

SAS Technical News

Summer 2003

*For Higher
Customer Satisfaction,
We Bridge
the SAS System
Between
Customer's World.*

CONTENTS

- I** SAS® Enterprise Guide® 2.0による記述統計の活用
- 6** Q&A
- 11** 新刊マニュアルのご紹介
- 12** SASトレーニングのお知らせ
- 12** 最新リリース情報

特集

SAS® Enterprise Guide® 2.0による記述統計の活用

1. はじめに

今回の特集では、前号に続き「SAS® Enterprise Guide® 2.0」(以下EG 2.0)について取り上げ、データの概要を把握する上で、簡単かつ大変有効な記述統計の手法を紹介します。SAS Enterprise Guideでは、データの入力や、要約統計量(平均値、標準偏差等)の計算を、マウスでポイント&クリックすることで実行できます。この特集では、分割表、要約統計量等の算出の手順を具体的に説明します。また、視覚的にデータを捉えることができるグラフについても説明しているので、データ分析を始める際に活用できる内容となっています。また、EG 2.0で実現可能な、高度な分析機能についても紹介します。なお、この特集ではデータセットの入力を前提としています。データセットの入力に関しては、EG 2.0のメニューから、[ヘルプ] [Enterprise Guideヘルプ]を選択し、[データの操作]に関する記述を参照してください。

2. EG 2.0におけるデータの種類の

2.1 EG 2.0の画面の説明

EG 2.0を起動して、新規データセットを入力、または既存のデータセットを読み込んだ場合、次ページ図1のような画面が現れます。図1の左上には「プロジェクト」ウィンドウが表示され、データセット、分析手法、結果、コード、ログ等のノードを統括的に管理できます。「プロジェクト」ウィンドウの下には「タスクリスト」ウィンドウが表示されています。このウィンドウには、利用可能である全てのタスクがカテゴリ別と名前別に表示されています。「データグリッド」と呼ばれる中央の部分には、データテーブル、出力結果等が表示されます。その他、画面下には「タスクステータス」ウィンドウ、右上には「バインダリスト」ウィンドウ、その下には「サーバーリスト」ウィンドウが表示されます。これらのウィンドウが表示されていない場合には、メニューの「表示」から選択できます。

ここでは、ライブラリSASHELPにあるデータセットCLASS(以下CLASSデータ)をサンプルデータとして使用しています。

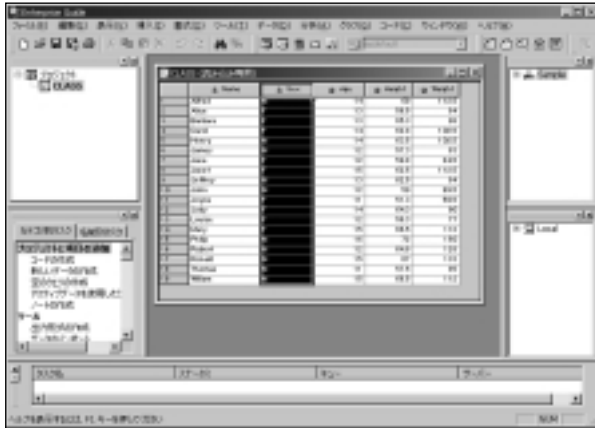


図1: データ入力後の画面

2.2 データテーブルにおける変数の種類

図1のデータテーブルで、変数名(列名)と共に記号が表示されていることに注目してください。EG 2.0では、以下の5種類の変数を、対応する記号と共に指定できます(図2参照)。変数の種類を変更するには、データテーブルで変数名のセルを右クリックして「プロパティ」を選択し、「列のプロパティ」の「種類」にて選択してください(データテーブルが読み込み専用の場合、メニューから「データ」→「プロテクト」を選択し、プロテクトを外しておく必要があります)。



図2: 変数の種類と記号の対応表

EG 2.0では、データテーブルでの変数の種類を表す記号を用いて、分析に用いる変数の選択が容易にできるようになっています。つまり、各分析もしくはグラフの機能を実行する際に、変数として選択できる変数の種類がこの記号を用いて表示されます。上記では5種類のデータを紹介しましたが、この特集では頻繁に用いられる文字データと数値データを使用し、記述統計の手法に関して説明します。記述統計とは、統計学の一部であり、データの概要を把握する上で大変有効な手段となります。記述統計における分割表、要約統計量を求めることにより、入力したデータの検証もすることができます。また、記述統計の数値を視覚的に表現する方法として、グラフの作成にも説明します。

3. 記述統計の手法

ここでは、数多くある分析手法のうち、基本的な記述統計に関して具体的な例を用いながら説明します。具体例としてすでに紹介したCLASSデータを用います。このデータには、19人の年齢、性別、体重、身長が記録されています。また、各個人を識別するために名前も変数として含まれています。このデータでは、性別(Sex)と名前(Name)は文字データ、年齢(Age)、体重(Weight)、身長(Height)は数値データとなります。以下の説明では、まず男女それぞれ何人ずついるかを調べます(度数表)。次に、年齢と性別を用いて分割表を作成します。さらに、性別ごとに要約統計量を求めるとともに、体重と身長の相関を求めます。

3.1 文字データに対する分析

性別などの文字データは質的変数として扱われます。この場合、平均、分散等の数値的な要約統計量は解釈することができません。たとえば、変数Sexにおける水準値、F(女性)とM(男性)の平均を算出することに意味がなく、またどのように算出するかについても不明です。そのため、変数の各水準におけるオブザベーション数がより有益な情報となります。つまり、19人のうち男女それぞれ何人いるかを調べることになります。このような場合、必要となる度数表は、EG 2.0において以下の選択をすることで作成できます。

一元度数表の作成

- 1)メニューから「分析」→「記述統計量」→「一元度数表」を選択する
- 2)分析変数を選択する

ここでは、文字データである変数Sexを分析変数として選択します(図3参照)。



図3: 「一元度数表」における「変数」タブの画面

図3の「一元度数表の変数の役割」における分析変数の前には赤い三角の記号があるため、文字データである変数Sexを分析変数に指定できます。しかし、度数変数に対する記号は青い丸なので、変数Sexを指定することはできません。

- 3) 「統計量」タブをクリックし、「度数表のオプション」で「度数とパーセント」を選択する(図4参照)

「統計量」タブでは、度数表のオプションを指定できるほかに、欠損値をどのように扱うかを選択することもできます。また、必要に応じて画面(図4)の該当箇所を選択することで、二項分布の割合の検定、2乗検定を実行することもできます。



図4: 「一元度数表」における「統計量」タブの画面

4) [グラフ] タブをクリックし、縦棒グラフを選択する (図5参照)



図5: [一元度数表] における [グラフ] タブの画面

5) [完了] をクリックする

以上の操作により、下記のように分析変数Sexに対する一元度数表と対応するプロットが作成されます (図6参照)。

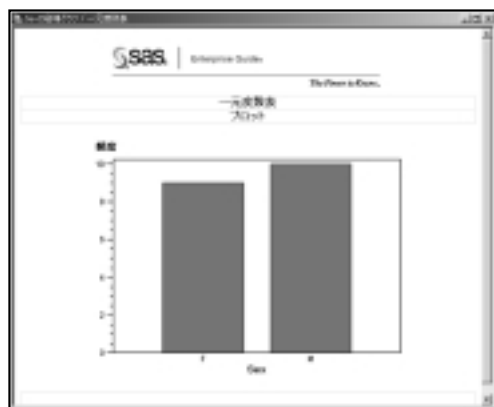


図6: 一元度数表の結果 (上) とプロット (下)

図6の一元度数表には、19人のうち、女子が9人で男子が10人であることが表示されています。また、それぞれの性別の全体に対する比率がパーセント列に表示されています。この比率を円グラフで視覚的に表すには、下記の手順で [グラフ] 機能を利用します。

円グラフの作成

- 1) メニューから [グラフ] [円グラフ] を選択する
- 2) [グラフギャラリー] タブにて [単純円] を選択する
- 3) [変数] タブをクリックし、グラフ変数を選択する (ここでは変数Sexをグラフ変数として指定)
- 4) [表示] タブをクリックし、作成されるグラフの色、レイアウト、タイトル等を指定する
ここでは、[レイアウト] として、名前 (ラベル)、円の割合、統計値 (度数) を円の外側に表示することを指定しています。
- 5) [完了] をクリックする

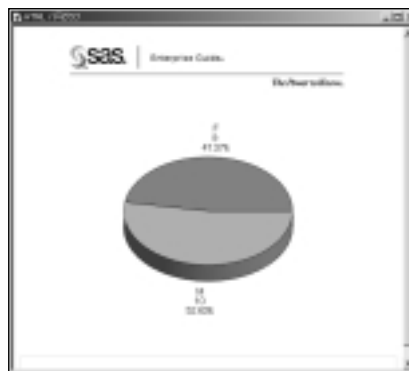


図7: 円グラフ

3.2 文字データと数値データに対する分析

「3.1 文字データに対する分析」では1つの変数に対して度数表を作成する方法を紹介しましたが、一般的にはより多くの変数の組み合わせに対しての度数表、つまり分割表を作成することが必要になります。分割表を作成することにより、変数のある組み合わせに対しては多くのオブザベーションが観測されている、もしくは全く観測されていないなどの情報も得ることができます。ここでは、数値データである変数Ageと文字データの変数Sexを用いて、具体的に以下のように選択してください。

分割表の作成

- 1) メニューから [分析] [表分析] を選択する

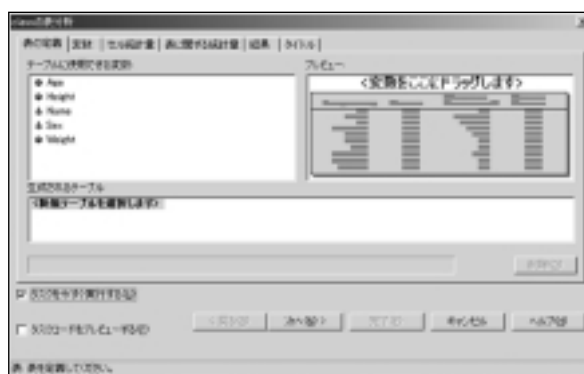


図8: [表分析] における [表の定義] タブの画面

- 2) [表の定義] タブをクリックし、調査対象の変数を [プレビュー] にドラッグする (図8参照)
CLASSデータの場合、変数Ageと変数Sexを [プレビュー] にドラッグし、性別と年齢の分割表を作成します (ここでは変数Age、Sexの順にドラッグします)。

3) 他のタブをクリックし、必要な項目を選択する(図9参照)

[セル統計量]タブでは出力される分割表の中にどのような数値を表示するかを指定できます。ここでは、[セル統計量]タブにて[列のパーセント]、[セルの度数]、[期待セル度数]を選択しています。また、[表に関する統計量]タブでは、分割表に関して統計的に推測をする際に用いられる多くの統計量を選択できます。

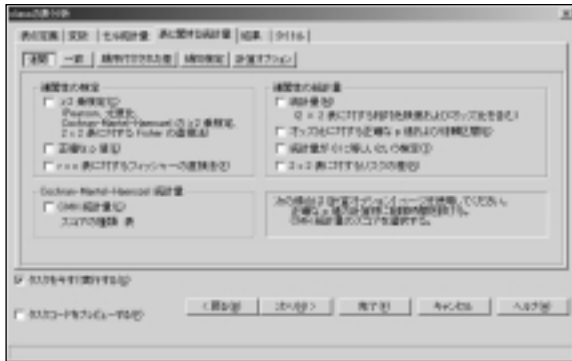


図9: [表分析]における[表に関する統計量]タブの画面

4) [完了]をクリックする



図10: 表分析結果の一部

数値データを扱う場合、上記のように文字データのように扱う場合のほかに、体重、身長のように数値データの値そのものを用いて分析する場合があります。次項では、文字データの各水準に対して、数値データの要約統計量を算出する方法を紹介します。たとえば、前述のCLASSデータでは、体重と身長の平均、分散等を求める場合に、性別に分けて算出したほうがより詳細な情報が得られます。要約統計量は、以下の手順で算出できます。

要約統計量の算出

- 1) メニューから[分析] [記述統計量] [要約統計量]を選択する
- 2) [変数]タブをクリックし、分析変数と分類変数を選択する
ここでは、変数Height、Weightを分析変数として選択します。さらにこれらの要約統計量を性別ごとに求めるために、分類変数として変数Sexを指定します(分類変数を指定しない場合、全体に対する要約統計量が算出されます)。
- 3) [統計量]タブをクリックし、平均値等、必要な要約統計量の項目を選択する
- 4) [グラフ]タブをクリックし、ヒストグラム、箱ひげ図を選択する
- 5) [完了]をクリックする

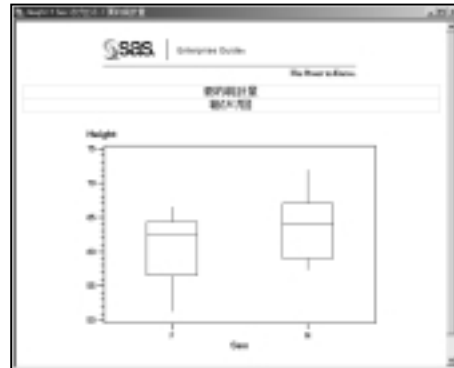


図11: 身長(Height)の性別ごとのヒストグラム(上)と箱ひげ図(下)

出力結果として、分類変数の水準ごとの要約統計量が算出されます。また、出力における数値のみではなく、グラフ(図11参照)をあわせて見ることで、平均値が同じであるかどうかの主観的な判断がしやすくなります。客観的な判断をするためには、統計的な検定が必要です。この場合、適切な手法であるt検定は、EG 2.0においてメニューから[分析] [分散分析] [t検定]を選択し、実行することができます。より具体的な内容に関しては次号以降でご紹介する予定です。

3.3 数値データに対する分析

複数の数値データがある場合、各変数に対する要約統計量と共に、変数間の関係も重要になります。たとえば、身長の高い人は体重も重い傾向にあります。手元にあるデータにそのような関係はあるのかどうかを検証する場合があります。このような数値データ間の相関は、下記の手順で求められます。

相関係数の算出と散布図の作成

- 1) メニューから[分析] [記述統計量] [相関]を選択する
- 2) [変数]タブで調査対象である変数を複数指定する
ここでは、変数WeightとHeightを相関変数として指定しています。他の変数の各水準において別々の分析を行なう場合には、グループ変数にその変数(ここでは変数Sex)を指定します。
- 3) [相関]タブにて相関係数のタイプを指定する
通常の数値データにおける相関係数として、デフォルトではPearsonの相関係数が選択されています。しかし、データによってはPearsonの相関係数が適していない場合もあります。そのため、EG2.0では、Hoeffding、Spearman、Kendallの相関係数を計算することができます。
- 4) [結果]タブで散布図を指定する
相関係数は変数間の連関を表す指標なので、データをグラフにして表示することから、相関係数の数値のみ以上の情報を得ることができる場合もあります(図12参照)。
- 5) [完了]をクリックする

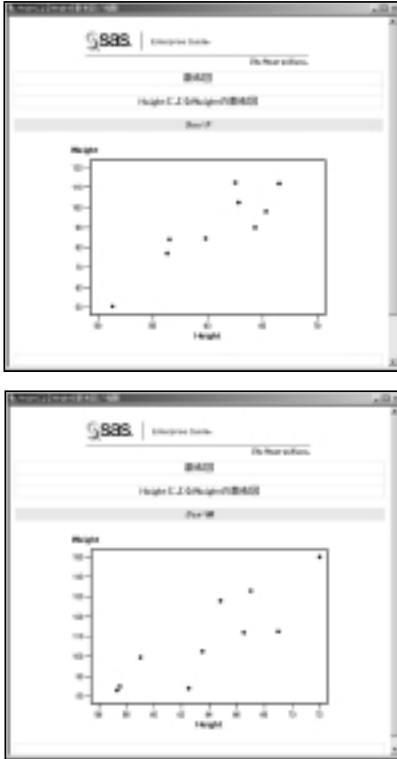


図12: 性別ごとの体重と身長散布図(上 - F、下 - M)

上記の性別ごとの散布図より、男女ともに体重と身長に線形的な相関があると推察できます。つまり、身長の高い人は体重も重いことがグラフから明確にわかります。また、男子のデータは女子よりも散らばっているとグラフから観察できます。

4. その他の機能

「3. 記述統計の手法」では、メニューから[分析]を選択し、[記述統計量]と[表分析]の一部を用いる手法を紹介しましたが、その他にグラフ機能、および各統計手法もプログラムを記述することなく、メニューからポイント&クリックで実行できます。

4.1 グラフ機能

すでに、[記述統計量]に関する説明で[グラフ]タブを紹介しましたが、メニューから[グラフ]を選択し、様々なグラフを作成できます。棒グラフ、円グラフ、散布図、曲面グラフ、等高線グラフ等が用意されているので、目的に沿ったグラフを選んで描画できます。また、[グラフ]機能では、図13の[グラフギャラリー]が表示されるため、名前とグラフのサンプルを参考に、視覚的にグラフを選択できます。



図13: [グラフギャラリー]の一部

4.2 その他の分析機能

メニューから[分析]を選択すると、多くの分析手法が表示されます。すでに[記述統計量]と[表分析]に関して紹介しましたが、その他の頻繁に用いられる手法の概要を説明します。

分散分析

2群における統計量の比較をする際に用いられるt検定から、より複雑な多くの実験計画に対する分析が可能です。また、ランダム効果を含む線形混合モデルについても分析可能です。

回帰

線形的なモデルを当てはめる回帰分析と共に、非線形的なモデルの分析も可能です。また、目的変数(応答変数)が二項である場合に用いられるロジスティック回帰に限らず、一般化線形モデルも扱うことが可能です。

多変量

各オブザベーションに対して複数の変数が観測されている場合、それぞれの変数に対して別々の分析をするよりは、複数の変数における相関も考慮する分析が望ましい場合もあります。このような分析手法は多変量解析と総称され、EG 2.0では、正準相関、主成分分析、因子分析、クラスター分析、判別関数を用いることができます。

これらの手法に関しては、次号以降で操作方法を含めてより具体的にご紹介する予定です。また、前述した手法のほか、生存時間解析、品質管理、時系列分析についてもメニューからマウスでポイント&クリックし、実行することができます。各手法を選択した後、タスクウィンドウが表示されるので、[記述統計量]、[表分析]と同じように、変数に付随する記号を参考にしながら変数を選択し、タブの切り替えで出力結果を操作することで、分析を実行できます。選択オプションの詳細に関しては、[Enterprise Guide ヘルプ]を参照してください。

5. おわりに

本特集では、データ分析の第一段階である記述統計の手法を、EG 2.0を使用して実行する方法を紹介しました。全てのデータ分析における最初のステップとして、簡単に算出できる要約統計量、分割表、グラフを活用してください。また、EG 2.0にて分析可能である高度な統計手法に関しては説明のみとなりましたが、「3. 記述統計の手法」と同様に、ポイント&クリックだけで、プログラムを記述することなく実行できるので、ご活用ください。

Q&A



実行中のマクロのマクロ名を取得する
 文字列の最後のnバイトを抽出する
 算術演算子について
 プログラムの呼び戻し
 欠損値と無効な数値データを区別したい
 任意の日付出力形式を作成したい
 サブセット化のWHEREについて
 UNIX上でのバッチプログラムに引数を渡す方法
 Enterprise Guideで存在しないLOCALサーバーを表示させない方法
 マハラビス距離に基づいたクラスター分析を実行する方法
 要約統計量のデータを用いての検定

Q 現在実行しているマクロのマクロ名を取得することはできますか？

A SAS 8.2から追加された自動マクロ変数SYSMACRONAMEを使って、マクロ名を取得できます。

```
%put Macro Program: &SYSMACRONAME;
```

Q 文字列の先頭の何文字かを抽出するには、SUBSTR関数を利用できます。では、文字列の最後の何文字かを抽出するにはどうすればよいのでしょうか？

A 文字列の最後から文字を抽出するには、文字列全体の長さを求めてからSUBSTR関数を使って最後のnバイトを抽出します。

例: 文字列の最後の3文字を抽出する

```
%let n=3;

data aaa;
  infile cards missover;
  input name $20. ;
  ext=substr(name,length(trim(name)) -&n + 1,&n);
  put ext=;
cards;
sas.exe
sample.txt
a.lzh
a aaaa.log
;
run;
```

Q 比較演算子を用いてデータを抽出する際、IFステートメントでは「=>」が使用できるのに、WHEREステートメントではなぜ「=>」が使用できないのですか？

IFステートメントを利用した例

```
1 DATA test_if ;
2     SET sashelp.class ;
3     IF age => 14 ;
4     RUN ;
```

NOTE: データセット SASHELP.CLASS から 19 オブザベーションを読み込みました。

WHEREステートメントを利用した例

```
6 DATA test_wh ;
7     SET sashelp.class ;
NOTE: SCL プログラムのテキスト
8     WHERE age => 14 ;
```

```
-
22
76
ERROR: WHERE 式の読み込み中に、構文エラーが発生しました。
```

A IFステートメントとWHEREステートメントは、SASプログラムに追加された経緯の違いがあるため、質問のような現象が発生します。

IFステートメント	WHEREステートメント
SAS System バージョン5で追加されたステートメントで、「<=」、「=<」、「>=」、「=>」の4つの比較演算子が、DATAステップ内でのみ利用可能な仕様となっていました。	SAS System バージョン6から導入されたステートメントで、DATAステップだけでなく、SQLプロシジャでも利用可能なステートメントを指したため、当時のANSI準拠に則り、「<=」、「>=」の2つの比較演算子のみを利用可能としました。

Q UNIX版SASをラインモードで使用しています。DMSモードで使用しているときに、一度サブミットしたプログラムを呼び戻して実行したいのですが、何か方法はありますか？

A SAS起動時、または環境設定ファイルでSPOOL オプションを指定すれば、プログラムを呼び戻して実行できます。UNIX版SASをラインモードで起動する際にSPOOLオプションを使用する場合は、下記のように実行します。

```
$/usr/local/sas82/sas_ja -nodms -spool
```

実際にプログラムを呼び戻して実行するには、%incステートメントの後に呼び戻して実行したいプログラムの行番号を指定します。実行したプログラムの行番号を知りたい場合は、%listステートメントを実行すれば、これまでに実行したプログラムの一覧を行番号付きで表示できます。

```
%list;
```

行番号の指定方法は、下記のように%incステートメントを実行します。

呼び戻す行	指定方法	例行番号を指定する
プログラムの任意の1行を呼び戻す	行番号を指定する	1行目を呼び戻す場合 %inc 1;
複数の連続する行を呼び戻す	「- (ハイフン)」でつないで指定する	1~5行目を呼び戻す場合 %inc 1-5;
任意の連続しない複数行を呼び戻す	「: (コロン)」で区切って指定する	2行目と4行目を呼び戻す場合 %inc 2:4;

実行例:

すでに、下記のプログラムが実行済みであると仮定します。

```
data test;
  x=1;
  y=2;
  z=3;
run;
```

%listステートメントを実行し、過去にサブミットしたプログラムの一覧を表示します。

```
6 %list; /* %listで一覧の取得 */

1 data test;
2  x=1;
3  y=2;
4  z=3;
5 run;
6 %list;

7 %inc 1-5; /* 1行目から5行目までを呼び戻して実行 */
```

NOTE: データセット WORK.TEST は 1 オブザベーション、3 変数です。

NOTE: DATA ステートメント 処理 :

```
処理時間      0.00 秒
CPU時間       0.01 秒
```

```
13 data test2;
14 %inc 2:4; /* 2行目と4行目のみを呼び戻して実行 */
17 run;
```

NOTE: データセット WORK.TEST2 は 1 オブザベーション、2 変数です。

NOTE: DATA ステートメント 処理 :

```
処理時間      0.00 秒
CPU時間       0.01 秒
```

Q

欠損値(.)や無効なデータ(たとえば「#」)が含まれているデータを、数値変数としてSASデータセットにしています。SASシステムでは、欠損値、無効なデータともに、欠損値として認識されます。これを区別したいのですが、可能でしょうか。

A

INVALIDDATA=システムオプションを利用すると、無効な数値データが現れた際に、任意の変数に値を割り当てることができます。INVALIDDATA=システムオプションには、アルファベット(大文字または小文字)、ピリオド(.)またはアンダーライン(_)のいずれか1文字を指定します。なお、デフォルト値はピリオドです。

以下のサンプルプログラムをご確認ください。

```
OPTIONS INVALIDDATA='E';

DATA data1;
  INPUT num1 num2;
CARDS;
. 2
3 4
# 5
;
RUN;

PROC PRINT DATA=data1;
RUN;
```

上記のプログラムの実行結果は、下記のとおりです。

OBS	num1	num2
1	.	2
2	3	4
3	E	5

なお、無効な数値データのみを抽出する場合は、以下のようにWHEREステートメントを記述します。

```
DATA result;
  SET data1;
  WHERE num1=.E;
  * IF num1=.E;
RUN;
```

Q

SAS日付値を、SASシステムが標準で提供していない出力形式で利用したいのですが、可能でしょうか。

A

SAS 8から、FORMATプロシジャのPICTUREステートメントにDATATYPE=オプションが追加され、任意の日付出力形式を作成できるようになりました。以下のサンプルプログラムをご確認ください。

```
DATA birthday;
  INPUT date YMMDD10.;
CARDS;
1976/04/05
1976/09/02
2002/09/03
;
RUN;

PROC FORMAT;
  /* PICTUREステートメントにて MM.DD YYYYというsas日付値の出力形式を作成 */
  PICTURE datefmt LOW-HIGH='%0m.%0d %Y'
(DATATYPE=date);
RUN;

PROC PRINT DATA=birthday;
  FORMAT date datefmt.;
RUN;
```

上記のプログラムの実行結果は、下記のとおりです。

```
04.05 1976
09.02 1976
09.03 2002
```

Q

サブセット化のWHEREステートメントで利用できる式には、どのようなものがありますか。

A

サブセット化のWHEREステートメントでは、通常の演算子のほか、次のような条件式が利用できます。

特定範囲を条件とする場合

•BETWEEN - AND

```
data test1;
  do x = 10 to 20;
    output;
  end;
run;
data _null_ ;
  set test1;
  where x between 15 and 18;
```

```
put _all_ ;
run;
```

値を含むものを条件とする場合

• IN

```
data test3;
  do x = 1 to 10;
    output;
  end;
run;
data _null_ ;
  set test3;
  where x in (1 4 6);
  put _all_ ;
run;
```

• CONTAINS

```
data test2;
  s = 'ABCD'; output;
  s = 'AB)D'; output;
run;
data _null_ ;
  set test2;
  where s not contains ')';
  put _all_ ;
run;
```

ブランク、欠損値を条件とする場合

• IS MISSING または IS NULL

```
data test4;
  s = 'A'; output;
  s = ''; output;
  s = 'C'; output;
run;
data _null_ ;
  set test4;
  where s is not null ;
  put _all_ ;
run;
```

特定パターンを条件とする場合

LIKE演算子では、「%」、「_」でパターンを作成可能です。「_」は任意の1文字、「%」は1文字以上の任意の文字列を表します。

• LIKE

```
data test5 ;
  s = 'AB11'; output;
  s = 'CD11'; output;
  s = 'EF12'; output;
run;
```



```
data _null_ ;
  set test5;
  where s not like 'EF__';
  put _all_ ;
run;
data _null_ ;
  set test5;
  where s like '%11';
  put _all_ ;
run;
```

Q バッチプログラムに複数の引数を渡す必要があります。良い方法はありますか？

A 通常、プログラムに引数を解釈させるためには SYSPARM オプションを利用します。ただし、SYSPARM オプションは一連の文字列をマクロ変数 SYSPARM に渡すため、引数が複数必要な場合、利用方法に工夫が必要です。SYSPARM オプション以外には、環境変数を利用する方法や引数を記述した専用ファイルの利用などが考えられます。パラメータが大量にある場合は専用ファイルなどの利用が効率的と思われるが、数個程度の引数の利用では環境変数の利用が手軽です。たとえば、環境変数を利用する場合、SET オプションと、SYSGET 関数を利用し、次のような連携が可能です。

実行コマンド

```
$ sas -set p1 This -set p2 is -set p3 TEST. program.sas
```

処理プログラム

```
data _null_ ;
  A = sysget('p1') ;
  B = sysget('p2') ;
  C = sysget('p3') ;
  put A B C ;
run ;
```

実行ログ

```
1      data _null_ ;
2          a = sysget('p1') ;
3          b = sysget('p2') ;
4          c = sysget('p3') ;
5          put a b c ;
6      run ;

This is TEST.
```

Q Enterprise GuideがインストールされているPCには、SAS システムがインストールされていません。存在しないLOCAL サーバーが選択されないように、リモートにあるSASサーバーを標準サーバーとして設定し、LOCALサーバーが表示されないようにするにはどうしたらよいでしょうか。

A Enterprise Guideでは、Administrator機能を利用して、標準サーバーの設定や、LOCALサーバーの設定が可能です。Enterprise Guideのリリースにより、若干手順が異なります。ご利用のリリース用の手順にて設定してください。

Enterprise Guide 1.2 での設定方法

1. Enterprise Administratorを起動する。
2. 画面左側[グループ]の項目で、該当のグループを選択する。
デフォルトでは[General]となる。
3. 画面左側のグループ名[General]を選択し、[ファイル] [プロパティ]を選択してプロパティ画面を表示する。
4. プロパティ画面の[サーバー]タブをクリックし、画面右側の選択済みリストから[LOCAL]を削除する。
5. 同じ画面で、左側の[すべて]の項目から、標準で使用するサーバーを選択して[追加]をクリックする。
6. 同じ画面で、右側に目的のサーバーが移動したら、該当のサーバーを選択して[標準]をクリックする。
このとき、標準で選択されたサーバー名は赤色で表示される。
7. Enterprise Administratorを終了する。

Enterprise Guide 2.0 での設定方法

LOCALサーバーの表示の削除、そして標準サーバーの設定の2段階の設定作業になります。

1. Enterprise Administratorを起動する。
2. [表示] [サーバー]を選択して、[接続なし]を選択する。
3. 画面右側のパネルから[LOCAL]サーバーを選択する。
4. [ファイル] [プロパティ]を選択し、[オプション]タブをクリックする。
5. [グループ単位にアクセスを制御する]をチェックして、[OK]をクリックする。
この操作で、LOCALを必要としないグループのユーザーにはLOCALが表示されなくなる。
6. [表示] [サーバー]を選択し、[標準]を選択する。
これでLOCALサーバーの表示が削除できる。
7. 続いて標準サーバーを設定する。まずEnterprise Guideを起動する。
8. [ツール] [アドミニストレータオプション]を選択する。
9. 画面の[標準 SAS サーバー]で標準に設定したいサーバーを選択し、[OK]をクリックする。

Q CLUSTERプロシジャを用いて、階層型のクラスター分析を行なっています。入力データセットは、座標データです。このとき、ユークリッド距離ではなくマハラノビス距離に基づいてクラスタリングを実行したいのですが、CLUSTERプロシジャにはそのためのオプションはないようです。何かよい方法はあるでしょうか？

A 座標データからなるデータセットをそのまま分析に用いた場合には、原則としてユークリッド距離に基づいて分析が行なわれます。マハラノビス距離に基づいたクラスター分析を実行する場合には、座標データに対して最初にPRINCOMPプロシジャを用いて主成分得点を計算し、その主成分得点を分析変数としてCLUSTERプロシジャを用いることにより、本質的に同じ分析を実行できます。なお、PRINCOMPプロシジャでは、あらかじめSTDオプションを指定しておく必要があります。以下にサンプルプログラムを示します。

```
/*サンプルデータの作成*/
data test;
  do id=1 to 100;
    x=rannor(12345);
    y=rannor(12345);
    z=rannor(12345);
    output;
  end;
run;

/*PRINCOMPプロシジャ*/
proc princomp data=test std out=out noprint;
  var x y z;
run;

proc cluster data=out method=ward outtree=tree;
  var prin; /*変数名がprinで始まるもののみを分析変数として用いる*/
  id id;
run;

proc tree data=tree;
run;
```

Q 生データはないのですが、各群における要約統計量は分かっています。理論上は要約統計量のみからもt検定は実行できますが、SASで同様に実行することは可能でしょうか。

A 各群における標本数、平均値、標準偏差(または分散)があらかじめ分かっている場合、生データを使用しなくてもt検定を実行することは可能です。具体的には、群を特定する変数(たとえば、group)、変数_STAT_(_STAT_)と要約統計量の変数(たとえば、summary)をデータセットに含める必要があります。ここで注意が必要なのは、変数_STAT_です。この変数名は必ず "_STAT_" と指定し、変数値をn、mean、stdとしてください。また、相当する要約統計量(順に標本数、平均値、標準偏差)を変数summaryにて入力してください。このデータセットに対して、以下のサンプルプログラムのようにTTESTプロシジャを実行することにより、t検定の結果を得ることができます。

```
data test; /* 要約統計量を含む擬似データ */
  input group _STAT_$ summary @@;
  cards;
1 n 10 1 mean 5.2235 1 std 0.8243
2 n 10 2 mean 9.7790 2 std 0.9980
;
run;

proc ttest data=test;
  class group;
  var summary;
run;
```

New Publication

新刊マニュアルのご紹介

「SAS言語:解説編 バージョン8」(日本語版)

注文番号:17375

価 格:15,000円

本書は、「SAS Language Reference: Concepts, Version 8」(英語版:注文番号 57375)の翻訳版で、バージョン8以降のSAS言語の仕様と概念について解説しています。既刊の「SASランゲージ リファレンス Version 6, First Edition」の「解説編」をバージョン8に対応させた内容となります。本書では、主に下記の項目について詳細に解説しています。

- ・ SASシステムの基本概要
- ・ SASシステムの処理
- ・ SAS言語における用語やSAS名の規則
- ・ SASデータセット
- ・ SASデータファイル
- ・ SASデータビュー
- ・ SASカタログ
- ・ SASデータライブラリ
- ・ SASデータセットオプション
- ・ SASシステムオプション
- ・ SASステートメント
- ・ SASフォーマット / インフォーマット
- ・ SAS変数とSAS CALLルーチン
- ・ SAS変数
- ・ 欠損値
- ・ SAS日付値 / 時間値 / 日時値と間隔
- ・ SAS出力
- ・ 生データの読み込み
- ・ DATAステップの処理
- ・ DATAステップ内でのBYグループ処理
- ・ SASデータセットの読み込み・結合・変更
- ・ WHERE式の処理
- ・ 配列処理
- ・ スタアドプログラムの作成と実行
- ・ DICTIONARYテーブル
- ・ DBMSデータへのアクセス
- ・ バージョン8と旧バージョンとの互換性
- ・ ファイル保護
- ・ SAS入出力エンジン
- ・ SASファイルの管理
- ・ 外部ファイル
- ・ 2バイト文字セット
- ・ 西暦2000年問題に関する注意点
- ・ エラー処理とデバッグ
- ・ システムパフォーマンスの最適化

SAS言語のそれぞれの言語要素の詳細は、「SAS言語:リファレンス編 バージョン8」(日本語版:注文番号 17239)を参照してください。なお、SAS 8.2導入パッケージに同梱されている「SASシステムV8e(リリース8.2)インストールシードキュメントメディア(CD-ROM)」には、リリース8.2を使用する上で必要な「SASシステム リリース8.1の新機能」、「SASシステム リリース8.2の新機能」の各PDF版、および本書と同内容のPDF版が含まれる予定です。

「SAS言語:リファレンス編 バージョン8」(日本語版)

注文番号:17239

価 格:28,000円

本書は、「SAS Language Reference: Dictionary, Version 8」(英語版:注文番号 57239)の翻訳版で、バージョン8以降のBase SASソフトウェアにおける、SAS言語のリファレンスです。既刊の「SASランゲージ リファレンス Version 6, First Edition」の「リファレンス編」をバージョン8に対応させた内容となります。本書には、SASデータセットオプション、SAS出力形式、SAS関数とSAS CALLルーチン、SAS入力形式、SASステートメント、SASシステムオプション、SAS言語におけるワードとSAS名に関する規則、SAS演算子、DATAステップデバッグに関する詳細な情報が含まれています。SAS言語の基本的な仕様や概念の詳細は、「SAS言語:解説編 バージョン8」(日本語版:注文番号 17375)を参照してください。なお、SAS 8.2導入パッケージに同梱されている「SASシステムV8e(リリース8.2)インストールシードキュメントメディア(CD-ROM)」には、リリース8.2を使用する上で必要な「SASシステム リリース8.1の新機能」、「SASシステム リリース8.2の新機能」、「SAS言語:解説編 バージョン8」の各PDF版が含まれています。本書と同内容のPDF版は次回送付予定「SASシステムV8e(リリース8.2)インストールシードキュメントメディア(CD-ROM)」に含まれる予定です。

「Data Mining Using SAS Enterprise Miner: A Case Study Approach, Second Edition」(英語版)

注文番号:59123

価 格:3,100円

データマイニングとは、データの中に隠された有用な情報や規則性を発見し、ビジネス上の問題点を解決に導く技術です。本書は、SAS Enterprise Minerを利用したデータマイニングについて、ケーススタディによって学習するための入門ガイドです。本書を通して、SAS Enterprise Minerを利用した予測モデル、クラスター分析、アソシエーション分析、そしてリンク分析といった数多くの複雑な解析の実際について学習できます。ケーススタディを多用しているため、データマイニングやSAS Enterprise Minerの初心者にも最適です。本書の課程を終えると、SAS Enterprise Minerのオンラインドキュメントに記載されている、より高度な統計解析を行えるようになります。

マニュアル販売係

T E L 03-3533-3835

F A X 03-3533-3781

E-mail JPNBooksale@sas.com

SASマニュアル注文用紙、および最新のPublication Catalog(マニュアル案内パンフレット)は弊社ホームページ(<http://www.sas.com/japan/manual/>)にて公開しておりますので、併せてご利用ください。

SAS Training

SASトレーニングのお知らせ

コース内容・料金改訂のお知らせ

10月より、SAS Institute Japan株式会社では、定期トレーニングの内容および料金を改訂します。詳細は弊社ホームページ (<http://www.sas.com/japan/training/>)をご参照ください。

特別コース開催のお知らせ

弊社では、目的・機能に特化した特別コースを随時開催しています。現在開催予定のコースは下記の通りです。

「項目反応理論入門」コース

日程:2003年9月8日(月)10:00~17:00 東京会場

価格:40,000円(テキスト代別/チケット捺印数2)

受講対象:1 項目反応モデルを利用したテスト開発を検討している方

2 数理モデルとしての項目反応モデルに興味のある方

前提知識:マウス操作などの基本的なコンピュータ操作が可能な方

担当講師:早稲田大学 文学部 心理学教室 豊田秀樹教授

学習内容:本セミナーでは項目反応理論を初めて勉強する方のための入門講座です。講義を主体としますが、受講者御自身が小さな心理検査を実際に体験し、SASのプログラムを使用して御自身で採点するなどの実習を通じて、項目反応理論に対する理解を深めていただきます。項目反応理論(IRT, item response theory)は項目応答理論とも呼ばれ、テストを作成・実施・評価・運用するための実践的な数理モデルです。項目反応理論は、米国はもとより、ヨーロッパの多くの国でもテスト理論のスタンダードとして不動の地位を築いています。中国や台湾など、アジア諸国の統一試験の運用にも使用されています。我が国の産業界も例外ではなく、留学のための語学試験や、入社のための筆記試験の一部は項目反応理論で運用されています。

テキスト:テキストとして、以下の書籍を使用します。お持ちでない場合は、お申し込み時にその旨お知らせください。

「項目反応理論[入門編]-テストと測定の科学-」

豊田秀樹著/朝倉書店発行

その他トレーニングに関する情報については、弊社ホームページ (<http://www.sas.com/japan/training>)をご参照いただくか、下記<トレーニング担当>までお問い合わせください。

トレーニング担当

TEL 03-3533-3835

FAX 03-3533-3781

E-mail JPNTraining@sas.com

Latest Releases

最新リリース情報

PCプラットフォーム

Windows版	SAS 8.2	TS2M0
----------	---------	-------

ミニコンピュータプラットフォーム

OpenVMS AXP版	SAS 6.12	TS020
--------------	----------	-------

OpenVMS VAX版	SAS 6.08	TS407
--------------	----------	-------

UNIXプラットフォーム

Tru64版	SAS 8.2	TS2M0
--------	---------	-------

ABI+版	SAS 6.11	TS040
-------	----------	-------

SunOS/Solaris版	SAS 8.2	TS2M0
----------------	---------	-------

HP-UX版	SAS 8.2	TS2M0
--------	---------	-------

AIX版	SAS 8.2	TS2M0
------	---------	-------

Linux版	SAS 8.2	TS2M0
--------	---------	-------

メインフレームプラットフォーム

IBM版(MVS)	SAS 8.2	TS2M0
-----------	---------	-------

富士通版(F4, MSP)	SAS 6.09E	TS470
---------------	-----------	-------

日立版(VOS3)	SAS 6.09E	TS470
-----------	-----------	-------

CMS版	SAS 6.08	TS410
------	----------	-------

SAS Technical News Summer 2003

発行
SAS Institute Japan株式会社

テクニカルニュースに関するお問い合わせ先

テクニカルサポートグループ

TEL:03-3533-3877

FAX:03-3533-3781

E-mail: technews@jpn.sas.com



SAS Institute Japan株式会社

www.sas.com/japan/

東京本社
〒104-0054
東京都中央区勝どき1-13-1
イヌイビル・カチドキ 8F
Tel 03 (3533) 6921
Fax 03 (3533) 6927

大阪支店
〒530-0004
大阪市北区堂島浜1-4-16
アークア堂島西館 12F
Tel 06 (6345) 5700
Fax 06 (6345) 5655

このカタログに記載された内容は改良のため、予告なく仕様・性能を変更する場合があります。あらかじめご了承ください。

SASロゴ、The Power to Knowは米国SAS Institute Inc.の登録商標です。その他記載のブランド、商品名は、一般に各社の登録商標です。Copyright©2003, SAS Institute Inc. All rights reserved.