

「TABULATEプロシジャを用いた アンケート集計の例題集と SAS次期バージョンの御紹介」

SAS Institute Japan Ltd.

営業本部

東 一成 (Azuma Kazunari)

Jpnkaa@jpn.sas.com

The Power to Know.

Copyright ©2000, SAS Institute Inc. All rights reserved.

Tabulateプロシジャとは？

- BaseSASにおけるレポート作成のためのプロシジャとなっており、オブザベーションの数あるいは数値変数の値を、さまざまな統計量で要約して集計表を作成します。
- 特徴としては以下のとおりです
 - 強力なクロス集計
 - 変数の値での分類と、階層表現も可能
 - タイトルの設定、出力値のフォーマット化など柔軟に行える
 - 様々な統計量を選択可能

The Power to Know.

Tabulate プロシジャの構文-1

```
PROC TABULATE <option-list>;  
  CLASS class-variable-list;  
  VAR analysis-variable-list;  
  FREQ variable;  
  WEIGHT variable;  
  FORMAT variable-list-1 format-1 <...variable-list-n format-n>;  
  LABEL variable-1='label-1' <...variable-n='label-n'>;  
  BY <NOTSORTED> <DESCENDING> variable-1  
    <...<DESCENDING> VARIABLE-N>;  
  TABLE <<page_expression,> row_expression,>  
    column_expression  
    </ table-option-list>;  
  KEYLABEL keyword-1 ='description-1'  
    <...keyword-n='description-n'>;
```

The Power to Know.

Tabulate プロシジャの構文-2

```
PROC TABULATE <option-list>;  
  CLASS class-variable-list;  
  VAR analysis-variable-list;  
  TABLE <<page_expression,> row_expression,>  
    column_expression </ table-option-list>;
```

The Power to Know.



サンプルデータとその項目

変数	ラベル	タイプ
Q1_1	年代	文字
Q1_2	睡眠時間	数値
Q1_3	家族人数	数値
Q1_5	居住地区	文字
Q2_1	仕事に対する考えかた	数値
Q2_2_1	関心(映画)	数値
Q2_2_2	関心(スポーツ)	数値
Q2_2_3	関心(健康)	数値
Q2_2_4	関心(ファッション)	数値
Q2_2_5	関心(ビジネス)	数値
Q2_2_6	関心(アウトドア)	数値
Q2_2_7	関心(コンピュータ)	数値
Q2_2_8	関心(住まい)	数値
Q2_2_9	関心(インテリア)	数値
Q2_2_10	関心(グルメ)	数値
Q2_2_11	関心(手料理)	数値
Q2_2_12	関心(マネー)	数値
Q2_2_13	関心(ダイエット)	数値
Q2_2_14	関心(読書)	数値
Q2_2_15	関心(音楽鑑賞)	数値
Q2_2_16	関心(ショッピング)	数値
Q2_2_17	関心(その他)	数値
Q2_3_1	消費者意識(情報収集を詳しくする)	文字
Q2_3_2	消費者意識(高級品を買いたくなる)	文字
Q2_3_3	消費者意識(流行に敏感である)	文字
Q2_3_4	消費者意識(店員の商品知識を重視する)	文字
Q2_3_5	消費者意識(納得するまで吟味する)	文字
Q2_3_6	消費者意識(安くなるのを待つ)	文字
Q2_3_7	消費者意識(店員のお勧め品を買う)	文字
Q2_4_1	コンピュータを利用する	文字
Q2_4_2	携帯電話を利用する	文字
Q2_4_3	インターネットのメールを利用する	文字
Q2_4_4	Webサイトの閲覧を行っている	文字
Q1_6A	住居区分	文字
ID		数値



サンプルプログラム1

```
proc tabulate data=mktwg.sample;
  class Q1_5;
  tables Q1_5;
run;
```

居住地区			
九州	西日本	東日本	北海道
N	N	N	N
102.00	86.00	84.00	228.00

分類変数のラベル
 分類変数の値
 統計量キーワード
 オブザベーションの集計値

上記は最も簡単なサンプルプログラムとなっている。
 var変数を指定しない場合は、統計量はNとなる。

The Power to Know.

サンプルプログラム2

```
proc tabulate data=mktwg.sample;
  class Q1_5 Q1_1;
  tables Q1_5 Q1_1;

run;
```

居住地区				年代			
九州	西日本	東日本	北海道	10代	20代	30代	40代
N	N	N	N	N	N	N	N
102.00	86.00	84.00	228.00	92.00	142.00	96.00	60.00

分類変数を区切って並べると、各変数ごとに度数分布表が出てきます。

The Power to Know.

サンプルプログラム3

```
proc tabulate data=mktwg.sample;
  class Q1_5 Q1_1;
  tables Q1_5,Q1_1;

run;
```

	年代					
	10代	20代	30代	40代	50代	60代
	N	N	N	N	N	N
居住地区						
九州	26.00	26.00	20.00	15.00	5.00	4.00
西日本	16.00	21.00	20.00	8.00	11.00	4.00
東日本	15.00	32.00	9.00	9.00	4.00	7.00
北海道	35.00	63.00	47.00	28.00	18.00	19.00

分類変数をカンマで区切ると、2次元のクロス集計表が作成されます。

The Power to Know.

サンプルプログラム4

```
proc tabulate data=mktwg.sample;
  class Q1_5 Q1_1;
  tables Q1_5 * Q1_1;
run;
```

居住地区							
西日本				東日本			
年代				年代			
20代	30代	40代	50代	60代	70代	10代	20代
N	N	N	N	N	N	N	N
21.00	20.00	8.00	11.00	4.00	6.00	15.00	32.00

分類変数をアスタリスクで区切ると、階層型の集計表 **2次元のクロス集計表** が作成されます。

サンプルプログラム5

```
proc tabulate data=mktwg.sample;
  class Q1_5 Q1_1;
  var Q1_2;
  tables (Q1_5 Q1_1) * Q1_2;
run;
```

居住地区				年代			
九州	西日本	東日本	北海道	10代	20代	30代	40代
睡眠時間	睡眠時間	睡眠時間	睡眠時間	睡眠時間	睡眠時間	睡眠時間	睡眠時間
Sum	Sum	Sum	Sum	Sum	Sum	Sum	Sum
299.00	280.50	226.00	641.10	403.00	497.10	247.50	93.00

上記のように記述すると、居住区 (Q1_5) と年代 (Q1_1) でそれぞれ睡眠時間を集計して並列で表示しています。Var変数を指定した場合は、SUMが自動的に設定されます。

サンプルプログラム6

```
proc tabulate data=mktwg.sample;
  class Q1_5 Q1_1;
  var Q1_2;
  keylabel SUM='合計';
  tables Q1_5 * Q1_2 , Q1_1/RTS=35;
run;
```

			年代				
			10代	20代	30代	40代	50代
居住地区	睡眠時間	合計					
九州	睡眠時間	合計	118.50	92.50	47.50	22.50	3.00
西日本	睡眠時間	合計	58.00	100.00	57.00	12.50	18.00
東日本	睡眠時間	合計	60.00	112.50	20.50	8.00	16.00
北海道	睡眠時間	合計	166.50	192.10	122.50	50.00	16.00

上記のように記述を行うと、住居区分別、年代別に睡眠時間の2次元集計表を作成できます。

The Power to Know.

サンプルプログラム7

```
proc tabulate data=mktwg.sample;
  where Q1_1='20代' or Q1_1='30代';
  class Q1_5 Q1_1;
  tables (Q1_5 ALL), (Q1_1 ALL) * (N PCTN)/RTS=20;
run;
```

			年代					
			20代		30代		All	
			N	PctN	N	PctN	N	PctN
居住地区								
九州	26.00	10.92	20.00	8.40	46.00	19.33		
西日本	21.00	8.82	20.00	8.40	41.00	17.23		
東日本	32.00	13.46	9.00	3.78	41.00	17.23		
北海道	63.00	26.47	47.00	19.75	110.00	46.22		
All	142.00	59.66	96.00	40.34	238.00	100.00		

行次元、列次元にオブザベーション数がわかる表を作成し、件数と同時に百分率も同時に並列で表示している。またWhere句で読み込む世代の条件設定を行っている。

The Power to Know.

サンプルプログラム8

```
proc tabulate data=mktwg.sample;
  class Q1_1 Q1_5;
  keylabel N='人数' PCTN='%';
  tables Q1_1=' ' ALL='合計',
         (Q1_5=' ' ALL='合計')*(N*F=6. PCTN*F=6.2)/RTS=25;
run;
```

	九州		西日本		東日本		北海道		合計	
	人数	%	人数	%	人数	%	人数	%	人数	%
10代	26	5.20	16	3.20	15	3.00	35	7.00	92	18.40
20代	26	5.20	21	4.20	32	6.40	63	12.60	142	28.40
30代	20	4.00	20	4.00	9	1.80	47	9.40	96	19.20
40代	15	3.00	8	1.60	9	1.80	28	5.60	60	12.00
50代	5	1.00	11	2.20	4	0.80	18	3.60	38	7.60
60代	4	0.80	4	0.80	7	1.40	19	3.80	34	6.80
70代	6	1.20	6	1.20	8	1.60	18	3.60	38	7.60
合計	102	20.40	86	17.20	84	16.80	228	45.60	500	100.00

左記のプログラムはほとんど、「サンプルプログラム8」と変わらないが、keylabelの指定やALLにラベルなどを当てて、表が見やすくなっている。この際に%を算出している母数は合計の500件である。

サンプルプログラム9

```
proc tabulate data=mktwg.sample;
  class Q1_1 Q1_5;
  keylabel N='人数' PCTN='%';
  tables Q1_1=' ' ALL='合計',
         (Q1_5=' ' ALL='合計')*(N*F=6. PCTN<Q1_1 ALL>*F=6.2)/RTS=25;
run;
```

	九州		西日本		東日本		北海道		合計	
	人数	%	人数	%	人数	%	人数	%	人数	%
10代	26	25.49	16	18.60	15	17.86	35	15.35	92	18.40
20代	26	25.49	21	24.42	32	38.10	63	27.63	142	28.40
30代	20	19.61	20	23.26	9	10.71	47	20.61	96	19.20
40代	15	14.71	8	9.30	9	10.71	28	12.28	60	12.00
50代	5	4.90	11	12.79	4	4.76	18	7.89	38	7.60
60代	4	3.92	4	4.66	7	8.33	19	8.33	34	6.80
70代	6	5.88	6	6.98	8	9.52	18	7.89	38	7.60
合計	102	100.00	86	100.00	84	100.00	228	100.00	500	100.00

左記のプログラムはほとんど、「サンプルプログラム9」と変わらないが%を算出している母数は、今度は居住地区別のQ1_1(世代)の合計となっている。

サンプルプログラム 10

```
proc tabulate data=mktwg.sample;
  class Q1_1 Q1_5;
  keylabel N='人数' PCTN='%';
  tables Q1_1=' ' ALL='合計',
         (Q1_5=' ' ALL='合計')*(N*F=6. PCTN<Q1_5 ALL>*F=6.2)
  /RTS=25;
```

run;

	九州		西日本		東日本		北海道		合計	
	人数	%	人数	%	人数	%	人数	%	人数	%
10代	26	28.26	16	17.39	15	16.30	35	38.04	92	100.00
20代	26	18.31	21	14.79	32	22.54	63	44.37	142	100.00
30代	20	20.83	20	20.83	9	9.38	47	48.96	96	100.00
40代	15	25.00	8	13.33	9	15.00	28	46.67	60	100.00
50代	5	13.16	11	28.95	4	10.53	18	47.37	38	100.00
60代	4	11.76	4	11.76	7	20.59	19	55.88	34	100.00
70代	6	15.79	6	15.79	8	21.05	18	47.37	38	100.00
合計	102	20.40	86	17.20	84	16.80	228	45.60	500	100.00

左記のプログラムはほとんど、「サンプルプログラム10」と変わらないが%を算出している母数は、先ほどとは逆で世代別のQ1_5(居住地区)の合計となっている。

The Power to Know.

サンプルプログラム 11

```
proc tabulate data=mktwg.sample;
  title 'コンピュータの利用状況';
  class Q1_1 Q1_5 Q2_4_1;
  keylabel N='人数' PCTN='%';
  tables Q2_4_1=' ' ALL='総合', Q1_1 = ' ' ALL='合計',
         (Q1_5 = ' ' ALL='合計')*(N*F=6. PCTN<Q1_1*Q1_5 Q1_1*ALL ALL*Q1_5 ALL*ALL>*F=6.2)
  /RTS=25 BOX=_PAGE_ CONDENSE;
```

run;

今使っていない	九州		西日本		東日本		北海道		合計	
	人数	%	人数	%	人数	%	人数	%	人数	%
10代	9	4.86	6	3.24	2	1.08	9	4.86	26	14.05
20代	11	5.95	8	4.32	12	6.49	18	9.73	49	26.49
30代	12	6.49	5	2.70	2	1.08	14	7.57	33	17.84
40代	5	2.70	3	1.62	3	1.62	13	7.03	24	12.97
50代	3	1.62	7	3.78	2	1.08	8	4.32	20	10.81
60代	2	1.08	1	0.54	5	2.70	8	4.32	16	8.65
70代	2	1.08	2	1.08	4	2.16	9	4.86	17	9.19
合計	44	23.78	32	17.30	30	16.22	79	42.70	185	100.00

左記のプログラムはほとんど、「サンプルプログラム11」に変数Q2_4_1を分類変数に加えて、より細かく集計を行っている。最後にBOX=オプションで追加した分類変数を表示している。またCONDENSEによりアウトプットでなるべく改ページを行わないようにしている。

The Power to Know.

サンプルプログラム 12

```
proc means data=mktwg.sample nway noprint;
  class q1_1;
  var q2_2_1 -- q2_2_17;
  output out=test(drop=_type_) sum=sum1- sum17;
run;
proc tabulate data=test;
  class Q1_1;
  var _freq_ sum1-sum17;
  tables Q1_1='',(sum1 sum2 sum3 sum4 sum5 sum6 sum7 sum8 sum9
    sum10 sum11 sum12 sum13 sum14 sum15 sum16 sum17)*
  (SUM='人数'*F=6. PCTSUM<_freq_>='%'*F=6.2)/RTS=10 CONDENSE;
run;
```

上記のプログラムではQ1_1が世代別でQ2_2_1-Q2_2_17はマルチアンサー形式となっていて、関心のあることに対して「1」のフラグがたっている。

今回の処理では、まずmeansプロシジャによって世代別に回答があった件数をそれぞれの質問に対して集計を行う。

次にtabulateプロシジャによって、それぞれの世代で何名が何に関心があるのかを、パーセンテージと件数で集計を行っている。

The Power to Know.

サンプルプログラム 12結果画面

	関心(スポーツ)		関心(健康)		関心(ファッション)		関心(ビジネス)		関心(アウトドア)		関心(コンピュータ)			
	人数	%	人数	%	人数	%	人数	%	人数	%	人数	%		
10代	50	54.35	56	60.87	26	28.26	32	34.78	34	36.96	38	41.30	26	28.26
20代	87	61.27	84	59.15	41	28.87	55	38.73	50	35.21	55	38.73	49	34.51
30代	57	59.38	72	75.00	34	35.42	46	47.92	38	39.58	39	40.63	45	46.88
40代	35	58.33	44	73.33	23	38.33	24	40.00	25	41.67	18	30.00	37	61.67
50代	19	50.00	26	68.42	10	26.32	10	26.32	10	26.32	11	28.95	8	21.05
60代	16	47.06	31	91.18	9	28.47	10	29.41	8	23.53	14	41.18	2	5.88
70代	15	39.47	30	78.95	13	34.21	9	23.68	10	26.32	17	44.74	3	7.89

(Continued)

The Power to Know.

 sas
e-Intelligence

 sas
e-Intelligence

The Power to Know™

The Power to Know.