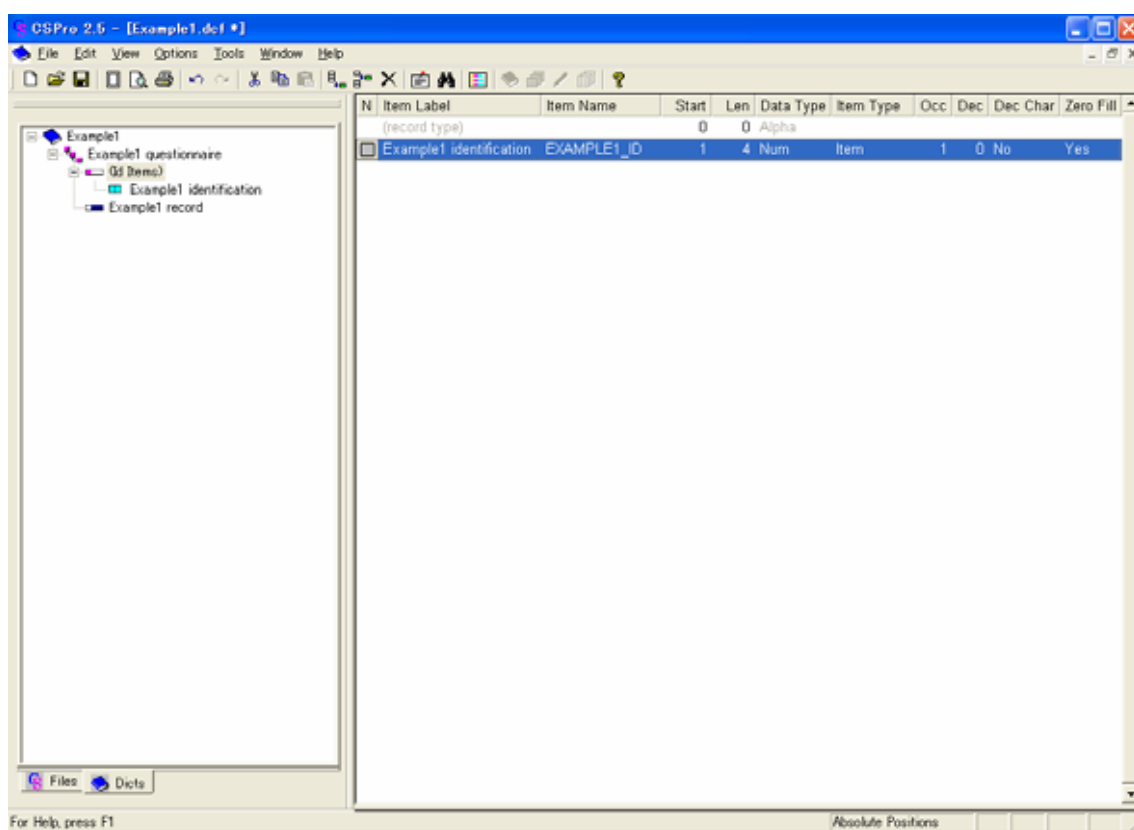


- 画面の右側の一番上、[Item Label]が「(record type)」となっている項目は、レコードの種類を識別するために使います。[Len]欄は、この項目が1文字分の長さであることを示していますが、サンプルデータは、1世帯を1レコードとしているため、レコードの種類を識別は必要ありません。長さを0文字に変更します。[Len]欄に表示されている“1”をダブルクリックします。長さ“1”が青くハイライトされるので、“0”と入力しEnterキーを押し

ます。

- すぐ下の「Example1 identification」で始まる行の [Start] 欄と [Len] 欄は、この項目が 2 桁目から始まり、1 文字分の長さであることを示しています。これを 1 桁目から始めて、4 文字分の長さに変更します。[Start] 欄に表示されている“2”をダブルクリックします。“1”と入力し Enter キーを押すと、[Len] 欄に移動します。“4”と入力し Enter キーを押します。Enter キーを 5 回押して、[Zero Fill] 欄に移動します。[Zero Fill] 欄を No から Yes に変更して、Enter キーを押します。

ここまでの操作が終わると、次のような画面になります。



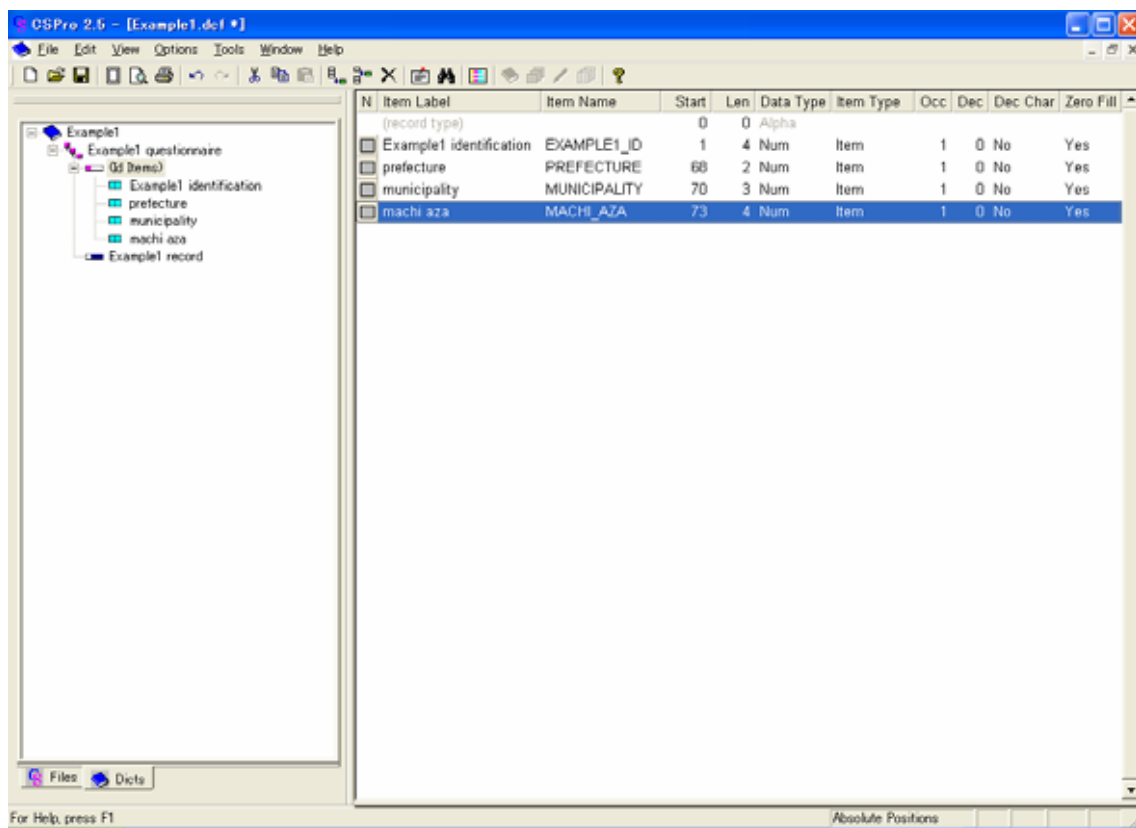
さらに、世帯の所在地を ID 項目として追加します。まず、都道府県番号を追加します。

- ツリーの [(Id Items)] を右クリックして、[Add Item] を選択します。
- [Item Label] として “prefecture” を入力し、Enter キーを押します。
- Enter キーを押して、[Item Name] を PREFECTURE とします。
- [Start] に “68” を入力して Enter キーを押して、[Start] 位置を決定します。
- [Len] に “2” を入力して Enter キーを押し、prefecture が 2 桁であることを指定します。
- Enter キーを 5 回押して、[Zero Fill] 欄に移動します。[Zero Fill] 欄を No から Yes に変更して、Enter キーを押します。

上記の操作を繰り返して、市区町村番号、町丁字等コードを追加します。[Item Label]、[Item Name]、[Start]、[Len]、[Zero Fill]の欄は、それぞれ、次の表の値に変更します。

Item Label	Item Name	Start	Len	Data Type	Item Type	Occ	Dec	Dec Char	Zero Fill
municipality	MUNICIPALITY	70	3	Num	Item	1	0	No	Yes
machi aza	MACHI_AZA	73	4	Num	Item	1	0	No	Yes

ここまでの操作が終わると、次のような画面になります。



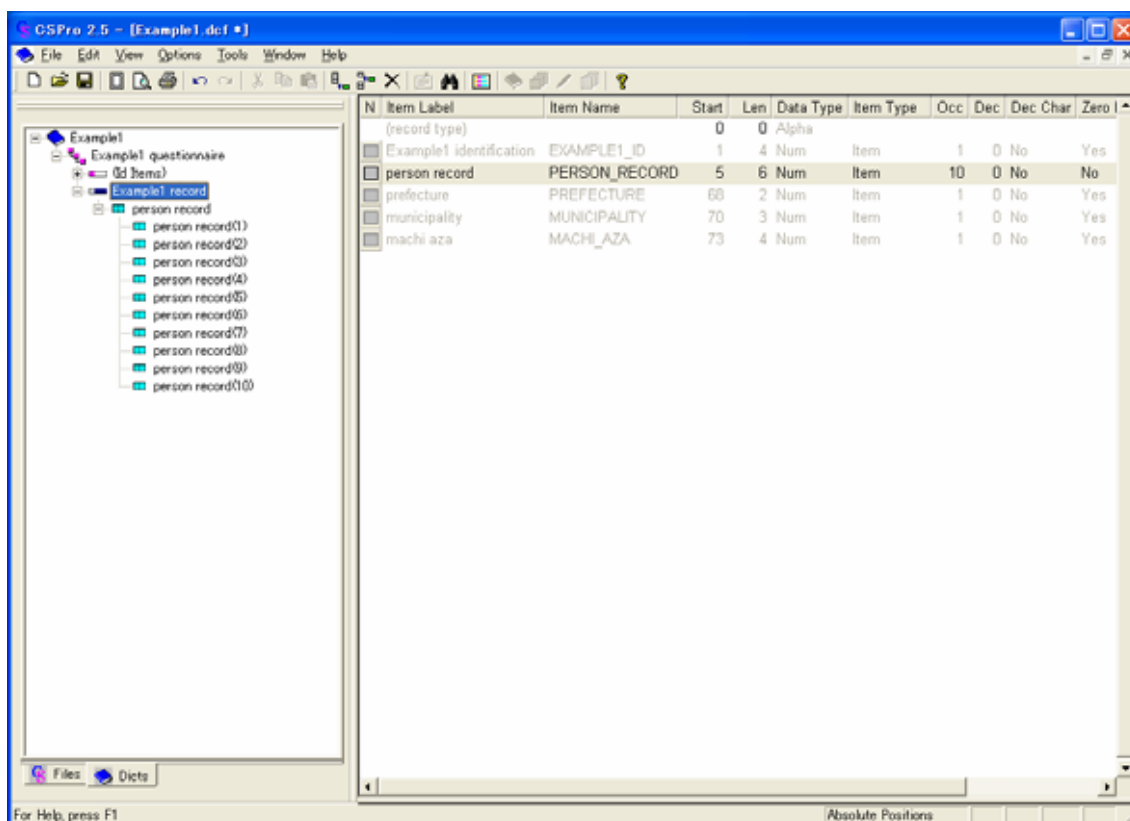
続いて、GSPro が自動的に作成したレコード「Example1 record」に項目を追加していきます。「person record」は世帯員に関する項目を示し、「person no」、「age」、「sex」、「marital status」の4つの小項目(subitem)に細分化されます。また、「Occ」に“10”を入力することで、同じ構造の項目を10個作成することができます。

- ツリーの[Example1 record]を右クリックして、[Add Item]を選択します。
- [Item Label]として“person record”を入力し、Enter キーを押します。



- Enter キーを押して、[Item Name]を PERSON\_RECORD とします。
- [Start]に“5”を入力して Enter キーを押して、[Start]位置を決定します。
- [Len]に“6”を入力して Enter キーを押して、「person record」を 6 桁に指定します。
- Enter キーを2回押して、[Occ]に移動します。“10”を入力して Enter キーを押して、10回の繰り返しを指定します。

画面の左側の [person record] の隣の [+] をクリックしてツリーを展開すると、下図のように person record(1) から person record(10) が作成されていることが分かります。



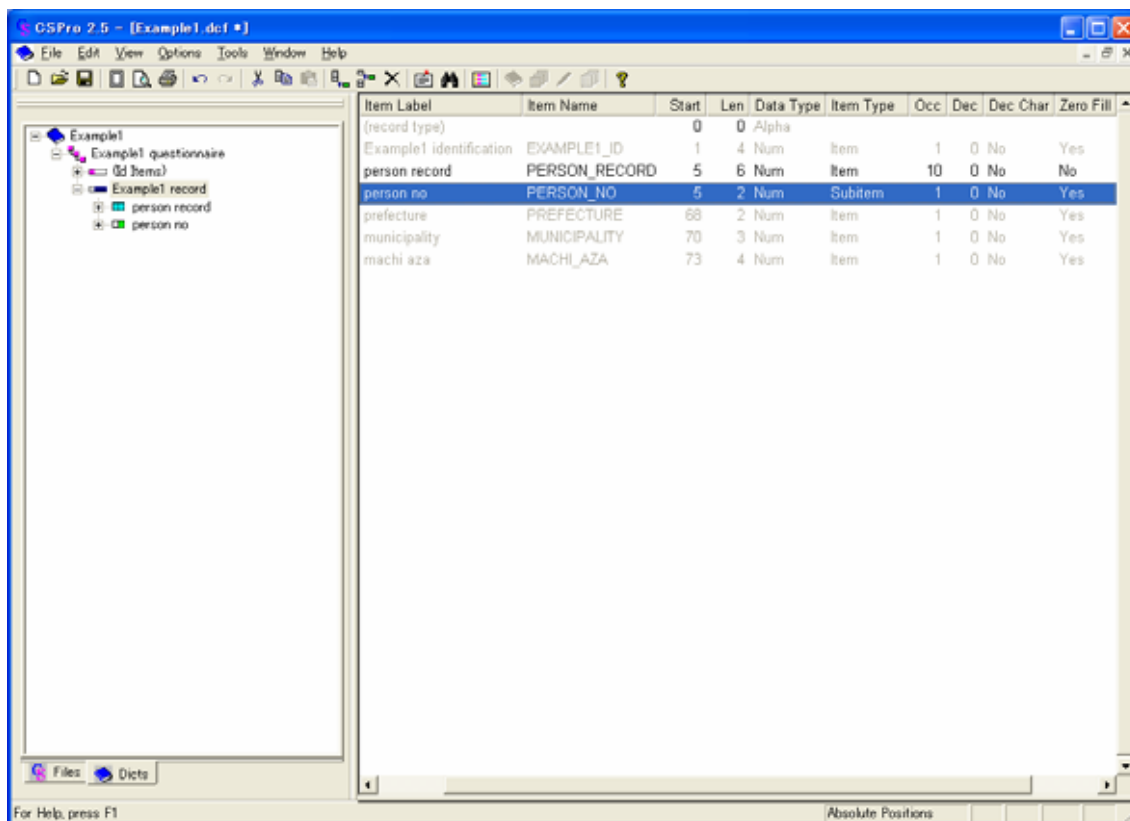
次に、小項目「person no」を作成します。

- ツリーの [Example1 record] を右クリックして、[Add Item] を選択します。
- [Item Label] として “person no” を入力し、Enter キーを押します。
- Enter キーを押して、[Item Name] を PERSON\_NO とします。
- [Start] に “5” を入力して Enter キーを押して、[Start] 位置を決定します。
- [Len] に “2” を入力して 「person no」 を 2 桁に指定します。これは、世帯員 1 から世帯員 10 までの最大 2 桁の番号を入力するためです。
- Enter キーを押して、[Item Type] へ移動します。Item から Subitem に変更し、「person

no」は、「person record」を細分化した小項目として指定します。

- Enter キーを押して、[Zero Fill]へ移動し、No から Yes に変更します。

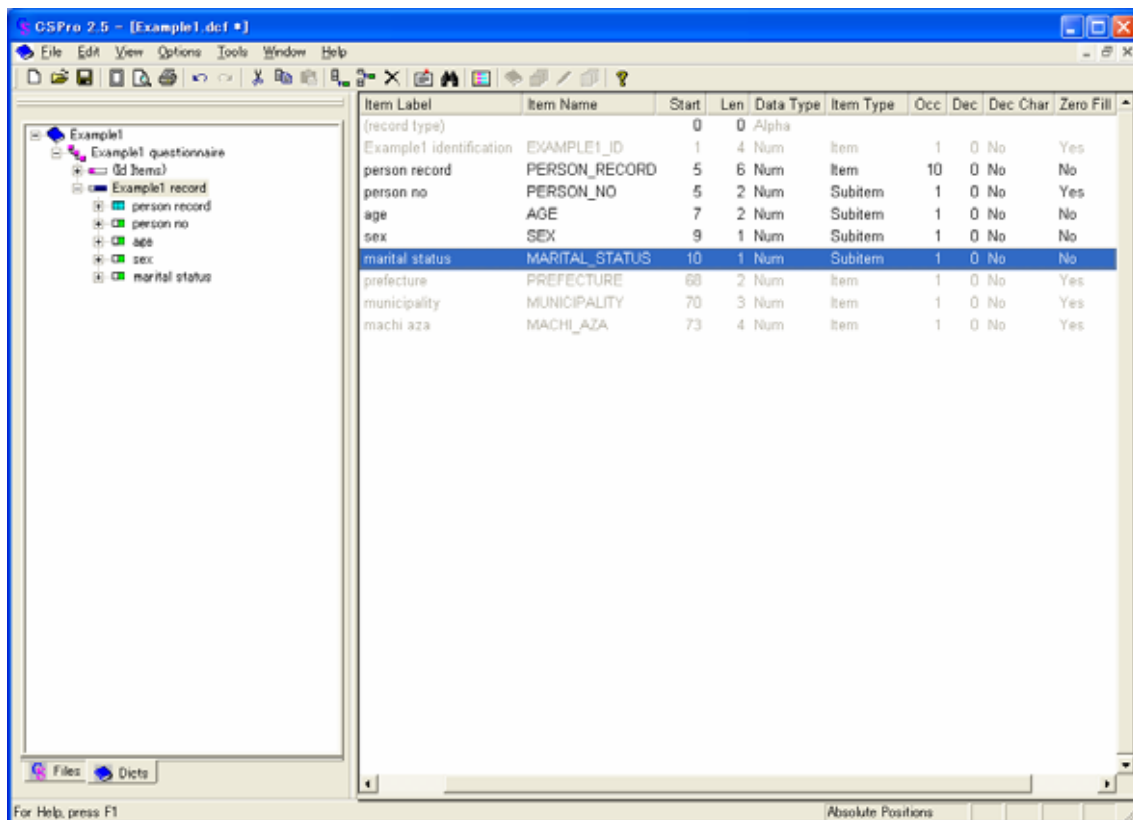
ここまでの操作が終わると、次のような画面になります。



上記の操作を繰り返して、年齢、性別、配偶関係を追加します。[Item Label]、[Item Name]、[Start]、[Len]、[Item Type]の欄は、それぞれ、次の表の値に変更します。

Item Label	Item Name	Start	Len	Data Type	Item Type	Occ	Dec	Dec Char	Zero Fill
age	AGE	7	2	Num	Subitem	1	0	No	No
sex	SEX	9	1	Num	Subitem	1	0	No	No
marital status	MARITAL_STATUS	10	1	Num	Subitem	1	0	No	No

ここまでの操作が終わると、次のような画面になります。

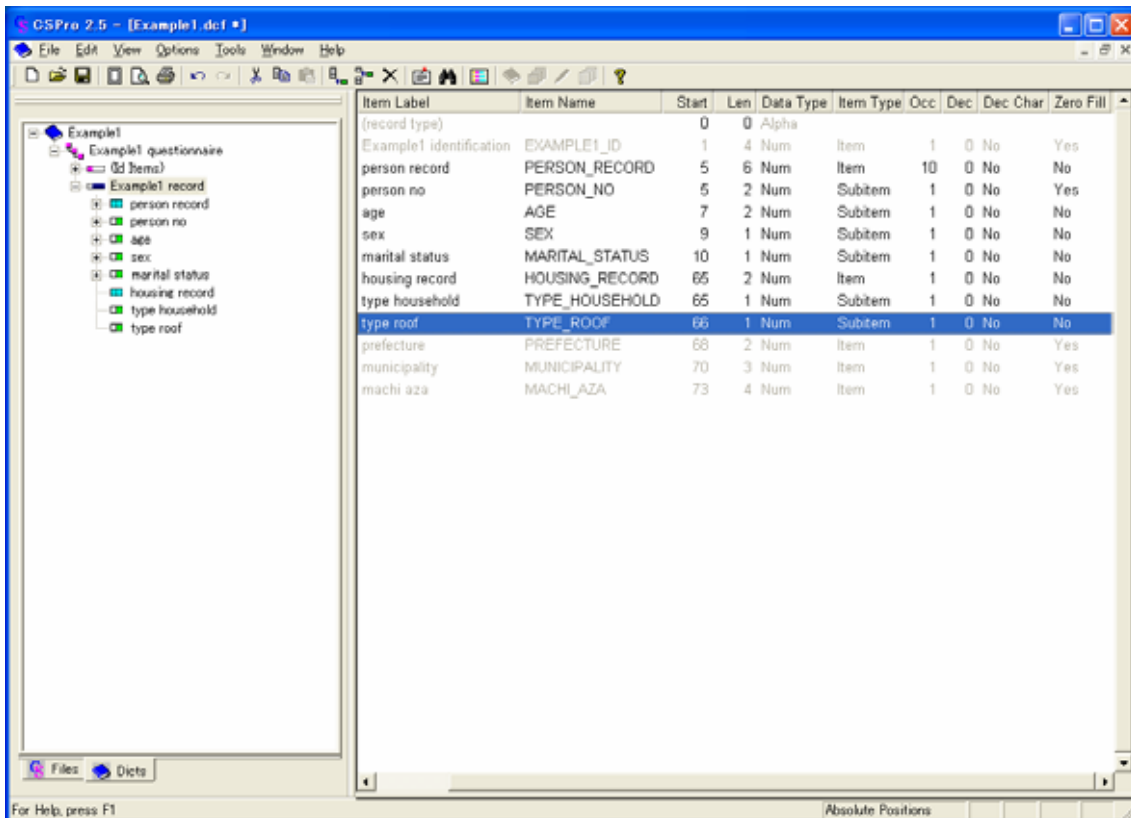



さらに:

- 「housing record」の項目を作成します。
- 「type household」の小項目を作成します。
- 「type roof」の小項目を作成します。
- Esc キーを押して、変更を終了します。

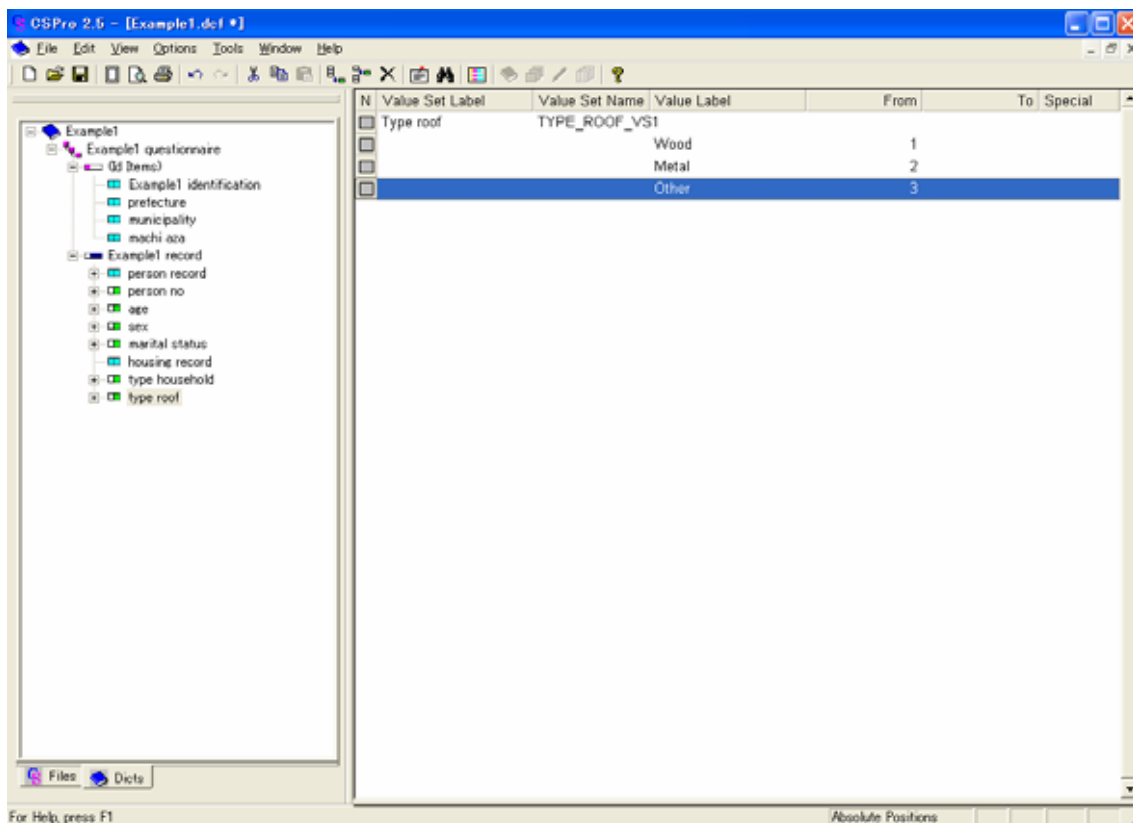
Item Label	Item Name	Start	Len	Data Type	Item Type	Occ	Dec	Dec Char	Zero Fill
housing record	HOUSING_RECORD	65	2	Num	Item	1	0	No	No
type household	TYPE_HOUSEHOLD	65	1	Num	Subitem	1	0	No	No
type roof	TYPE_ROOF	66	1	Num	Subitem	1	0	No	No


ここまでの操作が終わると、次のような画面になります。



- **重要:**ここまでの作業を保存しましょう。メインメニューの **File/Save** を選択するか、ツールバーの  ボタンをクリックします。

最後に、項目の値の定義を作ります。「CSPro はじめに」のチュートリアル課題2ステップ4にしたがって、「Age」、「Sex」、「Marital Status」、「type household」、「type roof」の値を追加してください。「type roof」の値の追加が終わると、次のような画面になります。




- **重要:**ここまでの作業を保存しましょう。メインメニューの **File/Save** を選択するか、ツールバーの  ボタンをクリックします。

次のステップへ続きます。

### ステップ3: データの集計

このステップでは、前のステップで作成したデータ辞書と、地域名が収録された外部ファイルを利用して、地域別のクロス集計を行います。


最初に、ウィザードを用いて、クロス集計アプリケーションを作成します。

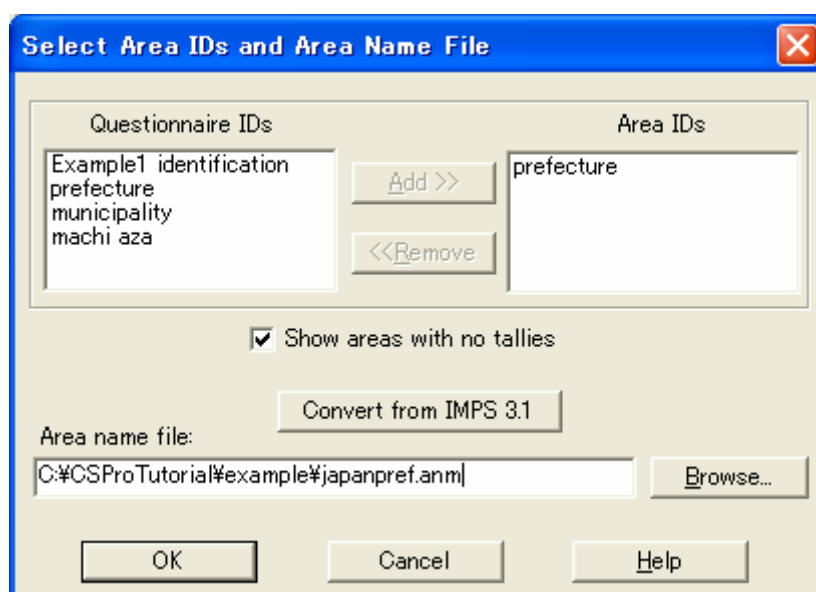
- メインメニューの **File/New** を選択するか、ツールバーの  ボタンをクリックします。
- [Cross Tab Application]をクリックします。
- アプリケーションの名前に“Example1Tab”と入力します。
- サンプルデータと同じフォルダ名 (“c:\%CSProTutorial¥example”)を入力します。

- [次へ]をクリックします。
- 次の画面で、データ辞書を選択するメッセージが表示されます。前のステップで作成したデータ辞書“c:\¥CSProTutorial¥example¥Example1.dcf”を入力して、[次へ]をクリックします。
- チェッカーフラッグが表示されます。[完了]をクリックして完成です。

これで表を設計する準備ができました。

地域別のクロス集計<sup>6</sup>を実行するために、まず、地域IDと地域名称ファイルを指定します。地域名称ファイルは、統計センターのホームページからダウンロードし、フォルダ“c:\¥CSProTutorial¥example”に、ファイル名“japanpref.anm”で保存してあるものとします。

- メインメニューの Edit/Area を選択するか、ツールバーの  ボタンをクリックします。
- [Questionnaire IDs]の中の[prefecture]を選択して[Add>>]ボタンを押し、[Area IDs]に[prefecture]を追加します。
- [Area name file]に“c:\¥CSProTutorial¥example¥japanpref.anm”と入力して、[OK]をクリックします。



続いて、地域別に「age」と「sex」をクロスした表を作成していきます。「age」を表側、「sex」を表頭に配置します。

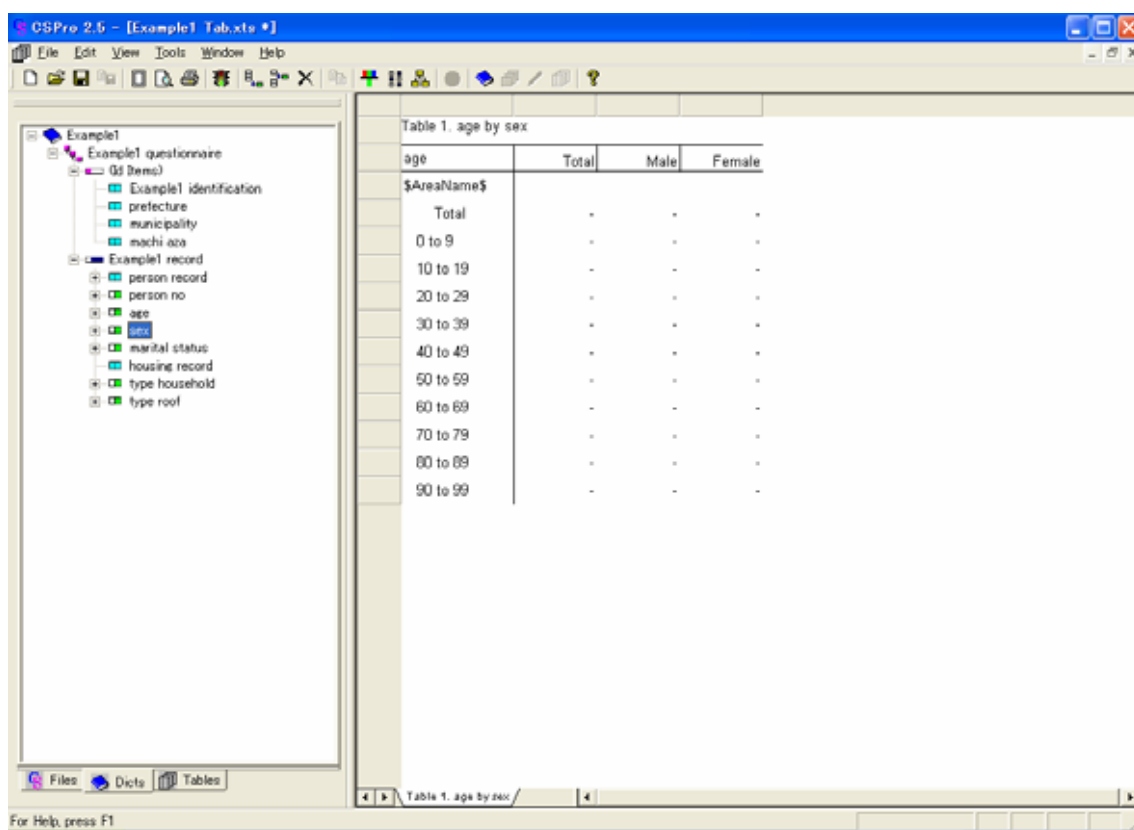
- [age]をツリーから表の左側へドラッグアンドドロップします。ドロップする場所は、横の灰


<sup>6</sup> 地域別のクロス集計においても、最大4変数までのクロス集計が可能です。

色の四角のとなりの白いスペースです。


- [sex]をツリーから表の上側ヘドラッグアンドドロップします。ドロップする場所は、上側の灰色の四角の下の白いスペースです。

ここまでの操作が終わると、次のような画面になります。



- **重要:**ここまでの作業を保存しましょう。メインメニューの **File/Save** を選択するか、ツールバーの  ボタンをクリックします。

ここで、集計を実行します<sup>7</sup>。集計の対象として、サンプルデータを指定します。

- ツールバーの  ボタンをクリックするか、Ctrl-Rを押します。
- 表を作成するためデータファイルの選択のメッセージが表示されたら、フォルダ名が“c:\%CSProTutorial%\example”になっていることを確認し、“example1.dat”を選択します。

<sup>7</sup> 実行の前にメインメニューの **Edit/Parameters** で、[Dump undefined(all tables)]のチェックをはずしてください。

- 集計値が表に表示されます。

次のような表が表示されます。

The screenshot shows the OSPro 2.5 interface with a pivot table titled "Table 1. age by sex". The table is organized by region (Japan and Hokkaido), then by age group, and finally by sex (Total, Male, Female). The data is as follows:

age	Total	Male	Female
<b>Japan</b>			
Total	23,694	11,514	12,180
0 to 9	7,168	3,585	3,583
10 to 19	6,090	2,938	3,152
20 to 29	3,847	1,815	2,032
30 to 39	2,785	1,347	1,438
40 to 49	1,610	794	816
50 to 59	1,079	526	553
60 to 69	713	334	379
70 to 79	315	134	181
80 to 89	76	39	37
90 to 99	11	2	9
<b>Hokkaido</b>			
Total	1,143	562	581
0 to 9	337	173	164
10 to 19	306	154	152
20 to 29	194	90	104
30 to 39	134	67	67
40 to 49	80	37	43
50 to 59	66	24	32
60 to 69	26	11	15
70 to 79	6	3	3
80 to 89	3	2	1
90 to 99	1	1	-

1つのアプリケーションで複数の集計表を作成することができます。ここでは、地域別に「age」と「marital status」をクロスした表を追加します。

- ツールバーの  ボタンをクリックします。

Table 2 のタブが作成され、新しい表を設計する準備ができました。

- [age]をツリーから表の左側へドラッグアンドドロップします。ドロップする場所は、横の灰色の四角のとなりの白いスペースです。
- [marital status]をツリーから表の上側へドラッグアンドドロップします。ドロップする場所は、上側の灰色の四角の下の白いスペースです。

ここまでの操作が終わると、次のような画面になります。



The screenshot shows the CSPro 2.5 interface. On the left is a tree view of the questionnaire structure. The main window displays a table titled 'Table 2. age by Marital Status'. The table has columns for 'age', 'Total', 'Married', and 'Not married'. The rows represent age groups from 'Total' down to '90 to 99'. All data cells in the table contain a hyphen (-).

age	Total	Married	Not married
\$AreaName\$			
Total	-	-	-
0 to 9	-	-	-
10 to 19	-	-	-
20 to 29	-	-	-
30 to 39	-	-	-
40 to 49	-	-	-
50 to 59	-	-	-
60 to 69	-	-	-
70 to 79	-	-	-
80 to 89	-	-	-
90 to 99	-	-	-



ここで、ツールバーの  ボタンをクリックするか、Ctrl-R を押し、集計を実行します。集計の対象として、前の集計と同じサンプルデータを指定します。次の図のように集計値が表に表示されます。Table 1 の集計も同時に行われます。Table 1 のタブをクリックすると、その結果が表示されます。

Table 2. age by Marital Status			
age	Total	Married	Not married
Japan			
Total	23,694	6,927	16,767
0 to 9	7,168	-	7,168
10 to 19	6,090	257	5,833
20 to 29	3,847	1,895	1,952
30 to 39	2,795	2,169	616
40 to 49	1,610	1,281	329
50 to 59	1,079	768	311
60 to 69	713	403	310
70 to 79	315	123	192
80 to 89	76	28	48
90 to 99	11	3	8
Hokkaido			
Total	1,009	303	706
0 to 9	311	-	311
10 to 19	259	14	245
20 to 29	186	108	78
30 to 39	104	86	18
40 to 49	61	50	11
50 to 59	40	24	16
60 to 69	32	17	15
70 to 79	14	3	11
80 to 89	1	1	-
90 to 99	1	-	1

最後に、表を保存します。

- メインメニューの **Save/Save Tables** を選択するか、ツールバーの  ボタンをクリックします。
- 保存する表を選択します。[Select All] ですべての表が選択できます。[OK]をクリックします。
- [名前を付けて保存]のダイアログが表示されます。ここで、ファイルの種類とファイル名を指定します。ファイルの種類は、CSPro 表形式、HTML、RTF、タブ区切りが選択できます<sup>8</sup>。例として、ファイル名“example1 tables.htm”、種類をHTMLとして保存します。

保存したファイルを Excel で開くと、次のような画面になります。

<sup>8</sup> CSPro表形式のファイルは、表ビューアで表示することができます。HTMLはホームページへの掲載や表計算ソフトで扱う場合に便利です。RTFはワープロソフトで扱う場合に便利です。タブ区切りは、表計算ソフトで扱う場合に便利です。

The screenshot shows a Microsoft Excel spreadsheet titled "Table 1. age by sex". The data is organized into three main sections: Japan, Hokkaido, and Aomori-ken. Each section includes a "Total" row and five age group rows (0 to 9, 10 to 19, 20 to 29, 30 to 39, 40 to 49, 50 to 59, 60 to 69, 70 to 79, 80 to 89, 90 to 99). The columns represent age groups, Total, Male, and Female.

age	Total	Male	Female
<b>Japan</b>			
Total	23,694	11,514	12,180
0 to 9	7,168	3,585	3,583
10 to 19	6,060	2,838	3,152
20 to 29	3,847	1,815	2,092
30 to 39	2,785	1,347	1,438
40 to 49	1,610	794	816
50 to 59	1,079	526	553
60 to 69	713	334	379
70 to 79	315	134	181
80 to 89	76	39	37
90 to 99	11	2	9
<b>Hokkaido</b>			
Total	1,009	484	525
0 to 9	311	155	156
10 to 19	259	123	136
20 to 29	186	80	106
30 to 39	104	56	48
40 to 49	61	29	32
50 to 59	40	21	19
60 to 69	32	14	18
70 to 79	14	5	8
80 to 89	1	1	-
90 to 99	1	-	1
<b>Aomori-ken</b>			
Total	235	119	116
0 to 9	61	26	35
10 to 19	59	30	29
20 to 29	43	23	20

おめでとうございます。あなたは、CSpro の補足の課題を完了しました。これで、入力済みの個票データを処理するために、自分のアプリケーションを作成して実行することができるでしょう。詳細は User's Guide を参照してください。CSPro のどのモジュールからでも、メインメニューの Help/Help Topics を選択すれば、User's Guide を参照できます。