

SAS Technical News

Winter 2004

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

CONTENTS

- 1 SAS® Enterprise Guide® 2.0による分散分析
- 6 Q&A
- 12 SASトレーニングのお知らせ
- 15 新刊マニュアルのご紹介
- 15 最新リリース情報

特集

SAS® Enterprise Guide® 2.0による分散分析

1. はじめに

この特集では、SAS Technical News 2003年夏号の特集記事「SAS Enterprise Guide 2.0による記述統計の活用」に引き続き、SAS Enterprise Guide 2.0(以下EG 2.0)で利用できる統計手法を紹介します。今回は、複数の平均値を比較・検討する手法に関して説明します。この手法は、たとえば男女間の身長、体重に差があるかを調べる場合などに活用できます。一般的には分散分析と呼ばれ、複雑な実験計画に対しても分析を行なうことができますが、ここでは簡単なモデルを用いてEG 2.0における具体的な指定・選択を交えて説明します。

サンプルデータとしてCLASSデータとFITNESSデータを用います。これらのデータは、EG 2.0の[バインダリスト]ウィンドウから[Sample]を展開して参照できます。CLASSデータには、19人の名前、年齢、性別、身長、体重が記録されています。また、FITNESSデータには多くの変数が含まれていますが、今回は酸素の摂取量を表す応答変数(Oxygen)と、各被験者が行なったエクササイズを表す変数(Group)を使用します。

2. 平均値に対する検定、および2平均値の比較

1つの母集団を仮定している場合、観測データから算出された平均値、要約統計量を元に、母集団における平均値が検定値(任意の数値)であるかを検討します。また、2つの母集団を仮定している場合には、母集団において平均値が等しいかを検討します。具体例としてCLASSデータにおける身長(Height)を用い、EG 2.0で検討する方法を紹介します。この他、データの観測値間に対応がある場合には、対応があることを考慮した上で、平均値に対する検討を行なうことがより適した手法となります。

2.1 平均値に対する検定

CLASSデータにおける身長(Height)の平均値は、EG 2.0のメニューから [分析] [記述統計量] を選択することにより、62.3インチ(約158cm)と表示されます。仮に、母集団(注1)における平均値が60インチ(約152cm)であるとされている場合、観測データから母集団における平均値が本当に60インチであるかを検証します。EG 2.0では以下のように選択してください。

注1 たとえば、ある中学校Aに在籍する生徒から19人を選択し、データを収集した場合には、母集団は中学校Aに在籍する全ての生徒となります。

1 標本に対する検定

- 1) メニューから [分析] [分散分析] [t検定] を選択する。
- 2) [t検定の種類] タブをクリックし、[1 標本に対する検定] を選択する。
- 3) [変数] タブをクリックし、分析変数を選択する。
ここでは、分析変数として変数Heightを選択します。
- 4) [分析] タブをクリックし、検定値(任意の数値)を「帰無仮説」にて指定する。ここでは、身長が60インチであるかどうかを検討するので、検定値として60を入力します。また、信頼水準(デフォルトでは95%)の変更、標準偏差に対して信頼区間を求める際の手法も選択できます。
- 5) [完了] をクリックする。



図1:平均値に対する検定結果の一部

図1では、変数Heightの平均値(Mean)が62.337インチ、標準偏差(Std Dev)が5.1271インチであることが表示されています。これらの要約統計量を元に、t検定における統計量(t Value)が1.99、対応するp値(Pr > |t|)が0.0624と算出されています。事前に有意水準値(α)を0.05と定めている場合には、算出されたp値が0.05以上ですので、母集団における身長の平均値が60インチであるという仮説を棄却することができません。つまり、身長が60インチであるという仮説を否定するには十分な確証がないと解釈します。

2.2 2平均値の比較

上記では、身長(Height)の平均値が60インチであるかどうかを検討しました。しかし、男子と女子における身長差の有無についてはわかりません。男子と女子の(標本における)身長の平均値は、それぞれ63.9インチ(約162cm)、60.6インチ(約154cm)ですが、母集団において男子と女子の身長に差があるかを検証するには、以下のように選択してください。

2 標本に対する検定

- 1) メニューから [分析] [分散分析] [t検定] を選択する。
- 2) [t検定の種類] タブをクリックし、[2 標本に対する検定] を選択する。
- 3) [変数] タブをクリックし、分析変数と、グループ化する基準を選択する。
ここでは、分析変数として変数Heightを選択し、グループ化する基準として変数Sexを指定します。

- 4) [分析] タブをクリックし、検定値(任意の数値)を帰無仮説にて指定する。男女間における平均値が同じであるかどうかを検討するので、検定値を0(デフォルト)に指定します。
- 5) [プロット] タブをクリックし、箱ひげ図または平均のプロットを選択する。平均のプロットを選択した場合には、プロットに関するオプションを選択できます。
- 6) [完了] をクリックする。



図2:2平均値の比較結果の一部

図2では、2つのt検定の結果が出力されています。最初のt検定は、男子と女子における分散が等しいことを仮定していますが、2目のt検定では分散が等しいことを仮定していません。どちらの統計量を用いるかは意見が分かれるところですが、t検定の種類を選択する時のひとつの指標として、等分散性の検定であるFolded F統計量を使用できます。ここでは、Folded F統計量から、分散が等しいと仮定(注2)し、最初のt検定によるp値を参照します。また、このp値0.1645が0.05(あらかじめ定めた有意水準値)以上なので、母集団の平均値に差がないということを棄却できません。つまり、CLASSデータには、男女の身長に差があるというには十分な確証がないと解釈します。

注2 正確には、Folded F統計量に対するp値(Pr>F)が大きい値(0.9527)なので、分散が等しいという帰無仮説を棄却することができないと解釈します。

2.3 対応のある観測値に対する検定

対応のあるデータ的具体例として、薬剤の投与前後における観測値が挙げられます。たとえば、投与前、投与後の観測値をそれぞれ変数Before、Afterとした場合、以下のように選択してください。

対応のある検定

- 1) メニューから [分析] [分散分析] [t検定] を選択する。
- 2) [t検定の種類] タブをクリックし、[対応のある検定] を選択する。
- 3) [変数] タブをクリックし、対応のある変数を2つ選択する。
対応のある変数として変数BeforeとAfterを指定します。
- 4) [分析] タブをクリックし、検定値(任意の数値)を「帰無仮説」にて指定する。2つの変数間における差異が検定値であるかを検討する場合に指定します。
- 5) [完了] をクリックする。



図3:対応のある検定結果(サンプル出力)

図3では、変数Beforeを変数Afterから引いた値の平均値2.6908が、その他の要約統計量とともに表示されています。2つ目の表では、対応のあるt検定の統計量(t Value)が11.21、p値(Pr > |t|)が0.0001より小さいと算出されています。そのため、母集団において変数Afterから変数Beforeを引いた値が0より大きいと解釈できます。

3. 一元配置分散分析

「2 平均値に対する検定、および2平均値の比較」では、1つまたは2つの平均値に対する検定を紹介しましたが、この節では3つ以上の平均値の比較に関して説明します。ここでは、データにおける標本を区別するための変数を要因とし、その変数の取り得る値を水準とします。CLASSデータの場合、要因は変数Sexであり、水準はF(女子)とM(男子)となります。

要因における水準値が2つである場合には、t検定で平均値が同じであるかを検証できます。水準の数が3つ以上である場合には、t検定を複数繰り返すことにより、各水準における平均値が同じであるかを検討することができます。しかし、有意水準が100%であるt検定を複数繰り返した場合、全体の検定としての有意水準が100%とは異なってしまいます。つまり、母集団における差がないにも関わらず差があると判断する過誤をしやすくなります。そのため、一度に平均値が全て同じであるかを検討する手法として、分散分析が用いられます。また、t検定を繰り返す場合にも全体の検定としての有意水準が100%となるように、調整する方法(多重比較法)があります。

3.1 1要因に対する分散分析

FITNESSデータでは、被験者を3つのグループ(変数Group)に分け、グループごとに異なるエクササイズを行ない、酸素の摂取量(変数Oxygen)を観測しています。ここでは、酸素の摂取量がグループによって異なるかを検証するために、以下のように選択してください。

一元配置分散分析

- 1)メニューから [分析] [分散分析] [一元配置分散分析] を選択する。
- 2) [変数] タブをクリックし、従属変数と独立変数(1つのみ)を選択する。
ここでは、従属変数として変数Oxygen、独立変数として変数Groupを指定します。
- 3) [検定] タブをクリックし、Welch検定(注3)を行なうかを選択する。
また、等分散性の検定を行なう場合には、検定方法を選択します。
- 4) [平均値] タブをクリックし、平均値に対するオプションを指定する。
 - a) [詳細] をクリックし、出力する要約統計量を選択します。
 - b) 「比較」に関しては、「3.2 多重比較法」を参照してください。
- 5) [プロット] タブをクリックし、箱ひげ図または平均のプロットを選択する。
平均のプロットを選択した場合には、プロットに関するオプションを選択できます。

6) [完了] をクリックする。

注3 一元配置分散分析では等分散性を仮定していますが、Welch検定ではこの仮定をせずに検定を行ないます。



図4:平均値の比較結果の一部

t検定では、検定統計量としてt値(t Value)が算出されましたが、一元配置分散分析(図4参照)では、F値(F Value)が検定統計量として用いられます。ここでは、F値(F Value)が0.94、対応するp値(Pr>F)が0.4040ですので、有意水準(たとえば0.05)と比較した場合、要因の水準における全ての平均値が同じであるという仮説を棄却することができません。つまり、FITNESSデータの場合、グループ間において酸素の摂取量が異なるというには統計的な確証がないと解釈します。一元配置分散分析の場合、1つの検定統計量にて全ての水準における平均値が同じであるかを検証することができます。検定の結果、算出されたp値が小さい場合には、少なくとも1つの水準における平均値が異なることは分かりますが、どの水準における平均値が異なるかに関しては分かりません。このような場合、ステップ5)の[プロット]タブでグラフを作成することにより、各水準での平均値の推移を視覚化することができます。(図5参照)

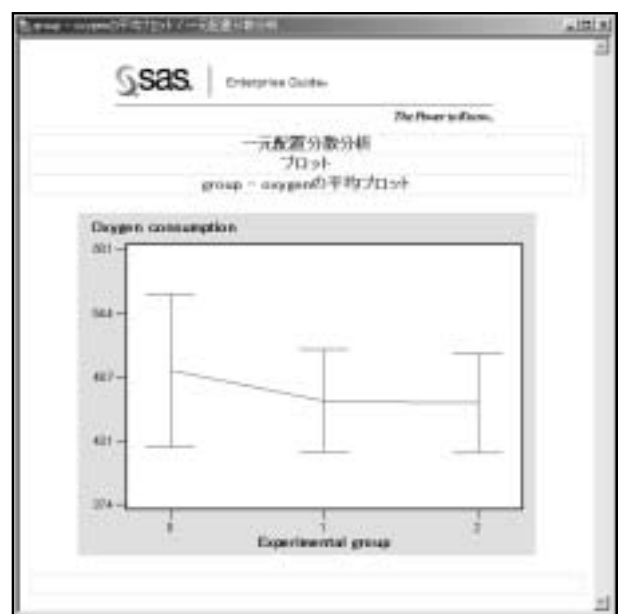


図5:平均値のプロット

3.2 多重比較法

「3.1 1要因に対する分散分析」では、各水準における平均値が同じであるかを検討しました。しかし、他の手法として水準を2つずつ選択し、比較を複数回行うことによって同じように全ての平均値が同じであるかを検討することができます。一般的には多重比較法と呼ばれる手法です。しかし、全体の検定としての有意水準を一定の数値(100%)にするために、調整が必要になります。EG 2.0における[一元配置分散分析]では、分析目的、データ等に応じたさまざまな調整法を選択できます。手法の選択は、「3.1 1要因に対する分散分析」で説明した[一元配置分散分析]におけるステップの4)-b)で可能です。



図6:多重比較法の選択画面

図6では、多重比較法として「Tukeyの多重比較」を選択しています。また、図6の右側にもありますように有意水準()を指定することができます。(デフォルトの値は0.05となります。)

group Comparison	Difference Between Means	Simultaneous 95% Confidence Limits
0 - 1	2.711	-2.197 8.618
0 - 2	2.886	-2.885 8.858
1 - 0	-2.711	-8.618 3.197
1 - 2	0.176	-5.596 5.947
2 - 0	-2.886	-8.858 3.085
2 - 1	-0.176	-5.947 5.596

図7:Tukey法に基づく多重比較の結果

図7では、水準の各組み合わせにおける平均値の差とともに、信頼区間が表示されています。信頼区間に0が含まれている場合には、対応する水準間には差があるというには十分な確証がないとして解釈します。図7では、全ての比較に対する信頼区間に0が含まれているので、全ての水準間において差があるとはいえないと解釈できます。

3.3 分散分析と多重比較法に関して

分散分析では、水準における平均値が全て同じであるかを1つの統計量に要約し、検定できます。しかし、少なくとも1つの水準における平均値が他とは異なると判断された場合、どの水準が違うかは分かりません。一方、多重比較法では各ペアにおける検定を行なっていますので、どの平均値が異なるかを検討することができます。分散分析と多重比較法では、同じ問題に対して別々の手法を用いているため、必ずしも同じ結果が得られるとは限りません。分散分析では少なくとも1つの水準における平均値が他とは異なる結果が得られている場合にも、多重比較法では全ての平均値が同じであるということもありえます。分析目的等を考慮した上で、適した手法を活用してください。

4. ノンパラメトリックな一元配置分散分析

[一元配置分散分析]における検定は、応答変数が正規分布に従うことを仮定しているため、データによっては分析手法として適切でない場合があります。この正規性を仮定しない分析手法として、ノンパラメトリックな手法があります。これらの手法では、観測値そのものではなく観測値をスコアに変換して分析を行ないます。スコアへの変換方法としてさまざまなものがあり、EG 2.0では多くの変換方法を選択できます。具体例として、FITNESSデータに対してノンパラメトリック手法の1つであるWilcoxon検定を行なうには以下のように選択してください。

ノンパラメトリックな一元配置分散分析

- 1) メニューから[分析] [分散分析] [ノンパラメトリックな一元配置分散分析]を選択する。
- 2) [変数]タブをクリックし、従属変数と独立変数(1つのみ)を選択する。ここでは、従属変数として変数Oxygen、独立変数として変数Groupを指定します。
- 3) [分析]タブをクリックし、検定に用いるスコアを選択する。ここでは、Wilcoxonスコアを用いるので、Wilcoxonを選択します。水準数が2つである場合には、2標本Wilcoxon検定が算出されますが、3つ以上の水準がある場合には、Kruskal-Wallis 検定となります。
- 4) [正確なp値]タブをクリックし、[分析]タブで選択したスコアに対して正確なp値を算出する場合には、該当する箇所を選択する。



図8:正確なp値の選択画面

図8の右側の注釈にあるように、正確なp値の算出には多くの時間とメモリを使用しますので、必要な場合のみ選択してください。

- 5) [完了]をクリックする。



図9: Wilcoxonスコアを選択した際の出力結果

FITNESSデータでは、変数Groupにおける水準数が3つあるので、Kruskal-Wallis検定の結果が図9に表示されています。ここでは、検定統計量であるカイ2乗値(Chi-Square)として0.3640、該当する漸近的なp値(Pr>Chi-Square)が0.8336と算出されています。「3.1 1要因に対する分散分析」と同じように、水準間における差があるとはいえないと解釈することができます。

5. 複数の要因を含めた分散分析

ここまでは、1つの要因がモデルに含まれている場合の分析手法に関して説明しました。しかし、EG 2.0では複数の要因をモデルに含めた分析もできます。これらの分析を行なうには、メニューから[分析] [分散分析]をクリックし、[線形モデル]を選択してください。[変数]タブにてモデルに含める要因の変数(量的変数、分類変数など)、従属変数を指定した後、[モデルの定義]タブにてモデルを構成することができます。(図10参照) また、[Post-Hoc検定]タブをクリックすることにより、各水準における平均値を算出する要因を指定し、多重比較の手法も含め平均値の算出に関するオプションを選択することができます。[線形モデル]と[混合モデル]における個々のタブに関する詳細に関しては、EG 2.0のメニューから、[ヘルプ] [Enterprise Guideヘルプ]をクリックし、[Enterprise Guideタスクの実行] [分散分析]における記述を参照してください。



図10: [モデルの定義]タブにてモデルの構成を指定

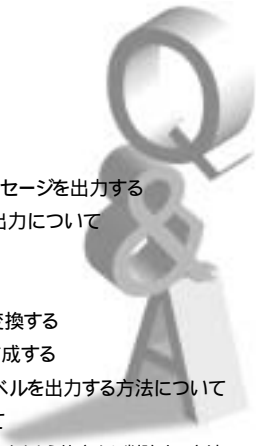
6. おわりに

今回の特集では、データ分析の一手法である分散分析をEG 2.0を使用して行なう方法についてご紹介しました。特に一要因をモデルに含む場合に関して説明をしましたが、「5. 複数の要因を含めた分散分析」で解説したように、より複雑なモデルの場合にも、ポイント&クリック操作で分析を行なうことができます。本特集における内容がデータ分析をする際の一手法として活用していただければ幸いです。

Enterprise
Guide 2.0



Q&A



データが0件のときOUTPUTウィンドウにメッセージを出力する
Microsoft Excelへのデータの複数シート出力について
データセットのサイズを見積もる方法
ログ出力のスクロール頻度を変更
「Jan-03」のようなデータをSAS日付値に変換する
SAS/GRAPHソフトウェアにて透過GIFを作成する
Microsoft Excelに変数名ではなく、変数ラベルを出力する方法について
文字変数値の一部を抽出する方法について
2つのSASデータセットのうち、片方に存在するデータをもう片方から削除する方法
SAS/AFのカスタムクラスの編集について
MULTTESTプロシジャ実行時のNOTE出力の意味について
LOGISTICプロシジャにおけるカテゴリカル変数のコード化について
SASのバージョン表記について

Q OUTPUTウィンドウへの出力があるプログラムを実行するとき、データが0件だと何も出力されませんが、任意のメッセージを出力するようにできますか？

A 次のようなプログラムでメッセージを出力できます。下記のサンプルプログラムでは、データのあるデータセットと0件のデータセットを作成し、データが0件のデータセットを処理した際にメッセージを出力するマクロを作成しています。

```

                                /* サンプルデータセット1 作成*/
DATA one;
  x=1;
RUN;

                                /* サンプルデータセット2(データ0件) 作成*/
DATA two;
  STOP;
RUN;

OPTIONS MPRINT;

%MACRO drive(dsn);

                                /* オブザベーション数のカウント */
DATA _null_;
  CALL SYMPUT('cnt',PUT(cnt,5.));
  SET &dsn NOBS=cnt;
RUN;

                                /* データがあるとき */
%IF &cnt NE 0 %THEN %DO;
  PROC PRINT DATA=&dsn;
    TITLE "データセット名 : &dsn";
  run;
%END;
%ELSE %DO;
  DATA _null_;
    TITLE;
    FILE PRINT;

```

```

PUT _page_;
PUT "データセット名: &dsn";
PUT "データセットにオブザベーションがありません。";

RUN;

%END;

%MEND drive;

%drive(one)
%drive(two)

```

Q 複数のSASデータセットを、1つのExcelファイル(ブック)内で複数のシートに分けて出力する方法を教えてください。

A EXPORTプロシジャでは、同一ブック内の複数シートへの出力はできません。次の2通りの方法で出力する方法があります。

DDE の機能(Base SAS)を使用してExcelシートに書き込む方法
SAS/ACCESS to ODBCを用いてExcelシートに書き込む方法

具体的な説明とプログラム例をご紹介します。

DDE の機能を使用してExcelシートに書き込む方法
DDEは、Dynamic Data Exchange の略で、Microsoft社が提供しているWindowsアプリケーション間でのデータ交換に広く使用されているプロトコルです(DDEの詳細については、Microsoft社提供、市販されている書籍などでご確認ください)。SASでは、このDDEプロトコルを利用して、Excelシートにデータを出力することが可能です(Microsoft ExcelがSASと同じPCにインストールされていることが必要です)。以下のプログラム例では、2つのシートに出力していますが、この部分を適宜変更することにより、さらに複数のシートに対して書き込むことができます。

```

OPTIONS NOXWAIT NOXSYNC;
X "start excel";

                                /* MS Excel を起動 */
DATA _null_;
  x=SLEEP(10);
RUN;

                                /* Excelの起動を10秒待機 */
                                /* 待機時間は環境に応じて変えてください */
FILENAME ddedata DDE 'excel|sheet1|rlc1:r100c3';
                                /* DDEを利用したエクセル出力 */
                                /* シート1枚目 */
                                /* 乱数を使用して、サンプルを作成します */
DATA _null_;
  FILE ddedata;
  DO i = 1 TO 100;
    x = RANUNI(i);
    y = 10 + x;
    z = x - 10;
    PUT x y z;
  END;

```

```

RUN;

                                /* Excelへの書き込みを待機 */
DATA _null_;
  x=SLEEP(10);
RUN;

FILENAME ddedata2 DDE 'excel|sheet2!r1c1:r100c3';
                                /* シート2枚目 */
                                /* 乱数を使用して、サンプルを作成します */
DATA _null_;
  FILE ddedata2;
  DO i = 1 TO 100;
    x = RANUNI(i);
    y = 20 + x;
    z = x - 30;
    PUT x y z;
  END;
RUN;
    
```

SAS/ACCESS to ODBCを用いてExcelシートに書き込む方法

- 1) データソースとしてシステムまたはユーザーDSNを事前に定義します。
 データソースの定義は、たとえばWindows 2000の場合は、[スタート] [設定] [コントロールパネル] [管理ツール] [データソース(ODBC)]の順で画面を開き、データソースを定義します。なお、詳細の定義方法につきましては、Microsoft社提供のヘルプ等をご参照ください。
- 2) ライブラリを割り当てます。

```

LIBNAME mylib ODBC DSN=Excel;
                                /*データソース名をExcel とします。*/
                                /*上記データソースの設定で定義した名前を指定してください。*/
    
```

- 3) 複数のシート別に出力します。下記の例では、sashelp.classのデータを利用しています。

```

DATA mylib.sheet1a;
  SET sashelp.class;
RUN;
                                /* 新しくsheet1a を作成 */
DATA mylib.sheet2a;
  SET sashelp.class;
RUN;
                                /* 新しくsheet2a を作成 */
    
```

- 4) SASを終了せずに作成したExcelファイルをExcelから利用するには、暗黙的に適用されたロックを解除するために、ライブラリの割り当てを開放する必要があります。

```
LIBNAME mylib CLEAR;
```

なお、Excelへの複数シート出力に関するサンプルのマクロを、下記のURLで公開しています。あわせてご参照ください。

<http://support.sas.com/techsup/unotes/SN/006/006695.html>

Q

新規に作成するデータセットに必要なディスク容量を推定するために、データセットのサイズを見積もる計算式はありますか。

A

データセットのサイズは、下記の計算式で概算することができます。

$$\text{データセットのページ数} = \text{切り上げ} \left(\frac{\text{オブザベーション数}}{\text{ページごとの最大オブザベーション数}} \right)$$

$$\text{データセットのサイズ} = \text{ヘッダー部分のサイズ} + (\text{データセットのページ数} \times \text{ページサイズ})$$

新規に作成するデータセットのサイズを見積もる場合には、まず必要な変数を含む1オブザベーションのみのダミーのデータセットを作成します。そしてこのデータセットに対しCONTENTSプロシジャを実行します。

```

                                /* サンプルデータセットを作成 */
DATA test ;
  INPUT a b c d e ;
CARDS;
  1 2 3 4 5
;
RUN;
                                /* CONTENTSプロシジャを実行 */
PROC CONTENTS DATA=test ;
RUN;
    
```

CONTENTSプロシジャの出力結果から、「ページごとの最大オブザベーション数」と「ページサイズ」が分かります。以下はCONTENTSプロシジャの出力結果の一部を抜粋したものです。

```

-----エンジン / ホスト関連情報 -----

データセットのページサイズ: 4096
データセットのページ数: 1
データページの先頭: 1
ページごとの最大 OBS 数: 101
.
.
    
```

データセットの「ヘッダー部分のサイズ」は、Windows版では256バイト、UNIX版では8K(=8×1024)バイトです。これらの数値と想定するオブザベーション数を、最初に示した計算式に代入すると、新規に作成するデータセットのサイズを概算できます。Windows版SASで、上記サンプルデータセットTESTと形式が同様でオブザベーション数が1000と仮定したデータセットのおよそのサイズは、次ページのようなプログラムで算出できます。

```

/* データセットのサイズを計算 */
DATA _null_ ;
  pagenum = 1 + FLOOR(1000/101) ;
  datasize = 256 + (pagenum*4096) ;
  PUT datasize= ;
RUN;
    
```

ログにdatasize=41216 と表示されます。計算したデータセットのおよそのサイズが41,216バイトくらいになることを表しています。

Q

ログウィンドウへの出力の際、自動スクロール頻度を変更することはできますか？

A

コマンドバーまたはコマンド行でAUTOSCROLLコマンドを実行することで、スクロールの頻度を変更することが可能です。コマンドの構文は、以下のとおりです。

```

AUTOSCROLL n
    
```

nには、受け取ったデータ行がウィンドウ内に収まらないときにスクロールする行数を指定します。自動スクロール頻度を少なくすれば、プログラムの実行速度が向上します。nの値を0に設定すると自動スクロールを行わず、処理終了後に一括してスクロールされます。また、nを1に設定すると1行づつスクロールされるため、処理時間は長くなりますが、実行中のプログラムの状況をログウィンドウで確かめることができます。次の例では、DMコマンドを使用して、プログラム上からAUTOSCROLLコマンドを発行しています。

```

DM LOG 'AUTOSCROLL 0'; /* AUTOSCROLLを0に設定します */

DATA _null_;
  DO i=1 TO 10000;
    PUT i 5. +10 i WORDS50.;
  END;
RUN;
    
```

同様の処理をデフォルトの状態で行なった場合の処理時間と、AUTOSCROLLコマンドを発行した場合の実行結果を比較すると、下記のようになります。

デフォルトの場合

```

NOTE: DATA ステートメント 処理 :
      処理時間          1.92 秒
      CPU時間           1.92 秒
    
```

AUTOSCROLL=0を発行した場合

```

NOTE: DATA ステートメント 処理 :
      処理時間          0.04 秒
      CPU時間           0.04 秒
    
```

注1 AUTOSCROLLのデフォルト値は、オペレーティングシステムにより異なります。
注2 処理時間は参考値です。お使いの環境により異なる場合があります。

UNIX版 SAS 8.2	1
Windows版 SAS 8.2	LOGウィンドウのデフォルトサイズ(行数)の1/2の値(LOGウィンドウのデフォルトサイズはディスプレイの解像度により異なる)。実際の値は、メニューバーで [ツール] [オプション] [プリファレンス] の [詳細] タブにて、[ログ] の項をご確認ください。

Q

「Jan-03」(2003年1月)のように、日付の値がないデータをSAS日付値として読み込みたいと考えています。この場合、値を文字列として読み込み、「01-Jan-03」のようにデータを加工した後、INPUT関数を使用しなければならないでしょうか。

A

MONYYインフォーマットを利用することで、「Jan-03」のようなデータを、直接SAS日付値として読み込むことが可能です。下記に例を示します。

```

DATA sample;
  INPUT date monyy6.; /* monyyインフォーマットを利用 */
  FORMAT date yymmdd10.;
  PUT date;
CARDS;
Jan-03
Feb-03
Mar-03
Apr-03
May-03
;
RUN;
    
```

上記プログラムを実行した際の、SASログへの出力結果は次のとおりです。

```

2003-01-01
2003-02-01
2003-03-01
2003-04-01
2003-05-01
    
```

ログへの出力結果で確認できるように、日付の値がないデータを読み込む場合、SASはその値を1日とします。

Q

SAS/GRAPHで、透過GIFフォーマットの画像ファイルを作成したいと考えています。このようなことはできますか？

A

SAS System 8から、TRANSPARENCYグラフィックオプションが追加され、SAS/GRAPHで透過GIFフォーマットの画像ファイルを作成できるようになりました。次にサンプルプログラムを示します。

```
GOPTIONS RESET=all;
FILENAME gsasfile 'C:\%temp%\test.gif';
GOPTIONS DEVICE=gif GSFNAME=gsasfile
GSFMODE=replace;
GOPTIONS TRANSPARENCY;          /* 透過GIFを作成 */

TITLE F=kanji '性別毎の身長合計';
PROC GCHART DATA=sashelp.class;
    VBAR sex / SUMVAR=height;
RUN;
QUIT;
TITLE;
```

なお、SAS/GRAPHで透過的な出力を作成する場合、GIF以外のデバイス名はサポートされていないのでご注意ください。

Q

SASデータセットをExcelシートにエクスポートする際、1行目に変数名ではなくラベルを出力することはできますか？

A

Windows版SAS System 8でSAS/ACCESS to PC File Formatsを利用している場合、Excel形式への出力にはEXPORTプロシジャを使用できますが、このプロシジャには変数名の代わりに変数ラベルを出力する機能はありません。このとき、変数ラベルを変数名に変換する操作を実施した後で、EXPORTプロシジャを利用する方法が考えられます。

変数名を変更するにはいくつかの方法が考えられますが、次の例ではビューを利用して出力用にデータセットを作成する方法をご紹介します。なお、変数名に日本語や空白文字を利用する場合、VALIDVERNAME=ANYオプション(評価版)を利用することで表現可能です。この際の変数名の指定方法は、下記のとおりです。

'変数名'n

変数ラベルが32バイトを超える場合、そのまま変数名には利用できず33バイト以降が切り捨てられます。このような場合は、適切な修正が必要となります。

/* 変数ラベルを変数名にしたデータセットを作成するマクロ */

```
%MACRO change(libnm,dsnm,odsn);
PROC SQL NOPRINT;
    SELECT COUNT(name) INTO :nobs
    FROM DICTIONARY.COLUMNS
    WHERE LIBNAME=UPCASE("&libnm") AND
MEMNAME=UPCASE("&dsnm");
    SELECT NAME INTO :name1-:name%LEFT(&nobs)
    FROM DICTIONARY.COLUMNS
    WHERE LIBNAME=UPCASE("&libnm") AND
MEMNAME=UPCASE("&dsnm");
    SELECT LABEL INTO :label1-:label%LEFT(&nobs)
    FROM DICTIONARY.COLUMNS
    WHERE LIBNAME=UPCASE("&libnm") AND
MEMNAME=UPCASE("&dsnm");
QUIT;

DATA &odsn / VIEW=&odsn;
SET &libnm..&dsnm(RENAME=(
%DO I=1 %TO &nobs;
    %IF "&label&I" NE "" %THEN %DO;
        &&name&I = "&label&I"N
    %END;
%END;
));
RUN;
%MEND;

/* 利用例 */
/* サンプルデータ */
/* 日本語のラベルがあるデータセット */

DATA test ;
    LABEL x='SAS Institute Japan' y='テクニカルサポート';
    x=1; y=2; z=3;
RUN;

/* オプションを有効にする */

OPTIONS VALIDVARNAME=ANY;

/* WORK.TESTから変数ラベルを変数名に変換した */
/* データセットWORK.TEST2を作成する */
%change(work, test, test2);

/* Excel形式ファイルにエクスポートする */
PROC EXPORT DATA=test2 DBMS=Excel FILE='c:\%temp%\test.xls' REPLACE;
RUN;

/* オプションをデフォルトに戻す */
OPTIONS VALIDVARNAME=V7;
```

Q 下記のようなデータ変換を行なうには、どのようにすればよいでしょうか。

元のデータの内容

変換後のデータの内容

```
男性(20才)
男性(30才)
女性(20才)
男性(40才)
```

```
男性
男性
女性
男性
```

A いくつかの方法が考えられますが、SCAN関数を利用し、「(」の前までの文字列を抽出する方法をご紹介します。SCAN関数は区切り文字で値を区切り、N番目の語を抽出します。また、2バイト文字を使用するには、KSCAN関数を使用します。次ページのようなプログラムでデータの抽出が可能です。

```
A test;
  INPUT dat $20.;
CARDS;
男性(20才)
男性(30才)
女性(20才)
男性(40才)
;
RUN;

DATA results;
  SET test;
  /* 全角 '(' で値を区切り、1番目の語を抽出 */
  dat=KSCAN(dat,1,'(');
  PUT dat=;
RUN;
```

Q 2つのSASデータセットがあります。片方のデータセットに含まれるレコード(オブザベーション)を、もう1つのSASデータセットから削除したいと思います。何かよい方法はありますか。

A SQLプロシジャのEXCEPT演算子を使用すると可能です。下記のサンプルプログラムは、TRANSデータセットに存在するオブザベーションをMASTERデータセットから削除します。

```
DDATA master;
  INPUT id name $;
CARDS;
1 A
2 B
3 C
4 D
5 E
6 F
7 G
```

```
;
RUN;

DATA trans;
  INPUT id name $;
CARDS;
2 B
3 C
5 E
;
RUN;

PROC SQL;
  CREATE TABLE master AS
  SELECT * FROM master
  EXCEPT
  SELECT * FROM trans
;
QUIT;
```

Q SAS/AFで、各画面で共通に使用するボタン等をクラスとして作成保存し、各画面でそのクラスを利用しています。作成済みのクラスの属性を変更したいのですが、どのようにすればよいでしょうか?

A 作成済みクラスの属性は、下記の手順で変更できます。

- 1) SASのエクスプローラウィンドウを表示する(表示されていない場合は、[表示] [エクスプローラ]で表示する)
- 2) 作成したCLASSエントリが保存されている該当のライブラリから[カタログ]をたどり、作成したCLASSエントリをダブルクリックして開く(または、マウスで選択後に右クリックメニューで[開く]を選択する)。
- 3) [クラスエディタ]で、該当のエントリに対する編集画面が起動するので、画面左側の[クラスプロパティ]から、[Attributes] [属性]などを選択する。
- 4) 画面右側に表示される各編集セルにある、[componentDefinition]項目の[初期値]を選択すると、枠内に[...]とボタンが表示されるので、そのボタンをクリックすると、クラスとして保存したフレームの作成時の画面が編集モードで開く。
- 5) 値を変更して上書き保存する。

Q MULTTESTプロシジャを利用して標本再抽出に基づくp値の調整を行なった際、以下のようなNOTEがログに出力されることがあります。これは、どのような意味でしょうか?

```
NOTE:The multiple testing procedure for this run
is not closed.In cases with badly heteroskedastic
data, tests for individual null hypotheses can
have inflated familywise Type I error rates.
```

A MULTTEST プロシジャにおける対比の指定がいわゆる「閉手順」の状況になっておらず、同時にBOOTSTRAP、またはPERMUTATIONオプションを指定して標本再抽出に基づく多重比較を行なうと、適切でない現象、すなわち「ファミリーワイズのTYPE I エラー」をコントロールできない場合があるとの研究があります。SAS System 8以降のMULTTEST プロシジャでは、このような場合にNOTEとしてログに出力するようになりました。「ファミリーワイズのTYPE I エラー」は、多重比較法を適用した際の「第1種の過誤の確率」に相当するものであり、この数値が有意水準以下にコントロールされる方法が適切な多重比較法と言われています。米国SAS社のWebサイトでは、「Resampling-Based Multiple Testing」の著者であるWestfallと、MULTTEST プロシジャの開発者であったWolfigerによる論文が公開されており、この問題について言及しています。

「Closed Multiple Testing Procedures and PROC MULTTEST」
Peter H. Westfall and Russell D. Wolfiger
URL://support.sas.com/documentation/periodicals/obs/obswww23/

また、下記の書籍においても「閉手順」に関する理論が紹介されています。

「統計的多重比較法の基礎」
永田靖・吉田道弘 著 サイエントリスト社

Q カテゴリカル変数(たとえばTREAT)を含むロジスティック回帰の分析をLOGISTICプロシジャにて実行しています。その際、GLMプロシジャ、GENMODプロシジャと同じように、CONTRASTステートメントを指定していますが、ログ画面に以下のメッセージが表示されます。このメッセージを回避する方法を教えてください。

```
WARNING: More coefficients than levels specified for effect
treat. Some coefficients will be ignored.
```

A LOGISTICプロシジャのCLASSステートメントでカテゴリカル変数を指定している場合、ダミー変数を作成するデフォルトの方法がEFFECTコーディングとなります。GLMプロシジャ、GENMODプロシジャでは異なる方法(GLMコーディング)でダミー変数を作成しているので、CONTRASTステートメントの係数の指定に注意する必要があります。ここでは、メッセージを回避する方法として2つご紹介します。なお、簡単な例として、水準数が3であると仮定します。

PARAM=オプションを用いてGLMコーディングに変更する例

```
PROC LOGISTIC DATA=one;
  CLASS treat / PARAM=glm; /* GLMコーディングを指定 */
  MODEL resp=treat;
  CONTRAST '1 vs 2' treat 1 -1 0;
  CONTRAST '1 vs 3' treat 1 0 -1;
  CONTRAST '2 vs 3' treat 0 1 -1;
RUN;
```

GLMコーディングを用いることにより、GLMプロシジャ、GENMODプロシジャと同じようにCONTRASTステートメントにおける係数の指定を行なうことができます。

デフォルトであるEFFECTコーディングに従い、係数の指定を行なう例EFFECTコーディングの場合、以下のように水準数より1つ少ない数のダミー変数(X_1,X_2)が作成されます。

TREAT	X_1	X_2
1	1	0
2	0	1
3	-1	-1

この場合、変数TREATにおける最初の水準(1)と2番目の水準(2)を比較する場合には、2行目の係数を1行目の係数から差し引くことにより、CONTRASTステートメントにて指定すべき係数を求めることができます。他の比較に関しても同じようにして求められます。その結果、下記のようにCONTRASTステートメントの係数を指定することになります。

```
PROC LOGISTIC DATA=one;
  CLASS treat;
  MODEL resp=treat;
  /* EFFECTコーディングにおける水準の比較 */
  CONTRAST '1 vs 2' treat 1 -1;
  CONTRAST '1 vs 3' treat 2 1;
  CONTRAST '2 vs 3' treat 1 2;
RUN;
```

Q 現在利用しているSASでは、起動時のログには「Release 8.2」と表示されますが、メニューバーの[ヘルプ] [リリース情報]を見ると「リリース 8.02」と表示されます。現在利用しているリリースはどちらになるのでしょうか。

A いずれの表記も同一のリリースを意味します。リリース番号についての正式決定が出荷直前まで遅れた経緯があり、リリース番号を取得する関数の出力情報に食い違いがあります。弊社における正式リリースは「8.2」としていますので、「8.02」を「8.2」と読み替えてください。なお、自動マクロ変数「SYSVER」と「SYSVLONG」の出力にも同様の差異がありますのでご注意ください。

実行例

```
%PUT &SYSVER;
%PUT &SYSVLONG;
```

ログには、それぞれ下記の値が出力されます。

```
8.2
8.02.02M0P012301
```

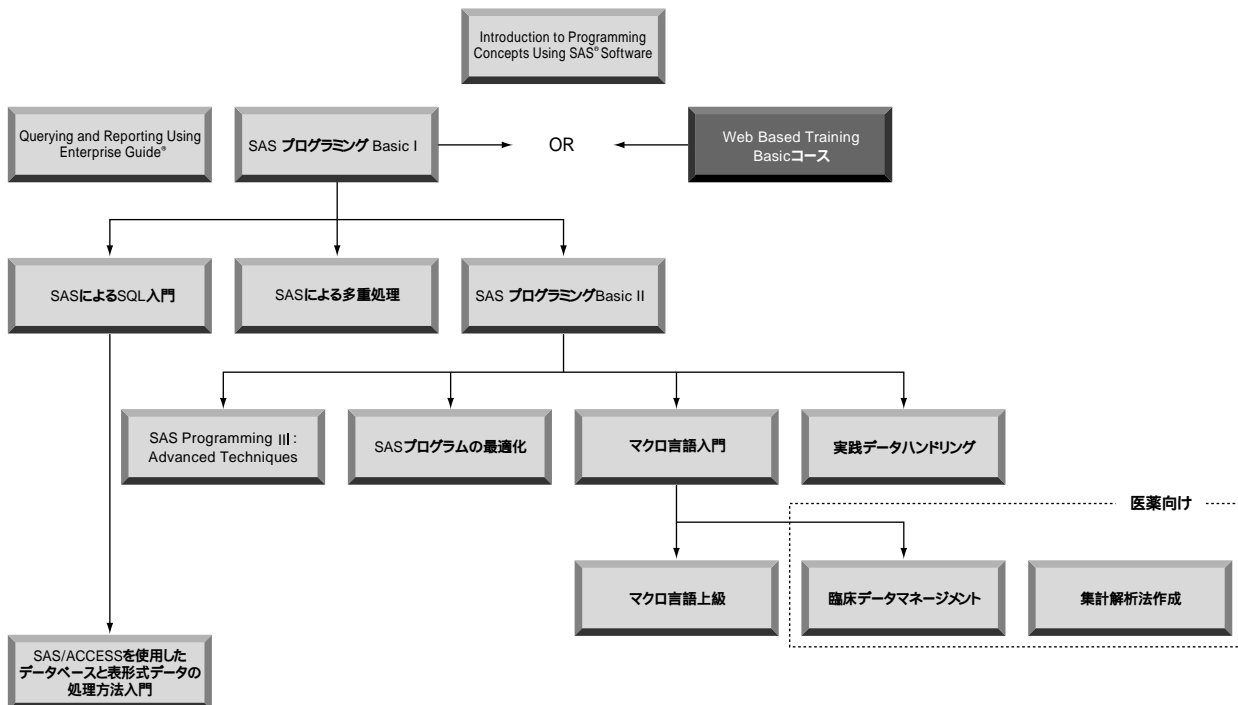
SAS Training

SASトレーニングのお知らせ

SAS Institute Japan株式会社では、SASのデータアクセス、管理、解析、そしてプレゼンテーションの機能をフルにご活用いただくために、さまざまなトレーニングコースを開催しています。このたび、8コースの定期トレーニングを2004年1月より順次新規開講します。また、特定の業務・業種・目的に合わせた「プロフェッショナルコース」についても順次開催します。

定期トレーニング 新コース開講

Accessing and Manipulating Data Curriculum
データへのアクセスと加工



「SASによるSQL入門」コース

日 程:2004年1月15日(木)~16日(金) 10:00~17:00(東京会場)
4月15日(木)~16日(金) 10:00~17:00(東京会場)
価 格:90,000円(チケット捺印数2)
学習内容:SASプログラムを使用したデータ検索ツールとしてのSQLプロシジャについて学習します。クエリの記述方法、複数のテーブルの結合、ビュー、インデックス、テーブルの作成方法、それらの更新や削除の方法について解説します。

「SAS/ACCESSを使用したデータベースと表形式データの処理方法入門」コース

日 程:2004年2月13日(金) 10:00~17:00(東京会場)
5月11日(火) 10:00~17:00(東京会場)
価 格:45,000円(チケット捺印数1)
学習内容:SAS/ACCESSを使用したリレーショナルデータベースやスプレッドシートのデータの読み込み、書き込みの方法について学習します。LIBNAMEステートメントのデータベースエンジン、SQLプロシジャのパススルー機能などについての解説も行ないます。

「SASによる多重処理」コース

日 程:2004年3月19日(金) 10:00~17:00(東京会場)
6月16日(水) 10:00~17:00(東京会場)
価 格:45,000円(チケット捺印数1)
学習内容:Windows環境を使用し、クライアント/サーバでの分散処理と多重処理の実装方法を学習します。
・多重処理の概念と設定方法
・クライアント/サーバ環境での多重処理
・データのダウンロード
・プログラムのリモート実行
・同期、非同期処理の利用方法

「SASプログラムの最適化」コース

日 程:2004年3月25日(木)~26日(金) 10:00~17:00(東京会場)
6月24日(木)~25日(金) 10:00~17:00(東京会場)
価 格:90,000円(チケット捺印数2)
学習内容:Windows環境を使用し、処理を最適化するためのSASプログラムの実装方法について学習します。

1日目

- ・メモリ使用量の最適化手法
- ・処理効率向上のためのプログラミング手法
- ・データエリの処理効率向上のための手法

2日目

- ・データソートの処理効率向上のための手法
- ・データ結合の処理効率向上のための手法
- ・データ保存領域削減のための手法

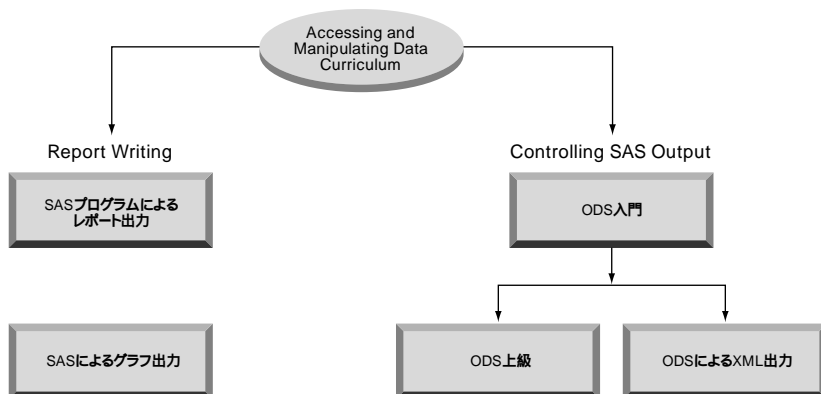
「SASプログラミング速習」コース

日程:2004年3月19日(金)/6月18日(金)10:00~17:00(東京会場)
2月19日(木)/5月20日(木)10:00~17:00(大阪会場)

価格:100,000円(チケットのお取り扱いはありません)

学習内容:「SASプログラミング Basic I」コース、「SASプログラミング Basic II」コース、そして「マクロ言語入門」コースの内容を1日で学習する速習コースです。プログラミング経験のある方を対象に、SASのデータセットの作成方法から基本的なプロシジャの使用法、データハンドリングの方法、マクロ言語の基礎について解説します。なお、自習教材として、「SAS Learning Edition リリース1.0」がセットとなっています(教材はトレーニング当日お渡しします)。

Data Presentation and Reporting Curriculum
データプレゼンテーション・レポートニング



「SASプログラムによるレポート出力」コース

日程:2004年3月18日(木)/6月17日(木)10:00~17:00(東京会場)

価格:45,000円(チケット捺印数1)

学習内容:REPORTプロシジャの詳細な使用法を、FORMAT、TABULATE、MEANSプロシジャの利用も絡めて解説します。また、明細表、集計表、クロスタブレポートなどの出力を、ODSを利用してHTMLファイルにする方法なども解説します。

「SASによるグラフ出力」コース

日程:2004年2月26日(木)~27日(金)10:00~17:00(東京会場)
6月3日(木)~4日(金)10:00~17:00(東京会場)

価格:90,000円(チケット捺印数2)

学習内容:Windows環境を使用し、カラーグラフィックの作成、編集、管理についての拡張機能を豊富なサンプルをもとに学習します。

1日目

- ・グラフテキストの作成と制御方法
- ・プロット図の作成と編集方法
- ・棒グラフ、円グラフの編集やデータ分類による統計量の出力
- ・マップデータを利用したブロックマップやプリズムマップの作成

2日目

- ・ODSを利用したHTML出力とポストスクリプトやPCLファイルの作成
- ・ANNOTATE機能を利用したプロット図、棒グラフ、マップの作成
- ・出力グラフの管理とテンプレートを利用したグラフ編集方法

「ODSによるXML出力」コース

日程:2004年1月29日(木)/3月23日(火)/4月27日(火)/6月10日(木)
10:00~17:00(東京会場)

価格:45,000円(チケット捺印数1)

学習内容:XMLへの導入とSASデータセットやプロシジャからXMLファイルを作成する方法を学習します。XML LibnameエンジンとODS XML(ODS MARKUP)ステートメントの比較、XMLタグセットの作成や更新の方法についても解説を行ないます。

プロフェッショナルコース

「医学研究における経時データ解析」コース

日 程:2004年3月11日(木)~12日(金)10:00~17:00(東京会場)

価 格:90,000円(チケット捺印数2)

受講対象:SASシステムによる統計解析に従事している方

前提知識:「医薬向け分散分析」コースを受講済みか、同程度の知識のある方

学習内容:医薬統計専門家により、以下の内容で開催します。

多くの医学研究においては、研究対象者ごとに時間を追って反応変数を繰り返し測定することがあり、このような研究は経時観察研究などと呼ばれます。その研究目的の一つは、反応変数と時間、治療効果あるいは他の説明変数との関係を定量的に評価することです。同一対象者内の反応変数には相関が生じることから、統計的に妥当な推論を行なうためには測定単位に起因する相関を考慮する必要があります。そのような反応変数間の相関を考慮する方法として、線型混合モデルおよび一般化推定方程式(GEE)による解析方法が近年注目を浴びています。このコースでは、線型混合モデルと一般化推定方程式の理論的背景について講義し、実際の医学データの解析を通してMIXEDプロシジャやGENMODプロシジャの使用法について学びます。

- ・ 線型混合モデルおよび一般化推定方程式(GEE)に関する基本概念の説明
- ・ MIXEDプロシジャやGENMODプロシジャの使用法
- ・ 医学分野における実データの解析

「SAS Enterprise Minerによる予測モデル」コース

日 程:2004年5月24日(月)~25日(火)(東京会場)

価 格:90,000円(チケット捺印数2)

学習内容:現在準備中です。詳細は右記<トレーニング担当>までお問い合わせください。

復活コースのご案内

下記の3コースについて、今秋より定期トレーニングとして復活しました。

- ・ 「ODS入門」コース
- ・ 「ODS上級」コース
- ・ 「SAS/INSIGHTによるビジュアルデータ解析」コース

今後も多岐にわたったトレーニングコースを追加していく予定です。日程等の詳細は順次弊社Webサイトに公開しますので、以下のURLをご参照ください。

<http://www.sas.com/japan/training/>

その他、トレーニングに関する情報については、上記のURL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

T E L 03-3533-3835

F A X 03-3533-3781

E-mail JPNTraining@sas.com

New Publication

新刊マニュアルのご紹介

「Instant ODS: Style Templates for the SAS Output Delivery System(CD-ROM)」

注文番号:58824

価格:5,500円(税抜)

「Instant ODS: Style Templates for the SAS Output Delivery System」で、SAS Output Delivery System(ODS)のフォーマット化機能をフルに活用できます。このCD-ROMには、50種類のODSスタイル・テンプレート集が収録されています。ユーザーは、業務にすぐに役立つスタイル・テンプレートを簡単に選択でき、また出力形式はHTML、PDF、そしてRTF形式から選べます。必要なテンプレートを選び、プログラムコードにテンプレート名を付加して実行すれば、最小限のプログラミングだけで視覚に訴えるアウトプットを得られます。このCD-ROMで、初心者から上級者まで、SASのすべてのユーザーがODSの便利さを実感し、業務に役立てることができるでしょう。

「Web Development with SAS by Example」

注文番号:58694

価格:9,200円(税抜)

本書はWebプログラミングにおける一般的な導入解説書であると同時に、SAS Webツールを使用するためのガイドブックとなっています。本書では、Webアプリケーション開発の総合的ツールであるSAS AppDev Studioを題材に、豊富な事例を通して、SASを使ったWeb開発の主要な課題を解説します。SAS AppDev Studioを利用すれば、ユーザーは新たに専門的なツールの難しい操作方法を学ぶことなく、これまでの経験を活かしてWebツールを開発・運用することができます。もちろん、SASが持つ、データの検索、分析、そしてレポートインギンといったSASのパワーをも、Webツールでフルに活用できるため、データ分析やプレゼンテーション作成の労力も軽減できます。

既にSASを活用しているユーザーは、本書のステップに沿って、Web上でのコンテンツ配布を簡単に実現できます。また、SASデータの活用に興味のあるWebサイト開発者は、SASが提供するソリューションのフレキシブルな操作性を実感できるでしょう。

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/>)にて公開していますので、併せてご利用ください。

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(OS/390, z/OS)	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 Winter 2004

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: JPNTechnews@sas.com



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
アクア堂島西館 12F
Tel 06 (6345) 5700
Fax 06 (6345) 5655

SAS Institute Japan株式会社 www.sas.com/japan/

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©2004, SAS Institute Inc. All rights reserved.