

# STN

## SAS Technical News

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



WINTER 2010

**特集** ..... 01  
**分析におけるグラフの活用**

**SAS Academic News** ..... 08

- 事例紹介
- コラム「SAS四方山話」
- カスタマー・リレーション

**Q&A** ..... 15

**SASトレーニングのお知らせ** ..... 17

**SAS Pressシリーズ**  
**売り上げTOP10のご紹介** ..... 19

**最新リリース情報** ..... 19

### 特集

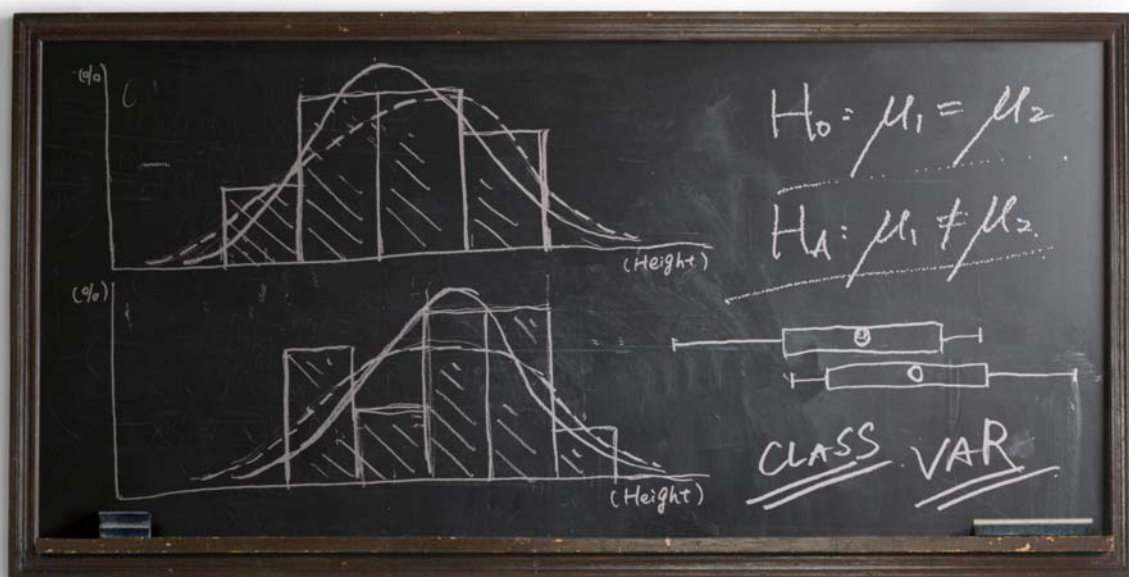
## 分析におけるグラフの活用

前々号では、SAS® 9.2におけるSAS/GRAPH®の新機能として、ODS統計グラフ(ODS Statistical Graphics)およびSAS/GRAPH ODS Graphics Editorなどを記述しました。前号では、Base SAS®に含まれている統計

プロシジャを用い、グラフ表示などをご紹介しました。

今号では、SAS/STAT®に含まれているプロシジャ、特に分散分析に関するプロシジャを用いながら、より簡略、きれいに作成できる

ようになったグラフの表示が、分析をすすめる上で活用できる点をご紹介します。また、SAS 9.2にてSAS/STATに追加されたプロシジャを、概略とともに紹介します。



Seeing is .  
Analyzing

## 分析をはじめる前に： グラフ表示

データの分析を始める前に、オブザベーションがどのような値となっているかを確認することは一つの重要なステップとなります。たとえば、性別として、男、女以外の値が含まれているか、体重、身長などの値として、0以下、もしくはとても大きな値が含まれているか、などの確認があげられます。また、他の値より大きく異なっているオブザベーション(外れ値)がないか、どのようにオブザベーションが散らばっているか(分布)を確認することも、用いる手法を選択、検討する上で、重要な情報となります。オブザベーション数がすくない場合、オブザベーション自体を確認することができます。しかしながら、非常に多い場合には、現実的な方法ではありません。このような場合、MEANSプロシジャを用いた要約統計量の算出、また、カテゴリカルデータの場合には、FREQプロシジャなどを用いた度数表の作成は一つの確認方法となります。

この他の方法としては、グラフ作成があります。グラフは、分析に用いているデータの確認ができるとともに、視覚的にデータを把握できるため、分析の結果を説明する上でもとても重要なツールとなりえます。これらの点に関し、2標本における平均値の比較検定、TTESTプロシジャを例として、より具体的に説明します。

### 1.1 2標本における平均値の比較検定

2つのグループにおいて、平均の値が同じであるかを確認する統計手法としては、2標本におけるt検定があります。TTESTプロシジャにて対応しており、以下のステートメントにて実行することができます。

※ここでは、性別ごとに体重、身長が観測されている、ライブラリSASHELPのCLASSデータセットをサンプルとしています。

#### 例

```
ODS GRAPHICS ON;
PROC TTEST DATA=sashelp.class;
  CLASS sex;
  VAR height;
RUN;
```

グループを表す変数SEX(性別)をCLASSステートメント、観測されている値の変数HEIGHT(身長)をVARステートメントにて指定しています。さらに、SAS 9.2から、正規の機能となりましたODS GRAPHICSステートメントは、TTESTプロシジャにても用いることができ、デフォルトにて以下のグラフ(図1参照)の他、QQプロットが作成されます。

※PROC TTESTステートメントのPLOTS=オプションにて、表示するグラフを指定することができます。例えば、PLOTS(ONLY)=BOXPLOTの記述にて、箱ひげ図のみを表示できます。

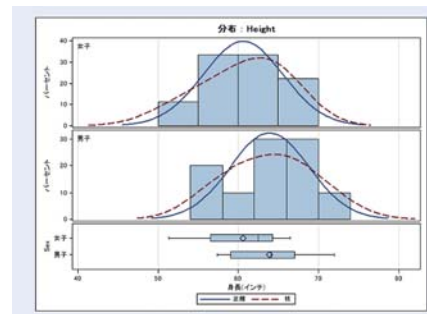
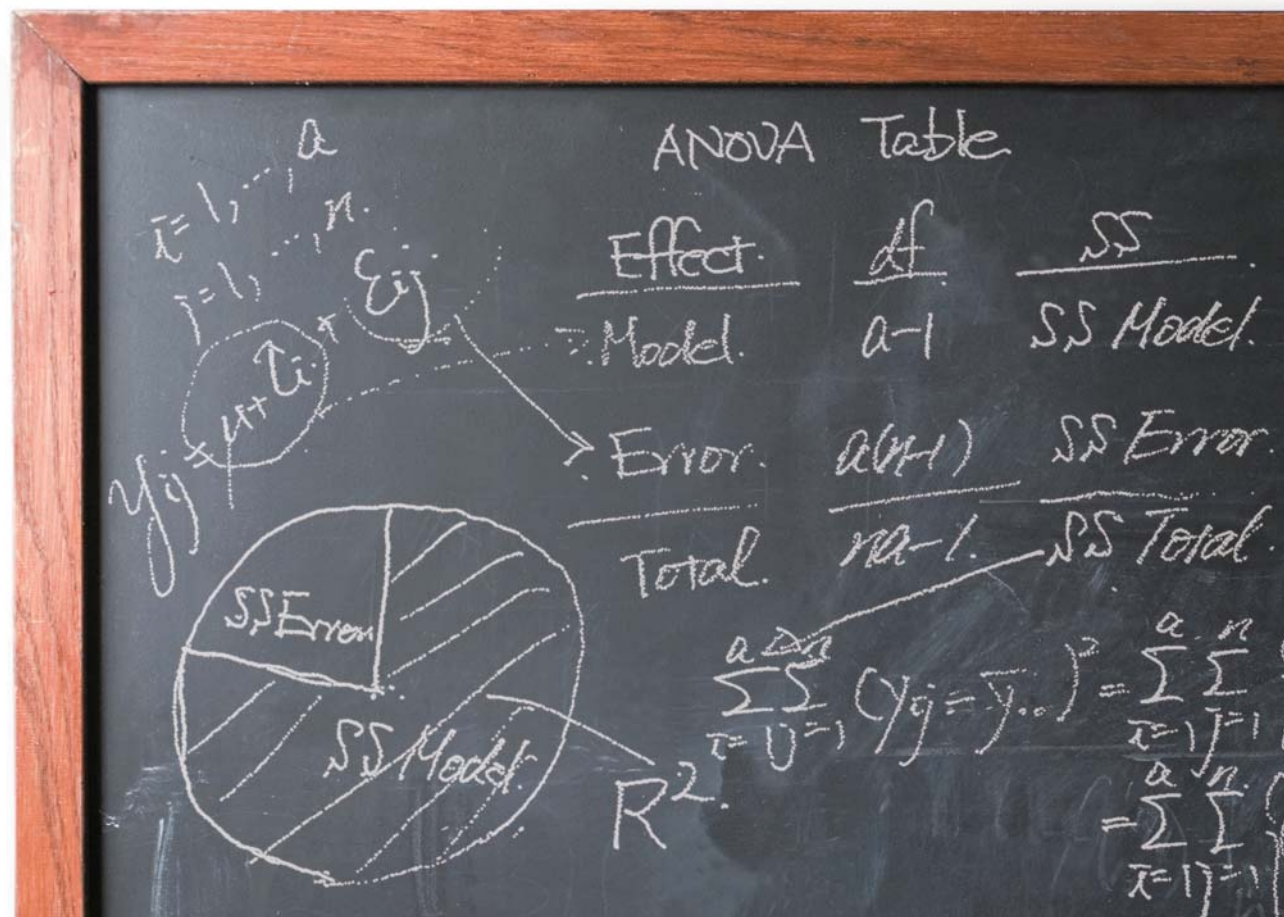


図1: ヒストグラムと箱ひげ図





グループごとのヒストグラムから、値がどのように散らばっているか、また、正規分布と比較し、大きく外れていないかを確認することができます。また、下の部分では、各グループの箱ひげ図が表示されており、四分位範囲である箱の長さより、各グループにおけるばらつきがほぼ同じであるかも確認することができます。平均値の位置はシンボル◇にて表示されており、グラフ上では違いがあるようにも見えます。しかし、分布のばらつき(分散)と考慮した場合、母集団における平均の値が異なっているかは定かではありません。

このように、グラフからは多くの情報を視覚的にとることができるのですが、人によってえられる印象が異なってくることがあります。このため、TTESTプロシジャのアウトプット、検定統計量(図2参照)を参照します。

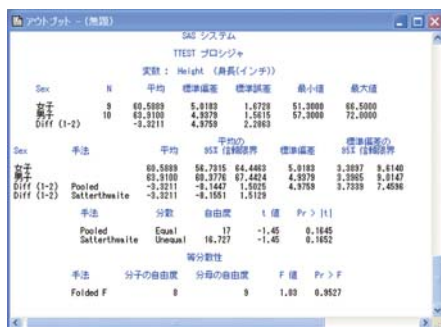


図2:TTESTプロシジャのアウトプット

最初に等分散性の部分では、Folded F統計量は $F=1.03(df=8,9)$   $p=0.9527$ の値となっており、グループにおける分散が異なるとは言えず、等分散性を仮定することができます。このため、検定統計量としては、手法Pooledの部分から検定統計量 $t=-1.45(df=17)$   $p=0.1645$ を参照します。この結果、データからは、母集団において、身長に性別差があるとはいえないことになります。

## 1.2 Enterprise Guide® 4.2での実行

TTESTプロシジャのプログラムを記述、実行することによって、グラフ、およびアウトプット(統計量)をえることができます。一方、プログラムを記述することなく、GUIにて同様の分析を行なうことができます。

Enterprise Guide® 4.2(以下、EGと略称)では、データをインポートした後、メニューより、「タスク」→「分散分析」→「t検定」を選択することで、平均値に対する検定を行なうことができます。t検定の種類にて用いる手法、データにて変数などを指定した後、グラフの箇所にてグラフの種類を指定できます。(図3参照)

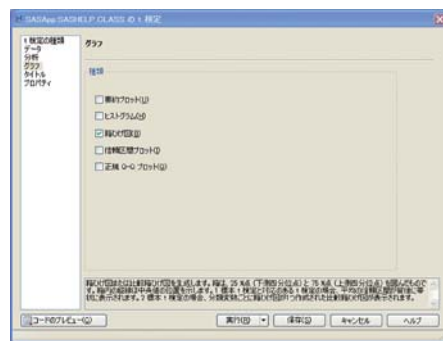
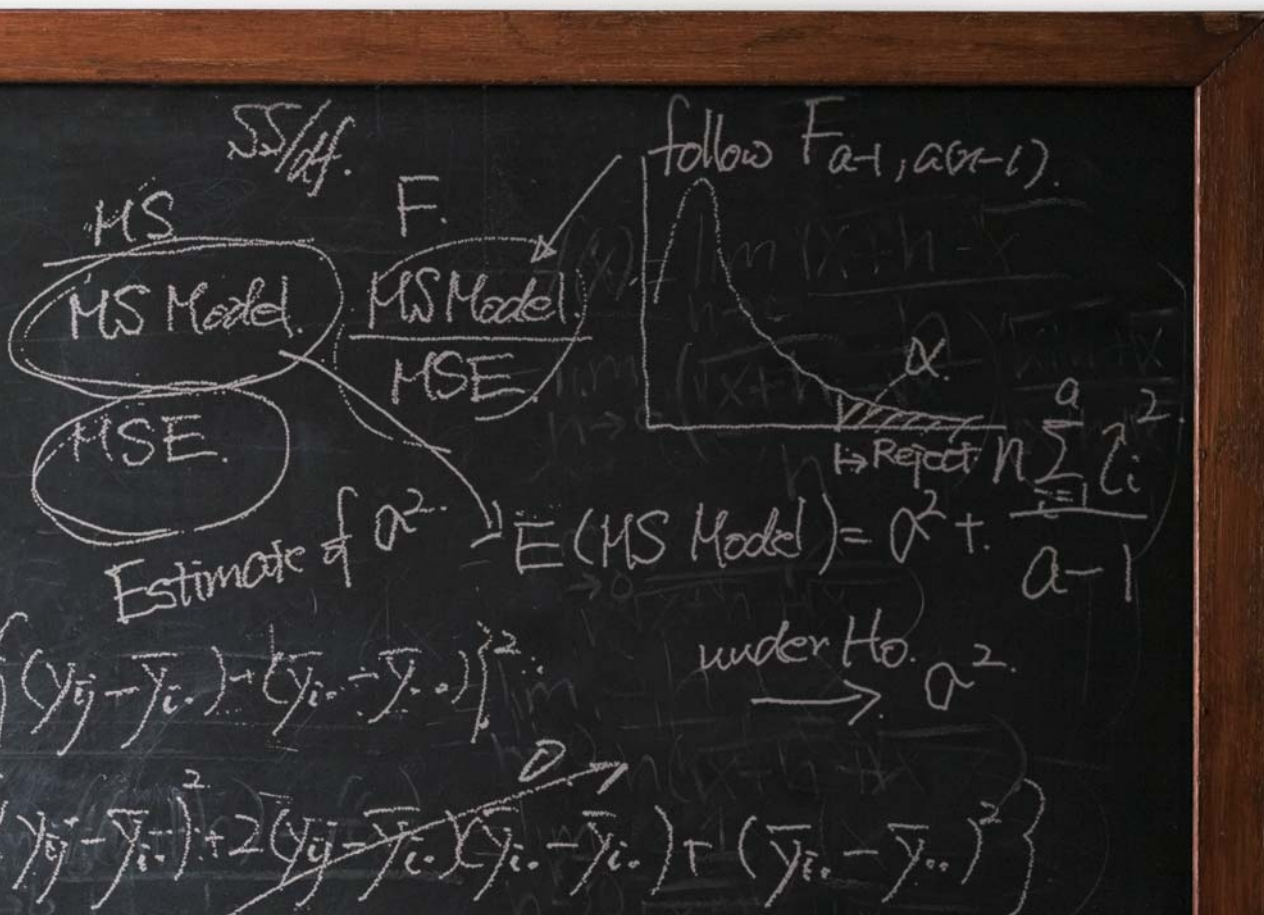


図3:EGにおけるグラフの選択画面



### 1.3 新機能：TTESTプロシジャ

TTESTプロシジャでは、すでに記述しましたODS GRAPHICSステートメントがサポートされている他、オプションがPROC TTESTステートメントに追加されています。

#### BYVAR | NOBYVARオプション

PAIREDステートメント、VARステートメントにて複数の変数が指定されている場合、要約統計量などはデフォルトでは個別の表として表示されます。NOBYVARオプションを用いることによって、一つの出力部分に複数の変数に対する結果を表示できます。

#### DIST=オプション

デフォルトではDIST=NORMALとなっており、分布としては正規分布を仮定しています。あらたにDIST=LONGNORMALの指定が可能となり、対数正規分布を仮定できます。

#### ORDER=オプション

CLASSステートメントにて指定した変数の水準順序を指定することができます。デフォルトではORDER=MIXEDとなっており、以前のリリースにおける水準の並びと同じになります。

#### SIDES=オプション

両側検定に加え、片側検定(上側、下側)をオプションにて指定できます。

#### TEST=オプション

デフォルトではTEST=DIFFとなっており、平均値の差に対する検定となります。新たにTEST=RATIOのオプション指定が可能となり、平均値の比に対する検定を行なえます。

#### TOSTオプション

SchurmanのTOST(Two One-Sided Tests)手法における、同等性の検定を行なえます。

また、VARステートメントでは、**CROSSOVER**=オプションが追加されており、2期2剤クロスオーバー分析に対応しています。

## 平均値の比較： GLMプロシジャ

TTESTプロシジャでは、2グループにおける平均値の比較に対応していますが、より多くのグループ平均の比較には対応していません。この場合、分析の一手法として、分散分析があります。

※ここでは、以下のプログラムにて作成される、4グループ(A、B、C、D)、各グループにおけるオブザベーション数が20のサンプルデータを用います。

#### サンプルプログラム

```
/* 出力形式の定義 */
PROC FORMAT;
  VALUE temp_g 1='A' 2='B' 3='C' 4='D';
RUN;

/* サンプルデータの作成 */
DATA sample;
  /* 初期シード値の指定 */
  CALL STREAMINIT(12345);
  DO group=1 TO 4;
    DO i=1 TO 20;
      /* RAND関数を用いた乱数の生成 */
      resp=RAND('NORMAL', 5-2*(group>2), 2);
      OUTPUT;
    END;
  END;

/* 出力形式の指定 */
FORMAT group temp_g.;
RUN;
```

### 2.1 2標本からの拡張

3つ以上のグループ平均値の比較を行う場合、すべての2グループの組み合わせに対して、TTESTプロシジャを繰り返し、行なうことが考えられます。しかしながら、すべての平均値が等しいという一つの帰無仮説に対し、複数の検定を行なっていることとなります。たとえば、3グループである場合には、2グループの組み合わせは3つとなり、3回の検定を行なうこととなります。このため、全体としての有意水準がより大きくなってしまいます。この点も含め、2標本t検定の繰り返しには問題があり、一度に3つ以上のグループ平均を比較するためには、他の手法、分散分析が必要となります。SAS/STATでは、ANOVAプロシジャ、GLMプロシジャなどにて対応しています。

### 2.2 GLMプロシジャの記述

線形モデル(General Linear Model)に対応しているGLMプロシジャにて、分散分析を行なうことができます。このプロシジャでは、CLASSステートメントにてグループ変数、MODELステートメントにて、応答変数およびモデル式を記述します。

#### 例

```
ODS LISTING SGE=ON;
ODS GRAPHICS ON;
PROC GLM DATA=sample;
  CLASS group;
  MODEL resp=group / SOLUTION;
RUN;
```

MODELステートメントでは、パラメータ推定値を表示するためのオプションSOLUTIONを追加しています。

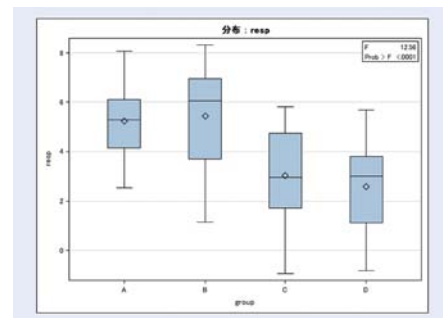


図4:箱ひげ図

TTESTプロシジャと同様、箱ひげ図(図4参照)が作成されます。各グループの平均値、オブザベーションのばらつきなどを視覚的に確認することができ、グループA、Bの平均とグループC、Dの平均がほぼ同じであるが、全体としてはグループごとの平均値は異なっているようです。また、箱の長さ(四分位範囲)より、各グループにおけるばらつき(分布)はほぼ同じであり、等分散性の条件を大きく外れていないと捉えることができます。

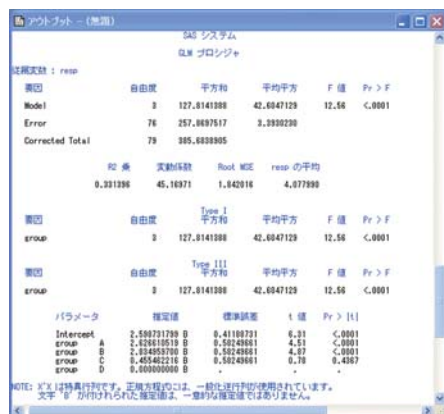


図5: GLMプロシジャのアウトプット(その1)

GLMプロシジャでは、最小二乗法を用いてモデルの推定が行なわれ、平方和、平均平方の値を含むANOVA表(図5参照)が出力されます。

この部分では、すべてのグループ平均が同じであるという帰無仮説に対して、F統計量、およびp値が算出されます。この例では、 $F=12.56(df=3,76)$ 、p値は0.0001以下となっていますので、母集団において、少なくとも一つのグループ平均が他のグループ平均とは異なっていることが、このデータからわかります。

SOLUTIONオプションを指定したので、アウトプットにパラメータ推定値が表示されます。GLMプロシジャでは、CLASS変数における最後の水準に対するパラメータを0とし、推定を行うので、各水準における推定値は、最後の水準と比較した場合の差となります。GLMプロシジャでは、対話式に分析をすすめることができ、さらに以下のステートメントを実行することによって、LS平均の値、およびグラフの作成(図6参照)を行なえます。

#### 例

```
LSMEANS group;
RUN;
```

ここでは、ODS Graphics Editorの機能を用い、SOLUTIONオプションにて表示されているパラメータ推定値を図示、理解しやすいようにしています。

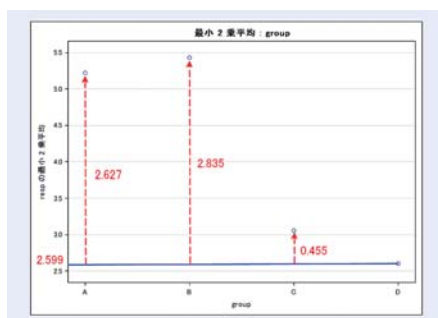


図6: LS平均のグラフ - SAS/GRAPH ODS Graphics Editorにて情報(赤、青の部分)を追加

### 2.3 多重比較

上記のアウトプットから、少なくとも一つのグループ平均が異なっていることがわかりましたが、どのグループ平均が異なっているかに関しては、わかりません。この点に関しては、LS平均値の比較、多重比較を行なうことが一手法となり、LSMEANSステートメントのPDIFオプション、また、多重比較の手法を指定するADJUST=オプションを用いることができます。

#### 例

```
LSMEANS group / PDIF ADJUST=tukey;
RUN;
QUIT;
```

※QUIT:ステートメントにて、GLMプロシジャのプロセスを明示的に終了としています。



図7: GLMプロシジャのアウトプット(その2)

各グループに対する最小二乗平均(LS平均)の値とともに、各組み合わせに対する比較の結果、p値が出力されています。p値は小さいほど、2つのグループ平均が異なっていると捉えることができますので、グループA、Bの平均値、グループC、Dの平均値は同じであるが、その他の組み合わせでは異なっていると解釈できます。また、ODS GRAPHICSステートメントにより、図8のグラフが表示され、視覚的に各平均値の関連を捉えることができます。

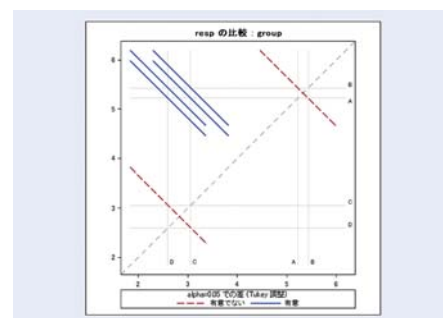


図8: 多重比較のグラフ

多重比較の手法としてTUKEYを指定していますが、この他、BON、DUNNETT、NELSON、SCHEFFE、SIDAK、SIMULATE、SMM|GT2、Tのオプションを指定できます。また、2つのLS平均の比較ではなく、グループAとグループBの平均がグループCとグループDの平均と同じであるかなど、対比を用いた検証の場合には、CONTRASTステートメント、ESTIMATEステートメントを用います。より具体的な指定、例題に関しては、以下のUsage Noteが参考となります。

#### Usage Note 24447:

Are there any examples of writing proper CONTRAST and ESTIMATE statements?

<http://support.sas.com/kb/24/447.html>

なお、以下のページでは、上記Noteを翻訳したものを記載しています。

CONTRAST、およびESTIMATEステートメントを記述する例題

<http://www.sas.com/jp/service/technical/faq/list/body/stat091.html>

### 2.4 EGでの実行

EGでは、データのインポートを行った後、以下の選択を行なうことによって、上記のGLMプロシジャにおける分析を実行することができます。

「タスク」-->「分散分析」-->「線形モデル」

表示されるウィンドウのデータにて、変数RESPを従属変数、変数GROUPを分類変数に指定します。モデルにてどの変数を効果としてモデル式に含めるかを指定します。この例では、変数GROUPを効果として追加しています。(図9参照)





図9: EGにおけるモデルのウィンドウ

サンプルプログラムでは、LSMEANSステートメントを記述し、LS平均のグラフ、また、多重比較の結果を出力しています。EGでは、**Post-Hock検定-最小二乗平均**の箇所に行なうことになります。最初に、変数GROUPを指定するために「追加」の部分をクリックし、図10のように選択することで、多重比較の結果などがえられます。



図10: EGにおける最小二乗平均のウィンドウ

※EG上では、CONTRASTステートメント、ESTIMATEステートメントに対応しているメニューがありません。

## 2.5 さらに複雑なモデルへ

複数のグループの平均値を比較する方法として、分散分析、対応しているGLMプロシジャを記述しました。このプロシジャでは、MODELステートメントにて複数の効果を指定することによって、多くの実験計画に対する分析を行なうことができます。しかしながら、モデルに含まれる変数は、固定効果として扱われ、分析が行なわれるため、基本的にランダム効果を含めた分析に対応していません。一方、**MIXED**プロシジャは、線形混合モデル(Linear Mixed Model)に対応しており、RANDOMステートメントでモデルに含めるランダム効果を指定することができます。また、オブザベーション間における相関を考慮するような反復測定分析の場合、REPEATEDステートメントにて相関構造を指定することができます。

※GLMプロシジャとMIXEDプロシジャの違いに関し、以下のUsage Noteがあります。

### Usage Note 22609:

**Is it possible to reproduce my PROC GLM analysis with PROC MIXED?**

<http://support.sas.com/kb/22/609.html>

なお、以下のページでは、上記Noteを翻訳し、記載しています。

### GLMプロシジャとMIXEDプロシジャの違いについて

<http://www.sas.com/jp/service/technical/faq/list/body/stat075.html>

観測されている値が数値である場合には、TTESTプロシジャ、GLMプロシジャ、MIXEDプロシジャなどを用いることができます。しかしながら、ある事象が生じたか否か(Yes or No)など、2値の場合には、用いている手法が適していません。2値データ(もしくは多項データ)の場合には、ロジスティック回帰分析など、二項分布(もしくは多項分布)を仮定した分析を用いることがあり、**CATMOD**プロシジャ、**LOGISTIC**プロシジャ、**GENMOD**プロシジャなどが対応しています。特にGENMODプロシジャでは、一般化線形モデル(Generalized Linear Model)に対応しており、二項分布、Poisson分布などを仮定した上で、モデルの推定、パラメータ推定値の算出を行なうことができます。

さらに、SAS 9.1.3より、二項分布、Poisson分布などを仮定し、ランダム効果をモデルに含む、一般化線形混合モデル(Generalized Linear Mixed Model)に対しては**GLIMMIX**プロシジャが正規版として追加されています。観測されているオブザベーションがどのような値であるか、どのような分布を仮定できるのか、また、ランダム効果を含むモデルの指定を行なうかなどを考慮し、用いるプロシジャを選択していただく形となります。

米国にて開催されているSAS Global Forumにて、これらの4つのプロシジャを比較している文献があり、以下のページにて公開されています。

### Comparison of Features and Applications of four Linear Models Procedures

<http://support.sas.com/resources/papers/proceedings09/258-2009.pdf>

## その他の新しいプロシジャ: SAS/STAT

SAS/STATは、多分野における分析手法に対応しているため、非常に多くのプロシジャを含むプロダクトとなっています。既存のプロシジャに関しては、一覧をSAS Technical News Summer 2007にて、概略とともに記載しています。

<http://www.sas.com/jp/periodicals/technicalnews/pdf/07sum.pdf>

すでにご紹介しました**GLIMMIX**プロシジャは、SAS 9.1.3にて当初、評価版のプロシジャでした。その後、正規版のプロシジャとなりましたが、依然、モジュールのダウンロード、インストールが必要でした。SAS 9.2では、GLIMMIXプロシジャは正規版のプロシジャとして含まれており、別途、ダウンロード、インストールの必要がありません。

この他、カテゴリカル変数を含む回帰分析、および変数選択に対応している**GLMSELECT**プロシジャ、中央値回帰(メディアン回帰、Median Regression)を含む分位点回帰分析(Quantile Regression)に対応している**QUANTREG**プロシジャもSAS 9.2から正規版のプロシジャとなっています。

また、ベイズ解析に対する評価版の**BGENMOD**プロシジャ、**BLIFEREG**プロシジャ、**BPHREG**プロシジャはWindows(32bits)環境SAS 9.1.3にて実行できます。SAS 9.2では、これらの機能は、GENMODプロシジャ、LIFEREGプロシジャ、PHREGプロシジャの**BAYES**ステートメントにて実行することができます。GLIMMIXプロシジャと同様、正規版の機能として追加されていますので、別途、モジュールをダウンロード、インストールすることなく、Windows環境に限らず、ご利用できます。

※SAS 9.2では、BGENMODプロシジャ、BLIFEREGプロシジャ、BPHREGプロシジャは実行できません。

新たに、多変量の正規分布に基づき、擬似乱数を生成する**SIMNORMAL**プロシジャが追加されています。このプロシジャでは、変数間の相関を考慮した擬似乱数の生成、および条件付きの擬似乱数の生成を行なうことができます。

ベイズ解析に関しては、GENMODプロシジャ、LIFEREGプロシジャ、PHREGプロシジャのBAYESステートメントに加え、より一般的な事前分布、尤度関数を用いることができるMCMCプロシジャがあります。また、逐次計画法に対するプロシジャとしてSEQDESIGNプロシジャ、SEQTESTプロシジャが追加されています。

※MCMCプロシジャ、SEQDESIGNプロシジャ、SEQTESTプロシジャは、SAS 9.2 TS2M2から正規版プロシジャとなります。

評価版のプロシジャとしてはHPMIXEDプロシジャ、TCALISプロシジャが含まれています。線形混合モデルの推定において、固定効果、ランダム効果の水準が非常に多い、オプザベーション数が非常に多い場合などに有用となります。ただし、MIXEDプロシジャにて推定ができる、すべてのモデルに対応はしていません。TCALISプロシジャは、共分散構造分析に対応しているCALISプロシジャを拡張しており、多母集団の同時解析なども含まれています。

新たに追加されたプロシジャ、正規版となったプロシジャに関し、簡略に記述しましたが、他のプロシジャにおいても新たなオプションなどが追加されており、機能が拡張されています。各プロシジャにおける拡張点に関しては、米国SAS社の以下のページにて記載されています。

<http://support.sas.com/documentation/onlinedoc/stat/index.html>

※What's New in SAS/STAT 9.2を選択します。

## 4 参考資料

今号も含め、過去のTechnical Newsでは米国にて開催されている“SAS Global Forum”の文献をいくつか紹介しております。文献に関しては、以下のページより、参照できます。

SAS Global Forum (米国)

<http://support.sas.com/events/sasglobalforum/previous/index.html>

日本にて開催された“SASユーザー総会”における発表論文の一覧、補足資料に関しては以下のページにて公開しています。

SASユーザー総会 (日本)

<http://www.sas.com/jp/usergroups/index.html>

また、今号にてご紹介しましたNoteをはじめ、多くのUsage Noteがあり、以下のページから探索できます。

SAS Usage Note (検索)

<http://support.sas.com/notes/index.html>

[Search only Samples and SAS Notes.]の部分に、プロシジャ名、手法名などを記述し、検索します。すでに、GLMプロシジャとMIXEDプロシジャによる違い(Usage Note 22609)をご紹介しましたが、その他にも多くの有用なNoteがあります。例えば、以下のUsage Noteでは、多くの手法のリストとともに、関連するプロシジャ、詳細なページへのリンクなどが含まれています。

Usage Note 30333:

FASTats: Frequently Asked-For Statistics

<http://support.sas.com/kb/30/333.html>

この他、日本語のFAQは、以下のページにて参照することができます。

<http://www.sas.com/jp/service/technical/faq/index.html>

## 5 おわりに

今号では、分析を行う際、プロシジャにて算出される統計量のみではなく、グラフを参照することを記述しました。このグラフ作成は、以前のリリースでは、別のプロシジャなどを用いることが必要でしたが、SAS 9.2ではODS統計グラフの機能を用いることによって、簡略に作成できます。また、SAS/STATにおけるプロシジャの拡張に関しても、記載しました。

分析の手法を検討、選択する上で、また、分析の結果を理解し、説明する場面にて、今号の内容が参考となれば幸いです。



# SAS SAS アカデミック・ニュース Academic News

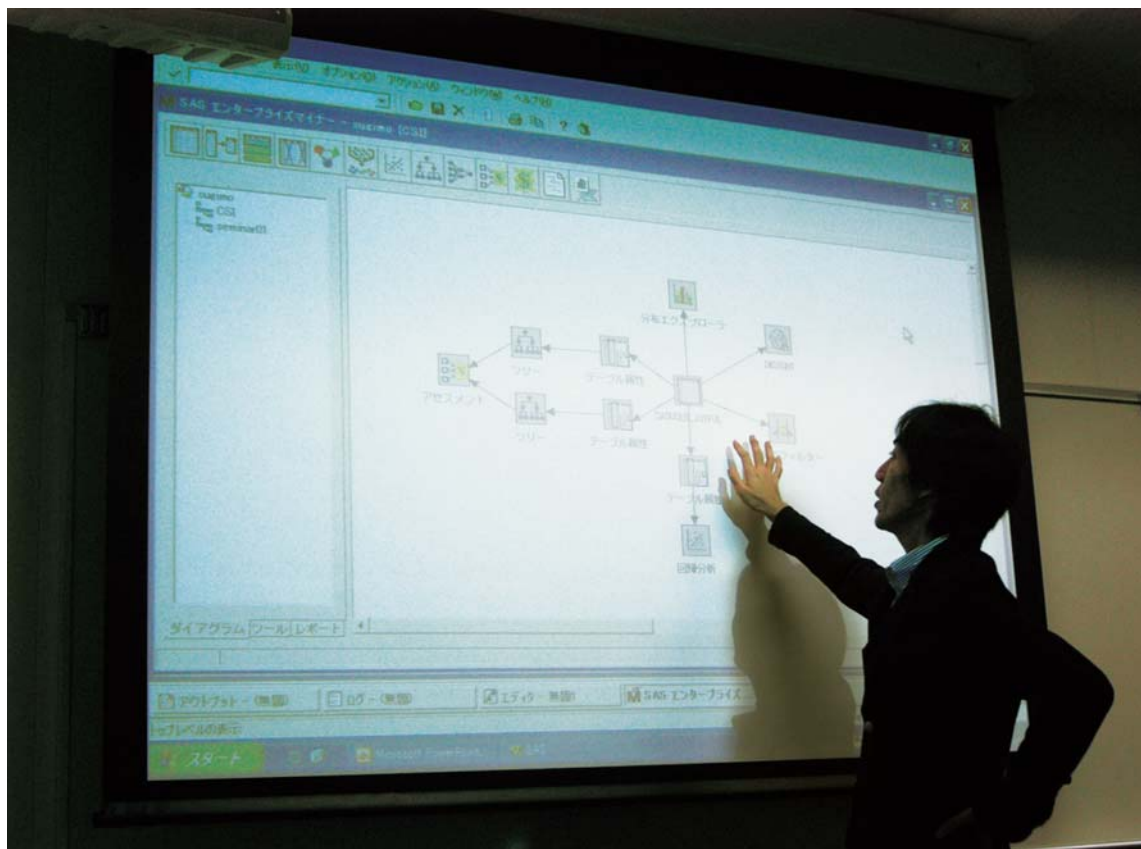
事例紹介はいよいよ佳境に入った福岡大学のEnterprise Guide活用編、SASコラム四方山話「アンケートデータの解析③」では、回帰分析および重回帰分析手法の説明とSASが出力した処理結果からの理解について、カスタマー・リレーションでは山形大学にて教員と学生向けSAS講習会開催の後半編を紹介しております。なお、事例紹介として投稿いただいております福岡大学のご紹介記事は、今号が最後となります。ご多忙の中、2009年春号から2009年冬号までの4号の記事をご提供いただきました福岡大学 杉本先生へは編集部より紙面を通じ改めてお礼申し上げます。

事例紹介

コラム  
「SAS四方山話」カスタマー  
リレーション

## 事例紹介

杉本宏幸  
福岡大学商学部准教授 博士(商学)



ました。

今回は一年弱の総括となりますので、講義の様子も踏まえながら、現状および現状の問題点、そして今後の課題についても触れさせていただければと存じます。

### 講義でのEGの使用

ゼミナールでの試験的な活用を経て、9月からのマーケティングリサーチの講義でEGを活用しています。本学商学部のマーケティングリサーチは2クラスが開講されていま

前回は、ゼミナールでの状況についてご説明いたしました。前回も触れましたが、本学商学部の学生達がプログラムベースのSASを使用するには一定程度の訓練が必要です。しかし、GUIに優れるSAS Enterprise Guide(以下、EGと略称)であれば、これを入り口としてデータ分析に入門できる可能性が高いことが、ゼミ生達の様子から明らかになってきました。

繰り返しの強調となってしまいますが、本学商学部ではデータ分析が必ずしもメ

ジャーでありません。しかし、ビジネスの現場でデータをどう扱うのか、それを意思決定にどのようにリンクさせていくかが重視される現状では、基礎的なデータ分析を避けて通れません。

こうした背景を鑑みれば、より多くの学生達にデータ分析に親しんでもらい、データ分析に一定程度の知識を持った学生の比率を今より上げて行くことは本学商学部にとって重視すべき課題と言えます。そのためのツールとしてEGの可能性を探ってき

ますが、EGを使用しているのは私の担当クラス(約40名)です。

講義では学生達数名でグループを編成してもらい、調査プロジェクトの立案、質問紙の作成と調査、データ分析(単純集計、度数分布、クロス集計分析、相関係数、回帰分析、分散分析、必要に応じて因子分析とクラスター分析)、そして調査結果のプレゼンテーションを実践してもらいます。半期週一回の講義で、かなりハードなスケジュールとなります。



受講生のデータ分析の経験はExcelによる表計算が主で、SAS等の統計ソフトに関する事前知識はほぼゼロです。因子分析等の分析手法までをクリックベースの簡単な操作だけで行なえるという条件でソフトウェアを選定すると、本学で利用可能なものはEGまたはSPSSが現実的な選択肢となります。プログラムベースのSASとR等も本学で利用可能ですが、講義でプログラムの学習に時間を割けないため、私の担当クラスでは選択肢に入れることが困難です。

講義はまだ半ばですが、学生達はあまり苦労せずEGを使っているように感じています。ゼミ生達への指導から、現在の学生達はパソコン操作にかなり習熟していること、操作方法を細かく指示しておけばEGの基本操作の習得は意外に易しいことがわかってきました。そのため、講義ではかなり細かい配布資料を用意して説明をしています。

さしあたって効果が見られているのは、講義中に「先生(ソフトが)動きません!」という声がほぼ聴かれなくなったことです。私が担当する限り、この声はほとんどの場合、Excel等で構築したデータをSAS等の統計ソフトで読み込む作業で発生します。これが起こると、私が教室中を右往左往したり、予め学生達のデータを預かってチェックしたりと時間のかかる作業が必要でした。結局、学生達も統計ソフトに対して「わけのわからないものをやらされた」とだけ認識して、講義やゼミナールを終えてしまっていたのではないかと考えます。

講義では分析結果の解釈も重視しています。このため、SAS環境が無い状況でもアウトプットを閲覧、編集できることが望ましいです。SAS環境が無くても、分析結果を確保してExcel等で図表を作成、加工できれば、分析結果をじっくり解釈できるはずです。そのため講義ではデフォルトのHTML形式だけでなく、必要に応じてRTF形式(またはテキスト形式)での出力とその保存を勧めています。少なくとも、本学商学部の学生達にとっては高度な分析を駆使したり、難しいプログラムを書いたりすることが必要ではありません。データ分析とその結果がマーケティング



やビジネスにどんなインプリケーションを有しているかを読み解くことが重要です。このために、学生・教員がともに苦戦していた作業を何段階か一挙に取り払ってくれたという点で、EGの活用は意義があると現段階では判断しています。

### 今後の活用方法

現在、本学商学部のマーケティングリサーチでは、私の担当クラスがEGを、もう一つのクラスがSPSSを採用しています。現状として、SASとSPSSはどちらかが良いとは言い切れない状況にあります。

今後EGを活用していくにあたって着目すべきと考えているのはETL(Extract Transform Loading)ツールとして利用可能という点です。データベースにアクセスし、分析データへ加工処理するシーンは、小売、通販企業、物流等といった大規模データを扱う企業では増えてきていると思います。

本学のマーケティングリサーチの講義でもデータベース処理の内容が含まれており、仮想の小さなFSP(Frequent Shoppers Program)データを構築、処理する演習を行っています。ただ、これは教職免許との関連からEGとは異なる別の指定ソフトウェアを利用しています。

しかし、今後、学内のカリキュラムが見直されていくとともに、講義で利用可能な

データが増えれば、流通や通販等の進路に向けて人材を育成していくことも可能だろうと思っています。理想的なのは、ビジネスで実際に使われている実データを開示してもらって講義等で活用することです。

これは単にソフトウェアの選択・使用に止まらず、本学と地元企業をはじめとしたビジネスの現場がどのような関係を構築できるかという問題、学生の進路・就職、さらには前々回に言及した私のゼミ卒業生をはじめとした卒業生達との関係づくりとも関わってくるため、重視すべき課題と認識しています。

また、EGをETLツールとして活用し、データの分析はSAS Enterprise Miner(以下、EMと略称)を活用すれば、EGとEMを連携したより効率的な分析が可能になると考えます。EGの分析機能は簡易なものなので、EGで分析に必要なデータを抽出し、より深く分析を行なう場合、EM等でデータを読み込んで分析すれば効率的だろうと考えます。

現在、私のゼミナールの学生達が、アンケートデータをEMで分析しようとしていたり、複数の政府統計の必要な部分だけEGで抽出しようとするなど、上記の部分的な試みを行っています。EGだけに止まらない分析を行うには、ゼミナールでも少しきちんと教育する仕組みを整える

必要があります。例えば、本学はSAS社と“Campus program”を契約している関係からSAS Trainer's Kit(Teaching Material)が利用可能なので、今後はこれをゼミ(や講義)で用することを検討しています。

### 残された課題

一年間を通じて改めて確認したことは、“データを読み解く目”を養うことが商学部のようなビジネスに近い学部では決定的に重要だということです。EGやEMを使えることはもちろん必要です。しかし、SASや統計ソフトはあくまでもデータ分析のツールなので、それを使うこと自体が商学部では目的とはなりえません。

解決すべきビジネス上の課題を設定する能力を持っている学生は、これが相対的に低い学生に比べ、分析結果の読み方が違っていると感じています。マーケティングやビジネス、さらに統計学等に関する知識があるか否か、もしくはそれらに対する強い関心があるか否かが“データを読み解く目”を養成する鍵になっているように思っています。

このためには、ゼミナールを核として、マーケティングに限定されないビジネスの知識、統計学・数学の基礎的な知識、共通教育科目(教養科目)、そしてパソコンの基礎的な操作等について体系的に履修するプログラムの開発とそれを1年次から学生達へ提供していくことが必須と考えます。

しかし、マーケティングやビジネスに関して学ぶことを重視すると、分析手法を身につける時間が無くなってしまうのが本学商学部の現状です。そうでなくとも、本学商学部の学生達は、データ分析、数学や数式を使うことに消極的な傾向があります。

マーケティングやビジネスの現場とデータ分析がどのように結びついているのか、ひいては将来の仕事にデータ分析がどう結びつくのか学生達は意識できていないのだと思います。このため、今まで以上に実務での活用事例や可能なら実データによる演習を取り入れていく必要性を感じています。

ビジネスの現場でデータ分析がどのよ

うな形で活用され、必要とされ、そして将来の仕事に関わってくるのか学生達が理解できるかは、今後、本学とビジネスの現場がどれだけ密接に連携できるかに大きく依存していると考えます。この意味で、産学官連携等、本学と外部との関わりは研究面だけでなく、教育上、重要なものとして認識すべきです。同時に、外部との関わりは学生達の進路に関わる出口教育とも密接に関わっているはずで、そして、出口教育を見据えた学部の専門教育として、データ分析の比重をどの程度に置くかは、今後の重要な検討課題とならざるを得ないと考えます。

EGとこれを入り口としたデータ分析の試みは、現在の本学商学部における様々な課題を「見える化」しつつあると考えます。福岡大学商学部の試みはまだ着手しはじめたばかりですが、今後は、ここで掲げた課題へ早急に手を打つことが必要になってくると考えております。





## コラム「SAS四方山話」

### 第8回 アンケート四方山話 (6) 「アンケートデータの解析③」

大橋 渉

東京医科歯科大学 歯医学総合研究科大学院 生命情報学  
情報医学センター 臨床・生物統計グループ 特任助教

秋号では相関について書かせて頂きましたが、アンケートデータ解析の部は今回で最終回ということで、同じく変数と変数の関係である回帰についてお話させていただきます。復習になりますが、相関とは「2つの変数の関連の強さを見ること」、回帰とは「一方の変数の値からもう一方の値を導くこと」が目的でした。いずれにしても相関係数や回帰式という形で何らかの結果(スコア)は出てきますが、求めること自体が目的でないことは両者に共通しております。確か相関係数(r値)には、「大きさのみにとらわれず、因果関係や合理性、時間性などの観点から、本当に求める意味があるのか」を考える必要がありますね? 実は回帰式も全く同様で、強引に求めることは可能ですが、あくまでも求めた回帰式を用いて予測を行なうことが最終目的なのです。

#### (1) 回帰分析の考え方

1つの変数からもう一方の値を導くことを、単回帰分析といいます。導きたい方の値(y)のことを目的変数、導くために用いる変数のことを説明変数と呼びますので、くれぐれも逆にしないようにしましょう。目的変数yに対し説明変数xが一つだけの場合で、単純なものは図1のように直線で表示されます(勿論、単純な直線でない場合もあります)。データを受け取ったら、まずはExcelでもproc gplot(筆者の苦手分野なので省略?)でも構いませんので、グラフを書いてみましょう。SASのgplotは散布図を描くためのプロシジャですが、Excelの場合はグラフのオプションから「散布図」を選んで実行してみましょう。

さて、Excelをはじめとした統計ソフトウェアでは、回帰直線を自動的に計算して表示してくれますが、実際にはどのように計算しているのでしょうか? 高等学校数学の復習になってしまいますが、1点を通過する直線は無数にあり、2点を通過する直線は必ず1本に決定されます。ならば3点以上を通過すると…残念ながら全てが一直線にならない限りは全ての点を通過させることは

出来ません。そこで妥協案ということではありませんが、全ての直線からの距離の2乗の和が、一番小さくなるように直線を決定します。

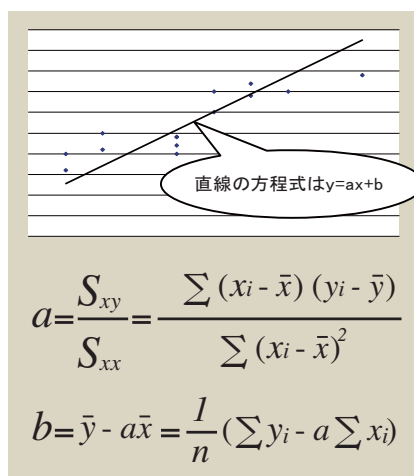


図1 回帰直線の求め方

#### (2) SASによる出力の見方と結果の理解

実行したら、次に回帰式のそれぞれの要素が有意であるかどうかを確認しなければなりません。変数TEMP(気温)の傾きはおおよそ2.87であり、この場合の帰無仮説は「H<sub>0</sub>:傾きは0である」です<sup>2)</sup>。さらにp<0.001と有意水準0.05を下回っているため帰無仮説を棄却、傾きが0ではないと見ることが出来ます。結果、気温は売上の個数に寄与していると言え、回帰式はy(売上個数)=2.87344x+43.13485ということになります。

2) 詳細は割愛しますが、標本の傾きaに対する母回帰係数βの95%信頼区間を求めます。もしも95%信頼区間が0をまたいでしまう場合、「傾きが0になる」可能性を否定できませんので、検定結果は有意でなくなるのです。

パラメータ推定値						
変数	ラベル	自由度	パラメータ推定値	標準誤差	t 値	Pr >  t
Intercept	Intercept	1	43.13485	13.01748	3.31	0.0062
TEMP	TEMP	1	2.87344	0.44323	6.48	<.0001

図2 気温(TEMP)の傾きと切片(Intercept)

では例題として、気温(TEMP:°C)からアイスクリームの売上個数(SALE:個)を予測する場合、SASでのオペレーションを考えてみましょう。この2つの変数を含むSASデータセットの名前をwk1として、まずは下記のプログラムNo1を実行します。最後のquit;ですが、これは「プロシジャの終了」という意味であり、無理に必要はありません<sup>1)</sup>。

1) glmやその他のいくつかのプロシジャの実行によって「SASの実行中」と表示されるのが気になるので、あくまで筆者が個人的に記載しているだけですが…。

傾きが2.87とは、気温が1度上昇するとアイスクリームの売上が2.87個増加することになるので、どうやら気温はアイスクリームの売上に関係があるということになります。つまり、気温が0°Cのときには、y(売上個数)=2.87344×0+43.1348543.13≒43.13個の売り上げがあり、1°C上昇ごとに2.87個の売り上げの増加が見込めます。この式を用いて、暑い日の仕入れ不足や寒い日の不良在庫の増加などを回避することこそが本当

#### <プログラムNo1> 単回帰分析

```
proc reg data=wk1 outest=wk2; /*傾き、切片をSASデータセットwk2に出力*/
  model SALE=TEMP; /*目的変数=説明変数*/
run;

quit;
```



の目的なのです。しかし実際の販売や企画、マーケティングなどの現場になりますと、何故か回帰式や相関係数を求めること自体が目的化してしまっている場面もたくさん見受けられますので、是非とも活用しましょう！

ならばTEMP(気温)のp値が有意でない場合( $p > 0.05$ の場合)はどのように判断すれば良いのでしょうか？ この場合は「 $H_0$ : 傾きは0である」を棄却できませんので、傾きが0である可能性も否定できません(もちろん、 $p < 0.05$ であっても完全に肯定はできませんが…)。もしも傾きがゼロの場合には、この回帰式は  $y = 43.13$  となり気温は売上には関係ない、つまり売り上げは一年中変わらず、暑かろうが寒かろうが一日43.13個であるということになります。さらには43.13の切片も有意でない場合は、暑かろうが寒かろうが43.13個という結論さえも出せなくなってしまう。

### (3) 回帰式の利用

要するに求めた回帰式は予測に使えるのだということ、求めっ放しではなく、もっと積極的に用いてみましょうということは分かりました。ですが実際のところ、この回帰式は実際どのくらい使い物になるのでしょうか？ アイスcreamの不良在庫を減少させるために、どのくらいこの店に貢献してくれるのでしょうか？ 回帰式が成立することは理解できたのですが、それがどのくらい使えるのかは定かではない、当てにならないというのでは困りものです。ということで、もう一点注目すべき場所は、図3のR2乗の大きさです。R<sup>2</sup>は「決定係数」もしくは「寄与率」と呼ばれるもので、たとえばこの場合であれば、気温はアイスcreamの売上の約77.8%を予測できるという意味になります。元となるrの値は気温とアイスcreamの売り上げ個数の相関係数で、0.7779の平方根でおおよそ $r \approx 0.8820$ となります。相関係数が0.88となりますと、非常に強い相関があることとなりますから、アイスcreamの売り上げ個数はかなり気温に左右されると言えるでしょう。さらに「調整済R2乗」なるものが表示されていますが、こちらはR<sup>2</sup>と表現され、

$$R^{*2} = 1 - \frac{\text{誤差の平方和/誤差の自由度}}{\text{全体の平方和/全体の自由度}}$$

によって示される値です。この値はR<sup>2</sup>値よりも小さくなりますが、予測の精度は調整済の方が高くなりますので、こちらの値を参照するようにして下さい。どうやら、「気温だけを見ていればおおよそ76%程度は予測できそうだが、気温以外の要因も24%程度存在する」ということで結論付けられそうです。その24%を知りたいというご意見もあるかも知れませんが、あくまで予測ですので、100%を説明することはまず不可能です。参考までに、もしも76%を説明できるモデルであれば、それは「極めて優秀なモデル」であると言えます。

誤差の標準偏差	3.67793	R2 乗	0.7779
従属変数の平均	127.28571	調整済 R2 乗	0.7594
変動係数	2.88950		

図3 調整済R2値の出力

### (4) 重回帰分析

駆け足で申し訳ございませんが、参考までに重回帰分析についてもお話させていただきます。重回帰分析は説明変数が2つ以上の場合であり、たとえばマンションの家賃が、駅からの徒歩距離、広さ、築年数、中心地からの距離など多くの要因によって決定されるような場合に用いられます。単回帰分析が $y = ax + b$ で表現されるのならば、重回帰分析は $y = ax_1 + bx_2 + cx_3 + \dots + mx_n + z$ のようなモデルで表現されます。Nは説明変数の数を意味しますが、ならば説明変数は多いほどいいのか…といえそうではありません。むしろ単回帰分析で説明が済むのであれば、単回帰分析を行なうに越したことは無いのです。先ほどの事例のように、仮に相関係数で0.88もあれば、重回帰を行なう必要などまずないでしょう。重回帰分析で気をつけるべきことは、説明変数間の相関が強いもの同士を同時に用いないということです。詳細は割愛致しますが<sup>3)</sup>、これは多重共線性(multi-collinearity: マルチコリニアリティ)と呼ばれる現象が発生するためです。これを防ぐためには、事前に

説明変数間の相関係数を求め、願わくば $r = 0.5$ 以上、もしくは $-0.5$ 以下(必ずしも明確ではありませんが)のペアは同時に用いないようにすることをお勧めします。特にrの値がいくつ以上ならダメというような明確なルールは存在しませんが、理想は各説明変数間の相関係数が $r = 0$ (まずありませんが)となることです。

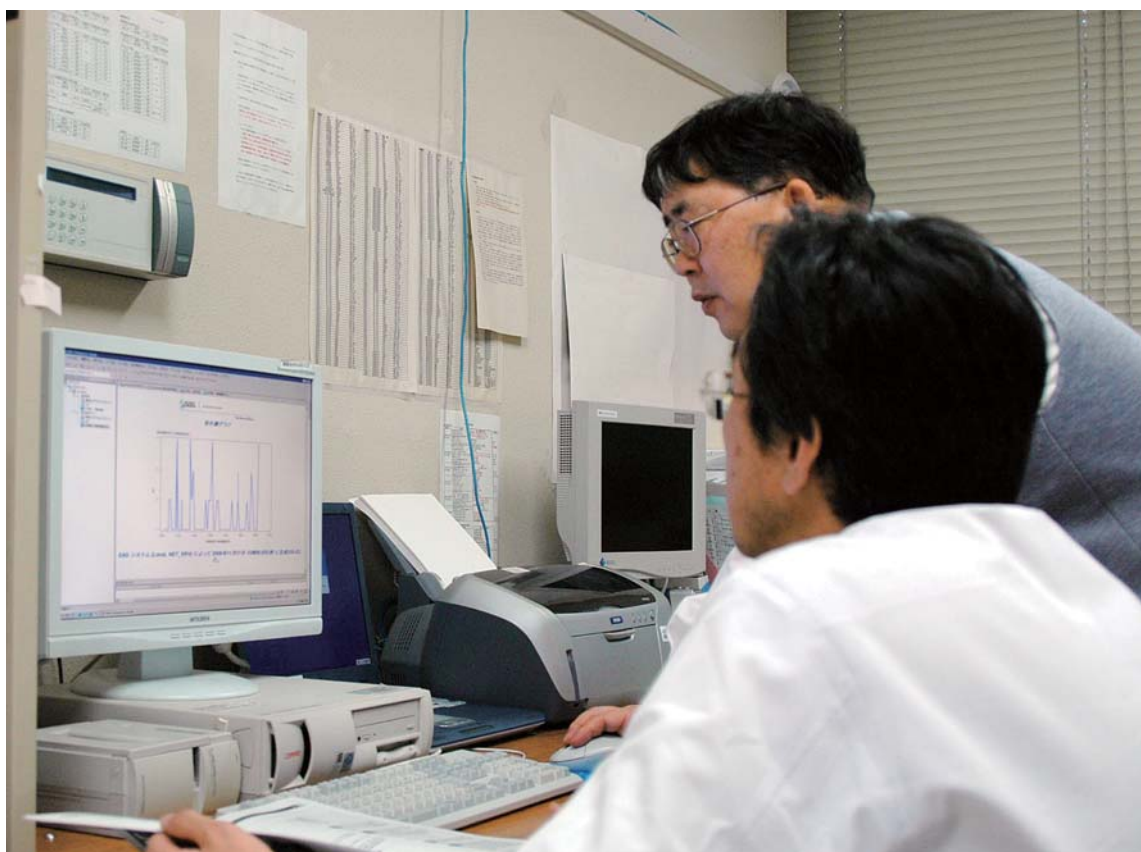
以上、アンケートデータの解析に必要と思われるお話について、おおよそ1年半に渡ってお話させて頂きました。p値の意味が、この連載でやっと分かったと言ってくださった読者の方のお声を頂きましたときには、改めて書いてよかったと思った次第です。回帰と相関の項目については、特にマーケティング業務の皆様を意識して書いたつもりです。商品の企画や開発、市場動向の予測など使いどころはたくさんござい

ますので、是非ともSASを用いて解析して頂ければと存じます。

また新シリーズにてお会い致しましょう。引き続き宜しくお願い申し上げます。

3) 多重共線性は、通常は偏回帰係数の符号の逆転により発見されることが多いのですが、説明変数間の相関行列を調べることで発見できます。

## カスタマー・リレーション

会田重信  
山形大学医学部情報基盤センター 技術専門職員

との問い合わせがありました。SAS契約に関与した大学側の担当教員とSAS Institute Japanの担当者間で協議がもたれ、飯田キャンパスのPCにもWindows版SASを導入することにゴーサインが出たことで、講習会を開催できる段階まで漕ぎ着けました。

**(4) 具体的な集客方法やその媒体について**

SAS講習会の日程は平成21年度5月と決まるとまずPRのため、大学院の講義の中で

上司(平成21年4月1日現在の学術情報基盤センター飯田分室長の一瀬教授)よりSASが簡単に利用できる講習会を開催することをPRしていただきました。その場で、詳細な内容と申込書が印刷したチラシを配布しました。

その後、5月の連休明けに電子メールで飯田キャンパス全員に向けた講習会案内を通知しました。返信メールで申込を受付することにしました。

**(5) 実際にセミナーで使われたパソコン環境や事前インストールなどの手順について**

講習会を開催する前までに、飯田キャンパスのWindowsパソコン約120台にSASソフトウェアをインストールしなければなりません。また、SAS講習会終了後に利用者に貸し出し出せるパソコンを準備

**II. 開催までの経緯****(2) 開催までの障壁になったこと**

講習会の開催は、当初2008年度中に実施する予定でしたが、開催されたのは2009年5月と、計画から実施に至るまで約1年5ヶ月を要してしまいました。

講習内容は2パターンで計画していること。1つ目の案は、特に遺伝子解析を得意とされる方から講義いただくこと。2つ目の案は、SASを一度も使っていない方を対象とすること。講師を推薦いただくことも含めてSAS Institute Japanの営業担当者に依頼しました。

回答は、1つ目の案については、講習内容は「SAS概要、医学分野でのSAS、遺伝子関連分野でのSASの使用」で、ご高名な先生より講義いただけること、費用は少しかさむこと。

2つ目の案については、講習内容は「SAS

概要、Enterprise Guide操作説明、事例紹介」という内容で、SAS Institute Japanの営業の方より講義いただけること、費用は交通費実費のみとのことでした。

当初は、1つ目の案の講習会を開催しようとしたが、大学内の各方面と具体的な調整が長引いてしまったため、2つ目の案の初心者向け講習会を実施することになりました。

**(3) 講習会を開催するにあたって次に生じた2つ目の障壁**

主たる講習内容は、「Enterprise Guide操作説明」となったのですが、Enterprise GuideはUNIX版SASには機能は無く、Windows版SASでないと動作しないことが判明しました。

この点についてはSAS Institute Japanから講習会のために一時的にWindowsパソコンにSASがインストールできないか

したいと考えました。この点についてはセンターでインストール費用と貸し出し専用パソコンを5台購入してもらえることになりました。

## (6) SAS以外のソフトでの講習会は頻繁に実施されているのか

今後はWindows版SASが利用してもらえる利用者を増やすための講習会を継続することになりました。他には卒業後に必須であると考えられるデジタル画像処理ソフトウェアの利用に関する講習は要望があれば随時実施しています。

### Ⅲ. 開催当日とその後

#### (1) 開催当日の問題点

参加申込を締め切ったところ申込者数は52名と予想以上の盛況ぶりに驚きました。また講習会の実際の参加者数は2日間を通して46名でした。講習会最中に発生した問題は以下の2つでした。

1. 講習会の事前案内において、パソコンを使うにあたり個人のUserIDとパスワード(アカウント)を用いるので覚えておくようお願いしてあったが、覚えていない方が多く見受けられた。このことはある程度事前に想定していた講習会用の臨時アカウントも用意していたが、一部のアカウントのパスワードに記載誤りがあったのでログイン時に混乱を生じた。

2. Enterprise Guide操作中(演算中)にアプリケーションが応答なしになったパソコンが3台程ありSASを再起動させた。

次に、講習会終了後に実施したアンケート調査の結果より注目すべき点を列挙します。

1. 今後も講習会を開催する場合どのような時間帯が参加しやすいか聞いたところ、夕方と回答した数が24名で約7割でした。

2. SASをいままでに使ったことがあるか聞いたところ、まったく使ったことがないと回答した数が31名で約9割、昔使ったことがあると回答した数が3名と、我々にとって大変ショッキングな結果となりました。

3. 主に使っている統計ソフトウェアについて聞いたところ、Excelと回答した数が19名で約5割強、次に同数でSPSSとStatViewが11名で約3割でした。StatViewは既にサポートが終了しているのに、早急にSASへ切替えてもらえるような体制づくりをしなければならぬことを痛感しました。

4. 今回の講習を受講してみた感想を聞いたところ、「なんとかついて行けた」、「理解できた」が同数で11名、約3割でした。約6割の方が『難なく講習できた』と答えられたのでほっとしました。

5. 今後どのような講習を受講したいか聞いたところ、「多変量解析(分析)全般について」が20名で最も多く、次に「統計的推定、統計的検定について」が17名、「遺伝子解析について」が13名の順でした。この結果により今後の講習会はどのような内容にすればよいか把握できました。

#### (2) 開催後の受講者や周囲からの反響や変化

講習会が終了して1ヵ月が経過したところからEnterprise Guideを介してSASが実際に利用されるようになりました。利用者が皆無にならずほっとしています。また、講習会で紹介してもらったテキストを大学内の書店で扱ってもらえるようお願いしました。多少高額な書籍だったのですが1ヶ月間ほどは全く売れませんでした。2ヶ月目になるとテキスト名を覚えてほしいと電話やメールがあり、書店から売れたことを聞かされて安堵しました。

#### (3) この会を通じての全体的な感想

我々は長年に渡り学内のSAS利用者から寄せられる質問に対して回答が出せない質問の時はSASテクニカルサポート部門に質問を仲介すると、常に的確な回答が返ってくるのでサポート体制には大変満足しておりました。この度の講習会では営業部門の皆様からハンズオンセミナー形式による講習をして貰い、講習中に参加者から質問ができれば的確な回答がその場で返されてくるので参加者の約6割の方から『難なく講習できた』と答えてもらえたと思います。一般的には、営業部門は技術的な

ことが分からないので技術部門に聞くようなことになりがちですが、SAS Institute Japanでは営業の方は講習もできてテクニカルサポートもできることに感心させられて、さらにSAS Institute Japanに対する信頼度が増しました。

営業部門の皆様には大変お世話になりました。講習会を終了することができました。感謝申し上げます。

また、当センターのSASの運用に関するさまざまな問題が発生する度に、的確なご支援をいただき、この場をお借りして感謝申し上げます。

#### (4) SASに今後期待すること

提供してきたプロダクトは、バージョンがアップされた段階で無くなってしまうようなことが無いようにしていただきたいと思います。



## Q&amp;A



- 複数変数のラベルおよび出力形式の一括削除
- 先頭オブザベーション内の変数をマクロ変数に割り当てる
- データにタブ文字が含まれるプログラムをバッチまたは %INCLUDEでサブミットするとデータセットが作成されない
- SAS Enterprise Guide Explorer上から SAS起動時ステートメントを設定する方法
- SAS 9.2以降のプログラムのリリース確認方法について
- マクロプログラムでの再帰処理について
- 多項分布に基づく擬似乱数の生成

**Q** 複数変数のラベルおよび出力形式を、一度に削除する方法はありますか？

**A** DATASETSプロシジャにて \_ALL\_ をご利用いただくことにより、全変数のラベルおよび出力形式が削除できます。

```
DATA sample;
  SET sashelp.prdsale;
RUN;

PROC DATASETS LIB=WORK NOLIST;
  MODIFY SAMPLE;
  ATTRIB _ALL_ LABEL='';          /* 全ラベルの削除 */
  FORMAT _ALL_ ;                  /* 全出力形式の削除 */
QUIT;
```

**Q** ある変数の先頭のオブザベーションをマクロ変数に割り当てたいのですがどのようにしたらよいでしょうか？

**A** CALL SYMPUT/SYMPUTXルーチンを使用すると、変数の値をマクロ変数に割り当てることができます。先頭のオブザベーションのみとするのであれば、DATASETオプションのOBS=オプションを利用して1オブザベーションだけ読み込むことで、余分な処理を省くことができます。

例

```
DATA sample;
  INPUT num @@;
CARDS;
1 3 5
;
RUN;
DATA _NULL_ ;
  SET SAMPLE(OBS=1);
  CALL SYMPUTX('first',num);
RUN;

/* 結果の確認 */
%PUT first=&first;
```

**Q** CARDSまたはDATALINESステートメントにタブ区切りの生データを記述したDATAステップをバッチサブミットや %INCLUDEで実行するとエラーになります。回避策はありますか？

**A** SASを起動し、プログラムエディタにてDATALINESステートメントにタブ文字を含むSASプログラムを読み込むと、自動的にタブ文字を空白に変換します。SASは空白を区切り文字として認識しますので、DATALINESステートメントに

タブ文字が使用されていても正常にデータセットが作成されます。しかし、バッチモードまたは %INCLUDEステートメントにて同じプログラムを実行するとプログラムエディタを介さないため、SASはタブ文字を区切り文字として認識せずエラーになります。そのため、バッチモードまたは %INCLUDEステートメントにてこのようなプログラムを実行するには、タブ文字を区切り文字として認識させる必要があります。

SAS 9.1以降のリリースを使用されている場合は、INFILEステートメントのEXPANDTABSオプションを使用するとタブ文字を空白8文字に変換しますので該当の事象を回避することができます。

但し、DATALINESステートメントにてデータの先頭、末尾に空白が含まれる場合、このオプションを使用するとデータに含まれる空白も区切り文字としてみなされ、これらの空白が削除されます。

記述

```
DATA dummy;          /* EXPANDTABSオプションを指定する */
  INFILE DATALINES EXPANDTABS;
  INPUT group sex data;
DATALINES;
1 1 100
1 2 110
2 1 150
2 2 160
3 1 180
3 2 200
;
RUN;
```

なお、SAS 9.1以前のリリースは、弊社テクニカルサポートの以下ページをご参照ください。

タブ区切りの外部ファイルの読み込み方法

<http://www.sas.com/jp/service/technical/faq/list/body/ba062.html>

**Q** SAS Enterprise Guide 4.1では、SAS Enterprise Guide Explorer上から起動時に実行する任意のプログラムを設定できましたが、SAS Enterprise Guide 4.2ではそのような機能が見あたりません。設定できないのでしょうか？

**A** SAS Enterprise Guide 4.2では設定箇所が変更されました。次の手順にて同様の設定が可能です。

1. メニューの「ツール」->「オプション」を選択します。
2. オプション画面左側から「SAS プログラム」を選択します。
3. 「その他のSASコード」内の「サーバーの接続時にSASコードをサブミットする」にチェックを入れ、「編集」ボタンを押します。
4. 実行するプログラムを記述します。

Q

SAS Enterprise GuideのHot Fixの一覧を確認したところ、一覧には4.2や4.21との記載があり、どちらを適用すればよいのか判りません。自分の環境がどのリリースか確認する方法はありますか？

A

SAS Enterprise Guide 4.2など、SAS 9.2を基本とした製品であれば、次の手順で、ご利用の製品のリリース情報を確認することが可能です。

※下記の作業は特定のフォルダへの書き込みが発生するなど、一定の権限を必要としますので、SASをインストールしたユーザーなど、管理者権限のあるIDで行なうことを推奨します。

1.ご利用の環境でのSASHOMEの場所を確認し、SASHOME/deploymntregディレクトリが存在するか確認します。フォルダが存在する場合は手順5から実行し、存在しない場合にはフォルダを作成して手順2から順に作業を行なってください。

※SASHOMEとは、SAS 9.2以降のSAS製品群をインストールする際のルートディレクトリです。標準では次の場所になります。

C:\Program Files\SAS

2. 次のSAS Notesを参照します。

<http://support.sas.com/kb/35/968.html>

- Notes内に記載されている“download sas.tools.viewregistry.jar”部分を選択するなどにて“sas.tools.viewregistry.jar”ファイルをダウンロードします。
- 上の「3」で入手した.jarファイルを、SASHOME/deploymntregディレクトリ下に保存します。
- エクスプローラなどから、“sas.tools.viewregistry.jar”をダブルクリックするなどにて実行します。
- 同一フォルダ内に“DeploymentRegistry.html”と“DeploymentRegistry.txt”のファイルが作成されます。どちらも同じ内容ですのでどちらかを開き、確認したい製品の部分を参照します。

次の例では、SAS Enterprise Guideのリリースが4.21であり、Hot FixのA32001が適用済みであることが確認できます。

```
Host: win
Product: eguide
Version: 4.21
Display Name: SAS Enterprise Guide
Display Version: 4.21
Hot Fix Entry: Hotfix A32001
```

上記情報は、SAS Enterprise Guideだけではなく、SAS Foundationや他のプロダクトでの詳細を確認する場合にも有用です。詳細につきましては、ファイルのダウンロード先として記載したSAS Notesにてご確認ください。

#### Usage Note 35968:

Finding out which product releases and hot fixes are installed on your system  
<http://support.sas.com/kb/35/968.html>

Q

マクロで再帰処理を記述することはできますか？

A

一般的に「再帰処理」とは関数の中で自分自身を呼び出す処理を指します。

SASのマクロ関数でも、マクロ関数の定義内でそのマクロ

関数自身を呼び出す、といった処理を記述することができます。以下のサンプルでは、再帰処理を利用して、指定した数値(正の整数)の階乗を求めています。

```
%MACRO test(n);
  %IF &n<=0 %THEN %DO;
    1
  %END;
  %ELSE %DO;
    %EVAL(&n*%test(%EVAL(&n-1)))
  %END;
%MEND test;

DATA _NULL_;
  x=%test(5);
  PUT x;
RUN;
```

/\* 5 の階乗を求める \*/

Q

試行回数が2以上の場合、多項分布に基づく擬似乱数は生成できますか？

A

二項分布の場合には、RANBIN関数、RAND関数、またはCALL RANBINルーチンにて試行回数を指定することができます。多項分布の場合には、RANTBL関数、CALL RANTBLルーチンを用いることができますが、試行回数は1となります。このため、試行回数が2以上の場合には、DATAステップにて、関数もしくはCALLルーチンを繰り返し、実行します。以下が一例となります。

#### 例

```
DATA ONE;
  ARRAY temp x1-x4;

  DO i=1 TO 5;
    /* サンプル数 (5) を指定 */
    /* 各サンプルごとに初期化 */
    DO j=1 TO DIM(temp); temp{j}=0; END;
    /* 試行回数 (10) を指定 */
    DO j=1 TO 10;
      /* 試行回数 1 の擬似乱数を生成 */
      xc = RANTBL(123,0.1,0.2,0.3,0.4);
      temp{xc}=temp{xc}+1;
    END;
    /* サンプルごとに出力 */
  OUTPUT;
END;
RUN;
```

SAS 9.2 SAS/IMLでは、多変量分布に基づく擬似乱数の生成に対応している、以下のモジュールが新たに追加されています。

RANDDIRICHLET	- Dirichlet 分布に基づく乱数生成
RANDMULTINOMIAL	- 多項分布に基づく乱数生成
RANDMVT	- 多変量 t 分布に基づく乱数生成
RANDNORMAL	- 多変量正規分布に基づく乱数生成
RANDWISHART	- Wishart 分布に基づく乱数生成

多項分布の場合には、以下の指定となります。

#### 例

```
PROC IML;

  /* 初期シード値の指定 */
  CALL RANDSEED(456);
  /* サンプル数 5、試行回数 10 の乱数生成 */
  x=RANDMULTINOMIAL(5,10,{0.1,0.2,0.3,0.4});
  /* 結果を表示 */
  PRINT x;
QUIT;
```

# SAS Training

## SAS トレーニングのお知らせ



### 特別トレーニング・コースの開催のご案内

#### ●「SASプログラミング:SAS 9.2拡張点」コース (1日間)

[日 程]

2010年1月18日(月) 10:00 ~ 17:00 (東京会場)

[価 格]

57,750円(税込)／チケット捺印数:1

[受講対象]

SAS 9.2 Foundationの新機能を学習したい方

[前提知識]

「SASプログラミングⅡ」を受講済みか、同程度の知識のある方  
本コースは、SAS初心者には適切ではありません。特に次のことができる必要があります。

- ・オペレーティングシステムファイル構造を理解し、基本的なオペレーティングシステムのタスクを実行することができる
- ・プログラミングロジックの概念を理解している
- ・DATAステップのコンパイルと実行処理を理解している
- ・外部ファイルからSASデータセットを作成するために入力のいくつかの方法を利用できる
- ・SASデータライブラリにアクセスするためのSASソフトウェアを使用できる
- ・SAS日付値の作成と利用ができる
- ・SASデータセットの読み込み、縦結合、マージ、マッチマージ、インターリーブができる
- ・DROP=, KEEP=, RENAME=データセットオプションが利用できる
- ・複数の出力データセットが作成できる
- ・配列処理と繰り返してデータを処理するDOループが利用できる

[学習内容]

本コースでは、SAS 9.2 Foundationの新機能を学習します。  
このコースは、演習問題や例題を通して、実業務でのシナリオで実施する、上級機能に焦点を当てています。

[コンテンツ]

- FCMPプロシジャでの関数とサブルーチンの作成
  - ・PROC FCMPを使用した関数の作成
  - ・PROC FCMPを使用したサブルーチンの作成
- XMLファイルとSAS 9.2 XML LIBNAMEエンジンでの作業
  - ・SAS XML92 LIBNAMEエンジン イントロダクション
  - ・XMLファイルへの書き出し
  - ・オプションでのXML92 LIBNAMEのカスタマイズ
  - ・XMLファイルの読み込み
- SASコードアナライザでの作業(SCAPROCプロシジャ)
  - ・新しいSASコードアナライザプロシジャの概念入門
  - ・SASコードアナライザでの情報のキャプチャ
  - ・キャプチャするコードのタイプの識別
  - ・グリッドコンピューティングでのジョブの最適化
- 新規チェックポイント/リスタート機能でのエラー処理とデバッグ
  - ・チェックポイント/リスタート機能の概念入門
  - ・チェックポイントのキャプチャ
  - ・リスタート機能での作業

#### ■ SAS言語とマクロ機能の拡張

- ・イントロダクション
- ・拡張されたプロシジャ
- ・一部の関数
- ・ステートメント
- ・オプション

#### ■ ODSとグラフの拡張

- ・イントロダクション
- ・SAS 9.2でのODS統計グラフ
- ・ODSグラフィックエディタ
- ・グラフテンプレート言語での作業
- ・ODSでの新規グラフプロシジャでの作業

#### ■ その他の重要な変更と拡張入門

- ・SQLプロシジャ
- ・ハッシュオブジェクト
- ・言語上の並べ替え

#### ●「SAS 9.2 SAS/STAT 拡張点」コース (1日間)

[日 程]

2010年3月12日(金) 10:00 ~ 17:00 (東京会場)

[価 格]

57,750円(税込)／チケット捺印数:1

[受講対象]

SAS 9.2における、Base SASやSAS/STATのODS統計グラフや、基本統計とカテゴリカルデータ解析の新機能を習得されたい方

[前提知識]

下記2コースを受講済みか、同程度の知識のある方

- ・「SASによる統計解析Ⅰ」
- ・「SASによる回帰分析Ⅰ」

また、SASプロシジャの構文を理解できる方

後半部を理解するためには、SASを利用した度数表の解析やロジスティック回帰分析の経験がある方

[学習内容]

- ・PROC UNIVARIATE(評価版)、PROC TTEST、PROC GLM、PROC REG、PROC LOGISTIC、PROC GENMOD(自己学習)、およびPROC FREQでODS統計グラフを生成します。
- ・PROC UNIVARIATEの新規ステートメントPPLOTを使用します。
- ・PROC TTESTでTOST同等性検定を行います。
- ・REGプロシジャでプロットを作成して、残差やモデルの適合度テーブルについて説明します。
- ・LOGISTICプロシジャのROCONTRASTステートメントを使用して、複数のROCモデルを比較します。
- ・交互作用が存在するときのオッズ比の計算や、オッズ比に関するプロットを行います。
- ・削除診断(deletion diagnostics)の実行、GEEモデルに対するプロット、ゼロ強調(zero-inflated)Poisson回帰モデルの実行、AICとQICというモデルの当てはまりに関する統計量を計算します。(自己学習)



## [コンテンツ]

- Base SASの統計プロシジャ
  - ・ODS統計グラフへの入門
  - ・PROC UNIVARIATEの新しいオプション

- T検定と分散分析
  - ・PROC TTESTの新しいオプション
  - ・分散分析
  - ・分散分析におけるポストホック検定
  - ・交互作用を含む分散分析

- 線形回帰分析
  - ・相関
  - ・回帰分析とREGプロシジャ
  - ・REGプロシジャにおけるモデル構築
  - ・回帰診断

- PROC FREQの新機能
  - ・PROC FREQにおけるODS統計グラフ
  - ・同等性検定と非劣性検定

- PROC LOGISTICの新機能
  - ・PROC LOGISTICにおけるODS統計グラフ
  - ・Firthのペナルティ付き最尤法

- PROC GENMODの新機能(自己学習)
  - ・PROC GENMODにおけるODS統計グラフ
  - ・ゼロ強調Poissonモデル

【注意】本コースは、英文書籍を使用し日本語で説明を行います。

## ●「医学研究におけるベイズ統計学：SAS 9.2によるアプローチ」コース (2日間)

## [日 程]

2010年3月18日(木)～19日(金) 10:00 ～ 17:00 (東京会場)

## [価 格]

147,000円(税込)／※チケットのお取り扱いはありません。

## [受講対象]

ベイズ統計を臨床試験のデザインに用いる方でSASを用いてベイズ流データ解析を実施したい方

## [前提知識]

SASによる回帰分析や生存時間解析の基本的な知識のある方  
基本的な生物統計学の知識のある方

- ・確率分布を知っている
- ・条件付分布、周辺分布を知っている
- ・ベイズの定理という言葉聞いたことがある

## [学習内容]

ベイズ統計の基礎を解説し、ベイズ統計の実際の臨床試験への適用事例を紹介します。

SASによる実際の解析手順、結果の解釈を説明します。

## [コンテンツ]

- ベイズ統計の基礎
  - ・ベイズ定理、(conjugate) 事前・事後分布、おもな確率分布、シングルパラメータモデル、マルチパラメータモデル

- ベイズモデル
  - ・回帰モデル、階層モデルHierarchical model、ベイズ流計算を行うSASプロシジャの紹介
- 計算方法
  - ・事後分布の計算、モンテカルロシミュレーション、MCMC、SASでの実現方法
- 事例紹介
  - ・ベイズ流試験デザインとSASを用いたデータ解析の事例紹介

SAS Institute Japan株式会社では、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。

コース内容・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/jp/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: 03-6434-3690

F A X: 03-6434-3691

E-mail: JPNTraining@sas.com

# SAS Publications

## SAS Pressシリーズ売り上げTOP10のご紹介

SAS Pressシリーズとは、世界中のSASユーザーが作成した専門的なSASマニュアルです。  
その中でも特にユーザーから人気の高いマニュアルをご紹介します。

順位	ISBN	タイトル
1	978-1-59994-725-9	The Little SAS® Book: A Primer, Fourth Edition
2	978-1-59047-333-7	The Little SAS® Book: A Primer, Third Edition
3	978-1-59994-165-3	Learning SAS® by Example: A Programmer's Guide
4	978-1-60764-038-7	Radical Action for Radical Times: Expert Advice for Creating Business Opportunity in Good or Bad Economic Times
5	978-1-59047-703-8	Predictive Modeling with SAS® Enterprise Miner™ : Practical Solutions for Business Applications
6	978-1-59994-657-3	The Global English Style Guide: Writing Clear, Translatable Documentation for a Global Market
7	978-0-470-46168-6	Credit Risk Assessment: The New Lending System for Borrowers, Lenders, and Investors
8	978-1-55544-279-8	Survival Analysis Using SAS®: A Practical Guide
9	978-1-59994-089-2	The Little SAS® Book for Enterprise Guide® 4.1
10	978-1-59047-508-9	CRM Segmentation and Clustering Using SAS® Enterprise Miner™

マニュアルの詳細は、下記URLよりご覧いただけます。

<http://www.sas.com/apps/pubscat/complete.jsp>

### ■ 購入方法

上記のマニュアルは、Amazon.co.jp (<http://www.amazon.co.jp/>) にて購入いただけます。

### ■ 日本語マニュアルについて

SAS Institute Japan株式会社では、現在日本語マニュアルを販売しております。

詳細は下記URLをご参照ください。

<http://www.sas.com/jp/manual/index.html>

また、日本語マニュアルの詳細および購入方法については、上記URLをご参照いただくか、下記マニュアル販売係までお問い合わせください。

### マニュアル販売係

T E L: 03-6434-3690

F A X: 03-6434-3691

E-mail: [JPNBooksale@sas.com](mailto:JPNBooksale@sas.com)

# Latest Releases

## 最新リリース情報

### PCプラットフォーム

Windows版 SAS 9.1.3 / 9.2  
64-bit Windows(Itanium)版 SAS 9.1.3 / 9.2

### メインフレームプラットフォーム

IBM版(OS/390,z/OS) SAS 9.1.3 / 9.2

### UNIXプラットフォーム

SunOS/Solaris版 SAS 9.1.3 / 9.2  
HP-UX版 SAS 9.1.3 / 9.2  
HP-UX(Itanium)版 SAS 9.1.3 / 9.2  
AIX版 SAS 9.1.3 / 9.2  
Linux(Intel) 版 SAS 9.1.3 / 9.2



### SAS Technical News入手方法

SAS Technical Newsは、右記のURLから入手できます。

<http://www.sas.com/jp/periodicals/technews/index.html>

発行:SAS Institute Japan株式会社



**STN**  
**SAS Technical News**

**WINTER 2010**

■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:03-6434-3680 FAX:03-6434-3681



SAS Institute Japan株式会社

本社  
〒106-6111  
東京都港区六本木6-10-1  
六本木ヒルズ森タワー 11F  
Tel 03(6434)3000  
Fax 03(6434)3001

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

[www.sas.com/jp](http://www.sas.com/jp)

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、一般の各社の登録商標です。  
Copyright©2010, SAS Institute Inc. All rights reserved.