

SAS Technical News

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



# **AUTUMN 2009**

10.00	4

SAS® 9.2における 分析機能の拡張点: Base SAS®

# SAS Academic News ····· 10

- 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 ······	18

# 最新リリース情報 ・・・・・・・・・ 20

# 特集

# SAS® 9.2における 分析機能の拡張点: Base SAS®

SAS 9.2は、これまでの計算機能を拡張・強化するとともに、プログラムの 記述を簡略化できるようになっています。たとえば、データをビジュアライ ズする際には簡略なプログラムで実行できるなど、より解析者の視点に 立った使いやすいバージョンに仕上がっているのではないかと自負して います。

今秋号の特集では、前号に引き続きSAS 9.2の新規拡張点について掲載。 SASシステムの根幹であるBase SASが有する記述統計において追加さ れた関数、さらに独自に作成できる関数やその関数のデータ ステップでの利用法など、ユーザーがそれぞれのゴールに近 づきやすくできるよう、豊富なサンプルと出力例を交えて紹 介していきます。





# Base SASの分析に関連する新機能

#### 1.1 関数とサブルーチン

分析の場面で利用できる関数やサブルーチンが数多く追加されています。

#### ● 双曲線関数の逆関数

双曲線関数(Hyperbolic function)に対する逆関数が用意されました。 ARCOSH、ARSINH、およびARTANH関数は、それぞれCOSH、SINH、 およびTANHの逆関数です。

#### ● lpノルムの計算

LPNORM関数は、第1引数をpとして、第2引数以降に対するlpノルムを 計算します。つまり、引数の絶対値をp乗してからその和を計算し、それ のp乗根を結果として返す計算を行ないます。SUMABS関数とEUCLID 関数は、それぞれp=1とp=2のときのLPNORM関数に相当します。

#### 例

```
x=LPNORM(1.5,1,-4); /*x=(1**1.5+4**1.5)**(1/1.5)=4.3267*/
y=SUMABS(3,-2,5);
                                            /*y=3+2+5=10*/
z=EUCLID(4,-2,3,1); /*z=sqrt(4**2+2**2+3**2+1**2)=5.4772*/
```

 $l_{\infty}$ ノルムはたとえばpとして非常に大きな値を指定する(例1)、または MAX関数とMIN関数を使って計算できます(例2)。

/\*例1\*/ 1\_infinity\_norm1=LPNORM(100000000,OF x1-x10); /\*例2\*/ 1 infinity norm2=MAX(OF x1-x10)<>-MIN(OF x1-x10);

#### ● 最大公約数と最小公倍数

GCD関数は引数の最大公約数を、一方LCM関数は最小公倍数を返し ます。

#### 例

x=GCD(12,48,16); /\*x=4\*/ y=LCM(4,6,18); /\*y=36\*/

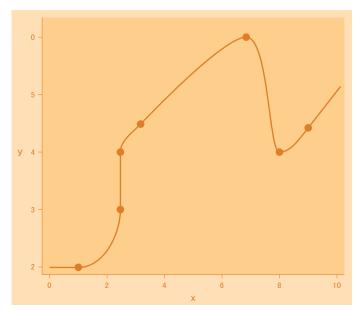
#### ● 単調性を保つスプライン補間

MSPLINT関数は、可能な限り単調性を保ちつつ、ノット(節点)として与え た点をすべて通るスプライン曲線の座標を計算します。

次のプログラムは、SASデータセットknotsに含まれる節点をベースとし て、単調性を保つスプライン曲線を作成する例です。



```
DATA knots;
  INPUT x y;
DATALINES:
1.0 2.0
2.5 3.0
2.5 4.0
3.0 4.5
7.0 6.0
8.0 4.0
9.0 4.5
RUN:
DATA plots;
  ARRAY _x{7};
  ARRAY _y{7};
  DO i=1 TO 7:
    SET knots;
    _x{i}=x;
    _y{i}=y;
    OUTPUT;
  END:
  DO x=0 TO 10 BY 0.01;
    y2=MSPLINT(x,7,of _x{*},of _y{*});
    OUTPUT:
  DROP _x: _y: i;
RUN;
PROC SGPLOT DATA=plots NOAUTOLEGEND;
  SCATTER Y=y X=x
/ MARKERATTRS=(SYMBOL=CIRCLEFILLED SIZE=8PT):
  SERIES Y=y2 X=x / LINEATTRS=(PATTERN=1);
RUN:
```



MSPLINT関数から作成した曲線の例

#### ● 順列や組み合わせの対数、およびべき乗の対数

ある種の計算の過程では、順列や組み合わせの数の対数、また階乗 の対数を求めなくてはならないものの、浮動小数点数の桁あふれの現 象が心配な場合があります。LPERM、LCOMB、およびLFACT関数は、 それぞれPERM、COMB、およびFACT関数で返される値の対数を直 接計算します。

#### ● 対数

LOG1PX関数は、引数xに対してloge(1+x)を計算します。これは、LOG 関数を使用したLOG(1+X)という演算と等価ですが、引数xが0に極め て近いときにはより良い精度をもつ値を返します。

#### ● Grayコード

Grayコードは、たとえば0と1からなる行を考えたとき、直前の行から1つ だけ要素を反転させることによって続きの行を作成した一連の結果とし て得られます。GRAYCODE関数とサブルーチンのCALL GRAYCODE は、Grayコード生成のために利用できます。

以下の例では、0と1からなる要素数が3の配列をGrayコードに従って 作成しています。

```
DATA _NULL_;
  ARRAY x[3];
  k=-1:
  DO i=1 TO 2**DIM(x);
      rc=GRAYCODE(k, OF x[*]);
      PUT i x[*] rc=;
  END:
RUN;
```

```
ログへの出力)
1 0 0 0 rc=0
2 1 0 0 rc=1
3 1 1 0 rc=2
4 0 1 0 rc=1
5 0 1 1 rc=3
6 1 1 1 rc=1
7 1 0 1 rc=2
8 0 0 1 rc=1
```

変数rcには、前のステップから置き換えられた場所を示す値が保存さ れています。

# ● ヨーロピアン・オプションの価格

> MARGRPTPRC

いくつかの代表的なモデルに対して、コール価格とプット価格を計算す る関数が用意されました。

Blackモデルにもとづくコール価格 > BLACKCLPRC Blackモデルにもとづくプット価格 > BLACKPTPRC Black-Scholesモデルにもとづくコール価格 > BLKSHCLPRC > BLKSHPTPRC Black-Scholesモデルにもとづくプット価格 Garman-Kohlhagenモデルにもとづくコール価格 > GARKHCLPRC Garman-Kohlhagenモデルにもとづくプット価格 > GARKHPTPRC Margrabeモデルにもとづくコール価格 > MARGRCLPRC Margrabeモデルにもとづくプット価格



#### ● 財務関数

FINANCE関数は、金融や財務に関するさまざまな計算をサポートしま す。これは、Microsoft EXCELで用意されているすべての財務関数をシ ミュレートしたものであり、第1引数はEXCELの関数名と対応しています。 次のプログラムは、EXCELのXIRR関数と等価な計算を行なう例です。

```
DATA NULL;
  INFORMAT d1-d5 NLDATE14.;
  INPUT v1 d1 v2 d2 v3 d3 v4 d4 v5 d5 @@;
  r = FINANCE('XIRR', of v1-v5, of d1-d5, 0.1);
 PUT r=:
DATALINES;
-10000 2008年01月01日
  2750 2008年03月01日
  4250 2008年10月30日
  3250 2009年02月15日
  2750 2009年04月01日
RUN:
```

ログのウィンドウには、rの値が表示されます。

```
r=0.3733625335
```

#### ● 順列と組み合わせ

変数値の全組み合わせを作成する関数とサブルーチンが複数用意さ れました。その利用目的によって、適宜使い分けることができます。

ALLCOMB関数は変数値から組み合わせを作成しますが、値の交換 をなるべく行なわないようなアルゴリズムにもとづいています。CALL ALLCOMBは、そのサブルーチン版です。

次のプログラムは、ALLCOMB関数の使用例とその結果です。

```
DATA NULL;
 ARRAY x[5] $6 ('朝日' 'いろは' '上野' '英語' '大阪');
 n=5; k=2;
                   /*5つの要素から2つを取り出す組み合わせ*/
 DO j=1 TO COMB(n,k)+1;
   rc=ALLCOMB(j, k, OF x[*]);
                                     /*ALLCOMB関数*/
   PUT j= rc= x1= x2=;
 END;
RUN:
```

ログのウィンドウには、次のように表示されます。

```
i=1 rc=0 x1=朝日 x2=いろは
j=2 rc=2 x1=朝日 x2=大阪
j=3 rc=2 x1=朝日 x2=英語
j=9 rc=2 x1=上野 x2=大阪
j=10 rc=1 x1=英語 x2=大阪
j=11 rc=-1 x1=英語 x2=大阪
```

一方、サブルーチンのCALL ALLCOMBIは変数値そのものではなく、そ のインデックスを返します。そのため、処理はCALL ALLCOMBより高 速です。なお、ALLCOMBI関数は用意されていません。LEXCOMB関数 とCALL LEXCOMBサブルーチンは、変数値から結果として異なる組み 合わせを辞書式に作成していきます。(ただし、日本語はその限りでは ありません。これは、LEXで始まる他の関数やサブルーチンでも 同様です。)結果のインデックスを返すバージョンが、LEXCOMBI関数と CALL LEXCOMBIサブルーチンです。

次のプログラムは、LEXCOMB関数の使用例とその結果です。

```
DATA NULL;
 ARRAY x[5] $6 ('あさひ' 'あさひ' 'うえの' 'えいご' 'えいご');
                 /*5つの要素から3つを取り出して作成する順列*/
  n=5;k=3;
  DO UNTIL(rc<0);
   rc=LEXCOMB(j, k, of x[*]);
                                     /*LEXCOMB関数*/
   PUT rc= j= x1= x2= x3=;
 END;
RUN:
```

ログのウィンドウには、以下のように表示されます。

```
rc=1 j=1 x1=あさひ x2=あさひ x3=うえの
rc=3 j=2 x1=あさひ x2=あさひ x3=えいご
rc=2 i=3 x1=あさひ x2=うえの x3=えいご
rc=2 j=4 x1=あさひ x2=えいご x3=えいこ
rc=1 j=5 x1=うえの x2=えいご x3=えいご
rc=-1 j=6 x1=うえの x2=えいご x3=えいご
```

一方、変数値の順列を作成する機能も追加されています。

ALLPERM関数は、利用する変数値のすべての順列を作成したいときに 利用できます。なお、同等の機能を持つサブルーチンのCALL ALLPERM はSAS 9.1から存在していました。一方、LEXPERM関数は変数値から結 果として異なる順列を作成して、辞書式で結果を返します。LEXPERK関 数は、m個の変数値からk個(k<=m)を取り出して同じ処理を行ないます。 それぞれに対応したサブルーチンが、CALL LEXPERKとCALL LEXPERMサブルーチンです。

# 1.2 ユーザー定義の関数作成機能

# 1.2.1 FCMPプロシジャ

FCMPプロシジャで作成した関数やサブルーチンは、SAS 9.1.3では一部 の分析プロシジャにおいて利用可能でしたが、SAS 9.2からはDATAス テップでも使用できます。

#### ● 関数の作成と利用

関数を作成するときに使用するFUNCTIONステートメントでは、ユーザー 定義の関数名と引数を与えます。ENDSUBステートメントまでの間に計算 処理のステートメントを記述して、戻り値をRETURNステートメントで指定 します。

以下のプログラムは、いわゆるBox-Cox変換を行なう関数の作成例です。

```
PROC FCMP OUTLIB=SASUSER.myfunc.test1;
  FUNCTION boxcox(x,lambda);
    IF x>0 THEN DO;
      IF ABS (lambda)>1e-4 THEN f=(x**lambda-1)/lambda;
      ELSE f=LOG (x):
    END;
    ELSE f=.;
    RETURN (f);
  ENDSUB;
RUN;
```

1行目のPROC FCMPステートメントでは、定義する関数の情報を保存するOUTLIB=オプションを指定しています。この例では、SASUSERライブラリのmyfuncというSASデータセットにエントリ名test1で保存されます。

2行目のFUNCTIONステートメントでは関数"boxcox"を定義していて、関数の引数であるxとパラメータlambda( $\lambda$ )が引数として与えられています。

3行目以降では、計算処理の内容を具体的に記述しています。パラメータlambdaが0から離れているときは指数演算を行ない、0に十分近ければ対数を計算しています。

8行目のRETURNステートメントでは、戻り値(f)を定めています。

定義した関数を使用するときは、OPTIONSステートメントのCMPLIB= オプションで、関数の情報を保存したSASデータセット名を事前に指定 しておく必要があります。

# ● サブルーチンの作成と利用

サブルーチンを作成する場合は、SUBROUTINEステートメントを使用します。次の例では、2つの変数値を交換するサブルーチン"swap"を作成して、それを実行しています。

```
PROC FCMP OUTLIB=SASUSER.mvsub.a;
  SUBROUTINE swap(a,b);
    OUTARGS a,b;
    temp=b;
    b=a;
    a= temp;
  ENDSUB;
RUN:
OPTIONS CMPLIB=SASUSER.mvsub;
DATA null;
  x=3; y=2;
                                          /*x=3 y=2*/
  CALL swap(x, y);
                         /*作成したサブルーチンswapの実行*/
                                          /*x=2 y=3*/
  PUT x= y=;
RUN:
```

#### ● FCMPプロシジャの中で利用できる機能

FCMPプロシジャにおける関数やサブルーチンの定義では、さまざまな機能を利用することができます。

#### [ 再帰呼び出し ]

関数やサブルーチンの定義中に、それ自身を呼び出す再帰処理が実現可能です。

次のプログラムは、階乗の計算を行なう関数factorialの作成例です。階乗の計算そのものはFACT関数で実現できますが、ここではFCMPプロシジャで再帰呼び出しを利用しています。

#### [SASデータセットへの入出力]

READ\_ARRAY関数を使うと、SASデータセットから配列としてデータ値を取り込むことができます。逆にSASデータセットへデータ値を出力するには、WRITE\_ARRAY関数が利用可能です。

#### [基本的な行列演算]

行列の加算や乗算、逆行列や転置行列の作成、またCholesky分解など、分析の計算処理において頻繁に登場するような行列演算を行なう CALLルーチンが用意されています。

#### [ 方程式の解法 ]

SOLVE関数を利用すると、方程式を満たす実数解を(それが存在すれば)見つけ出すことができます。たとえば、FCMPプロシジャのドキュメントにはインプライド・ボラティリティの計算例が記載されています。

次のプログラムは、3次方程式x3+3x2+4x+2=0の解を見つけ出す例です。

```
PROC FCMP;

FUNCTION equation3(a,b,c,d,x);

f=a*x**3+b*x**2+c*x+d;

RETURN (f);

ENDSUB;

y=0;

x = SOLVE("equation3",{.},y,1,3,4,2,.);

PUT x=;

RUN;
```

#### [SASマクロや外部SASプログラムの実行]

RUN\_SASMACRO関数を利用して、FCMPプロシジャの外で定義されているSASマクロを取り込むことができます。その処理を行なって作成した関数をDATAステップで使用すれば、DATAステップの中で別のDATAステップやプロシジャを実行するような操作を実現できます。

また、RUN\_SASFILE関数を使うと、SASファイルを関数やサブルーチンの中に取り込むことができます。

以上のように、FCMPプロシジャは関数やサブルーチンを作成するだけにとどまらず、これまでのSASでは難しかった処理を効率的に行なうことができます。ただし、以下の点に留意してください。

- ・同等のSASコードを直接DATAステップに記述した場合と比べて、FCMP プロシジャによる関数やサブルーチンの実行速度は一般に低下します。
- ・配列を引数として渡すときは、一時配列の形態でなくてはなりません。



#### 1.2.2 SAS Fcmp Function Editor

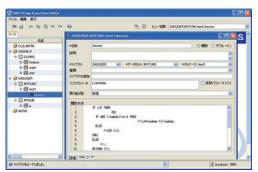
SAS Fcmp Function Editorは、ユーザー定義の関数とサブルーチン を管理するJavaアプリケーションです。

このアプリケーションは、メニューから「ソリューション」>「データ解析」 と選んで現れるオプションリストの最下段から起動できます。



SAS Fcmp Function Editorの起動方法

アプリケーションにおける左のペインでは、FCMPプロシジャで作成さ れた関数やサブルーチンの一覧が表示されており、その上でダブルク リックをすると、それに対応する機能の詳細を含むウィンドウが立ち上 がります。



前項で作成した関数boxcoxを開いたときに現われるウィンドウ

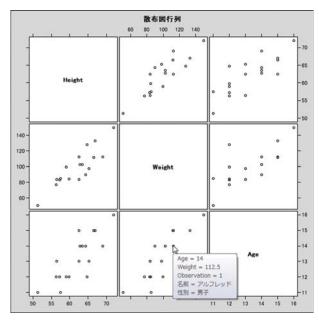
このウィンドウの中では、作成されていた関数とサブルーチンの修正 や、新規でそれらを作成することができます。FCMPプロシジャを直接 記述することに比べると、より簡単かつ直接的にユーザー定義の関数 やサブルーチンを編集することができるでしょう。

# 1.3 既存分析プロシジャにおける新規機能 1.3.1 CORRプロシジャ

IDステートメントが機能するのは、HTMLファイルを出力先としていて、同 時にODS統計グラフを利用した場合です。このとき、IDステートメントで変 数を指定すると、画像ファイルの点の上にマウスを置けば、それらの変数 値が「チップ」として表示されます。ただし、ODS GRAPHICSステートメントで IMAGEMAPオプションを指定しておかなくてはなりません。

```
ODS GRAPHICS ON / IMAGEMAP:
ODS HTML;
PROC CORR DATA=sashelp.class PLOTS=(MATRIX);
  ID name sex;
  VAR height weight age;
```

```
RUN:
ODS HTML CLOSE:
```



散布図行列でチップを表示させた例

#### 1.3.2 FREQプロシジャ

プロシジャ出力と統計機能の両方に対してさまざまな拡張が行なわれま した。

# ● クロス集計表のカスタマイズ機能

クロス集計表をカスタマイズするテンプレートが用意されました。 TEMPLATEプロシジャに追加されたDEFINE CROSSTABS、DEFINE CELLVALUE、またはCELLVALUEステートメントを利用すると、それぞ れクロス集計表に対するテンプレートの作成、集計表のセルにおける出 力の形式の定義、または出力の並びの制御が可能となります。これは、 多くのユーザーの皆様から要望のあった機能です。

次のプログラムは、FREQプロシジャによるクロス集計表において、割 合の出力に対してユーザー定義の出力形式(フォーマット)を与えて表 示する例です。また、「列のパーセント」に相当する出力を抑制しており、 行と列の変数はヘッダーで表示しています。

```
DATA class;
 INPUT gender$ age @@;
DATALINES;
女性 21 女性 23 女性 26 女性 43
女性 32 女性 54 女性 30 女性 40
男性 23 男性 32 男性 38 男性 30
男性 24 男性 38
RUN;
PROC FORMAT;
 PICTURE pctfmt (ROUND) OTHER='009.9%';
 VALUE agefmt LOW-29='30才未満
               30-39="30代"
               40-HIGH="40才以上";
RUN:
```

```
PROC TEMPLATE:
 DEFINE CROSSTABS Base.Freq.CrossTabFreqs;
    DEFINE HEADER tabletitle;
     TEXT "行変数:" _ROW_NAME_ " 列変数:" _COL_NAME_;
    CELLVALUE Frequency Percent Rowpercent;
    DEFINE Frequency;
     FORMAT=8.;
     HEADER='頻度';
    END;
   DEFINE Percent;
      FORMAT=pctfmt.;
     HEADER='割合';
    DEFINE Rowpercent;
     FORMAT=pctfmt.;
     HEADER='行における割合';
   END:
 HEADER tabletitle;
 END;
RUN:
PROC FREQ DATA=CLASS;
 TABLES gender*age;
 FORMAT age agefmt.;
RUN;
```

	行変数:ge	nder	列変数:age		
   頻度   割合	 				
行における割合					
女性		2	3		
	: :		21.4%		
	37.5%   +	25.0% 	37.5% 	-	
男性	2	4	0	6	
	14.3%	28.6%	0.0%	42.9%	
			0.0%		
 合計	+1 5	6	·+	14	
H #1	_	-	21.4%	100.0%	

この機能に関する詳細は、"SAS 9.2 Output Delivery System User's Guide "の"TEMPLATE Procedure: Creating Crosstabulation Table Templates"の項をご参照ください。

また、米国SAS Institute Inc.のWebサイトで公開されている下記の文献では、いくつかの具体例が記載されています。

# "Customizing FREQ Procedure Output in SAS 9.2"

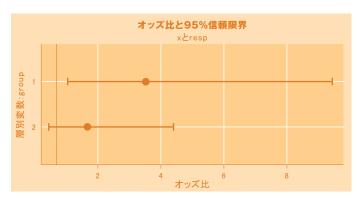
http://support.sas.com/resources/papers/freg92.pdf

# ● グラフ作成機能

ODS統計グラフの機能を利用して、さまざまなタイプのグラフを描画可能となりました。TABLESステートメントでPLOTS=オプションを使用すると、指定したテーブルやオプションに応じて、度数、累積度数、偏差、オッズ比、そしてカッパ係数に関するプロットを描くことができます。

以下のプログラムは、オッズ比に関するプロットをグループ毎に描画する例です。

```
DATA data2;
  INPUT group x$ resp freq;
DATALINES:
1 A 1 18
1 A 0 22
1 B 1 8
1 B 0 32
2 A 1 24
2 A 0 16
2 B 1 18
2 B 0 22
RUN;
ODS GRAPHICS ON;
PROC FREQ DATA=data2 ORDER=DATA;
  TABLES group*x*resp/RELRISK PLOTS=(ODDSRATIOPLOT);
  WEIGHT freq;
RUN:
```



FREQプロシジャによるオッズ比に関するプロットの例

# ● 同等性、非劣性、および優越性検定

BINOMIALオプションによる二項割合に関する検定や、RISKDIFFオプションによる二項割合の差に関する検定において、標記の内容がサポートされました。具体的には、以下のサブ・オプションを利用します。

> 同等性検定 EQUIVまたはEQUIVALENCE > 非劣性検定 NONINFまたはNONINFERIORITY > 優越性検定 SUPまたはSUPERIORITY

また、各検定におけるマージン(Δ)はサブ・オプションMARGIN=で、検定で用いる分散の計算方法をVAR=で指定できます。リスク差に関しては、検定の手法も数種類用意されています。次の指定例は、マージンを0.1としたリスク差に関する非劣性検定を行なうものです。

TABLES drug\*respnse / RISKDIFF(NONINF MARGIN=0.1);

# ● 二項割合に対する信頼区間

TABLESステートメントでBINOMIALオプションを利用すると、二項割合に対する信頼区間が算出されます。SAS 9.2では、その新たな計算法としてAgresti-Coull、Jeffreys、およびWilson(スコア法)にもとづく手法がサポートされました。それぞれAGRESTICOULLまたはAC、JEFFREYSまたはJ、およびWILSONまたはWというサブ・オプションを指定します。

STN

旧来から存在する漸近分布にもとづくWald流の計算はWALDの指定で、正確分布にもとづく方法もEXACTまたCLOPPERPEARSONの指定で個別に算出できます。なお、すべての種類の信頼区間を一度に出力するには、ALLというサブ・オプションを与えてください。

次のプログラムは、WilsonとAgresti-Coullの方法にもとづく信頼区間を計算する例です。関心の対象である変数xの水準は、LEVEL=サブ・オプションの指定によって"1"と定めています。

```
DATA data3;
    INPUT x freq 00;
DATALINES;
0 12 1 18;
;
RUN;

PROC FREQ DATA=data3;
    TABLES x / BINOMIAL(LEVEL="1" WILSON AGRESTICOULL);
    WEIGHT freq;
RUN;
```

```
      x = 1 の二項分布の比率

      比率
      0.6000

      漸近標準誤差
      0.0894

      タイプ
      95% 信頼限界

      Wilson
      0.4232
      0.7541

      Agresti-Coull
      0.4229
      0.7544
```

# ● 正確分布にもとづくリスク差の信頼区間

EXACTステートメントでRISKDIFF統計オプションを指定すると、正確分布にもとづくリスク差の信頼区間を計算できるようになりました。

```
DATA data4;
INPUT drug$ resp freq;
DATALINES;
A 1 6
A 0 9
B 1 4
B 0 11
;
RUN;
PROC FREQ DATA=data4 ORDER=DATA;
TABLES drug*resp / RISKDIFF;
WEIGHT freq;
EXACT RISKDIFF;
RUN;
```

		列 1リスクの推り		
	リスク	漸近標準誤差	(直 <del>)</del> 信頼	接確率) 95% 限界 
行 1	0.4000	0.1265	0.1634	0.6771
行 2	0.2667	0.1142	0.0779	0.5510
合計	0.3333	0.0861 (中野	各) 0.1729	0.5281
差	0.1333	0.1704	-0.2515	0.4908

新たに計算されるようになった統計量は、下線部の出力です。 この計算方法の詳細は、FREQプロシジャに対するドキュメントの以下 の項に記載されています。

Details: FREQ Procedure-->Statistical Computations-->
Risks and Risk Differences

#### ● Zelenの正確検定

ZELENまたはEQORを統計オプションとしてEXACTステートメントで指定すると、オッズ比の均一性に対するZelenの正確検定を行なうことができます。この検定の詳細については、FREQプロシジャに対するドキュメントの以下の項に記載されています。

Details: FREQ Procedure—>Statistical Computations—>
Cochran—Mantel—Haenszel Statistics

#### 1.3.3 MEANSプロシジャ

PROC MEANSステートメントで統計キーワードMODEを指定すると、最頻値(モード)が計算されるようになりました。度数が2以上である最頻値が複数存在するときは、その中で最も小さな値が出力されます。ただしUNIVARIATEプロシジャとは異なり、WEIGHTステートメントを利用して重み変数を指定した場合は、最頻値は計算されません。また、統計キーワードPRTをPROBTの代わりに利用できるようになりました。

#### 1.3.4 RANKプロシジャ

同順位データの扱い方を指定するTIES=オプションでDENSEまたは CONDENSEと指定すると、同順位の有無にかかわらず1、2、3…と順 位が割り当てられます。

以下のプログラムは、TIES=DENSEを用いた計算例です。

```
DATA data5;
INPUT value @@;
DATALINES;
1.5 1.5 2.3 6.4 5.9 5.9 4.5;
RUN;

PROC RANK DATA=data5 OUT=out5 TIES=DENSE; /*TIES=DENSE*/
VAR value;
RANKS denserank;
RUN;

PROC PRINT DATA=out5;
RUN;
```

OBS	value	denserank	
1	1.5	1	
2	1.5	1	
3	2.3	2	
4	6.4	5	
5	5.9	4	
6	5.9	4	
7	4.5	3	

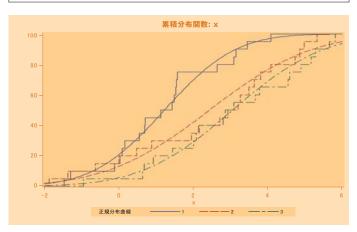
#### 1.3.5 UNIVARIATEプロシジャ

#### ● プロットの作成に関連する機能

CDFPLOTステートメントを利用して経験分布関数のプロットを、また PPPLOTステートメントを利用していわゆるP-Pプロットを描画可能に なりました。既存のHISTOGRAMステートメントなどと同様に、正規分 布などにおけるパラメータの推定や、ヒストグラムへ重ね合わせて密度 関数を描くこともできます。またHISTOGRAMステートメントでは、 Johnson分布システムSBとSUを扱うことが可能となりました。それぞれ SBとSUをオプションとして指定します。このステートメントでは、 Sheather-Jonesプラグイン法によるカーネル密度推定に対する「幅」 の算出もサポートされました。グラフを描画する5つのステートメントで は、ODS GRAPHICSステートメントの併用によって、ODS統計グラフに もとづいたグラフを描くことができます。なお、現在のリリースではこの 機能は評価版の扱いです。

次のプログラムは、変数groupの水準毎に求めたxの経験分布関数と 推定された正規曲線を、すべて重ねて表示するものです。

```
DATA data6;
  CALL STREAMINIT(12345);
  DO group=1 to 3;
    DO i=1 TO 20;
      x=RAND('NORMAL',group,2);
      OUTPUT;
    END:
  END;
  DROP i;
RUN;
ODS GRAPHICS ON;
PROC UNIVARIATE DATA=data6;
  CLASS group;
  CDFPLOT x /NORMAL OVERLAY;
RUN:
```



CDFPLOTステートメントによるグラフの例

# ● SASデータセットへの統計量の出力

PROC UNIVARIATEステートメントでOUTTABLE=オプションを使用する と、各分析変数に対するさまざまな要約統計量を一度にSASデータセッ ト化することができます。それらの統計量はOUTPUTステートメントを 利用してSASデータセットとして保存することもできますが、OUTTABLE= オプションは指定が特に変数が数多く存在するときに有効です。ただし、 出力できる統計量は、以下のものに固定されています。

・非欠損オブザベーション数 ・10パーセント点(\_P10\_) (\_NOBS\_) 第1四分位点、25パーセント点 ・欠損オブザベーション数 (Q1) ・中央値、第2四分位点、50パーセ (NMISS) ント点(\_MEDIAN ) ·重みの合計(\_SUMWGT\_) ·合計(\_SUM\_) 第3四分位点、75パーセント点 ·平均(\_MEAN\_) (\_Q3\_) ·標準偏差(\_STD\_) ・90パーセント点(\_P90\_) ・95パーセント点(\_P95\_) ·分散(\_VARI\_) ·歪度(\_SKEW\_) ・99パーセント点(\_P99\_) ·尖度(\_KURT\_) •範囲(\_RANGE\_) ·最小值(\_MIN\_) ·四分位範囲(\_QRANGE\_) ·最大値(\_MAX\_) ·最頻値(\_MODE\_) ・1パーセント点(\_P1\_) ・符号検定における検定統計量 ·5パーセント点(\_P5\_) (\_SGNRNK\_)

次のプログラムは、OUTTABLE=オプションでSASデータセットを作成 して、その内容を確認する例です。

```
PROC UNIVARIATE DATA=SASHELP.CLASS OUTTABLE=out7;
  VAR NUMERIC;
RUN;
PROC PRINT DATA=out7 LABEL;
RUN:
```

Variable Name Variance	Variable Label			Standard Mean	Deviation
Age Height Weight	年齢 身長(インチ) 体重(ポンド)		13.316 62.337 100.026	1.4927 5.1271 22.7739	2.228 26.287 518.652
_					1st
Skew	mess Ku	rtosis	Mi	nimum	Percentile
,		.11093		11.0	11.0 (後略)
-0.2 0.1		.13897 .68336		51.3 50.5	51.3 50.5



Base SASのドキュメントは、SAS 9.2を導入していない場合でも、下記の Webサイトからご参照いただけます。

#### Base SASに関連するドキュメント一覧

http://support.sas.com/documentation/onlinedoc/base/index.html

前号の特集記事とともに、この特集がSAS 9.2のさらなる活用に結びつ くと幸いです。

SAS アカデミック・ニュース

# **Academic News**

今号では、カスタマー・リレーションシップとして、本年春に山形大学とSAS Institute Japan 株式会社が共同で学生および教員向けの講習会を開催しましたので、

その模様を秋冬号でご紹介いたします。また、事例紹介は引き続き福岡大学にて SAS Enterprise Guide®を活用した教育および学生からの反応、

コラムSAS四方山話では「アンケート・データの解析②」アンケート・データからの 相関関係と因果関係などをSASを使用してどう分析していくかを紹介しています。



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)



報として活用すべき かが重要視されて いる現状では、学 生達はデータ分析 を避けて通ることは できません。ただ、 少なくとも本学商学 部の学生達が、統 計学や高度なデー タ分析を自身の専 門とする必要はな いとも考えます。重 視すべき点は、基 礎的なデータ分析 を日常的に使用す る手段の一つとして 身につけること、そし

てデータ分析の結果をビジネスに生かす ための"データを読み解く目"を養うことだ と考えています。

前回は、本学商学部商学科における教 育方針やその背景等についてご説明いた しました。今回は、私のゼミナールの状況 についてお話したいと思います。

現在、福岡大学商学部では2年次後期か ら4年次終了まで2年半のゼミナールを開 講しています。学生達は2年半の間、ゼミに 所属してそれぞれの分野について学習・研 究します。私のゼミナールでは、主にマーケ ティング戦略とマーケティング・リサーチを 扱っており、現在、4年生が17名、3年生が17 名(加えて聴講生1名)、2年生が18名です。

前回、前々回でもお話いたしましたよう に、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。商学に関するデータ 分析の経験をほぼゼロで卒業してしまう学 生達、数学や数式を使うことに極めて消極 的な学生達が決して少なくないことも事実 です。

しかし、ビジネスの現場でデータをどう扱 い、それをどのように意思決定に有益な情

# SASESAS Enterprise Guide の試験的導入

昨年10月頃、現3年ゼミ生の2名がデータ 分析とその技術を自分の武器として身に つけたいと希望してきました。早い時期か ら学んでもらえばSASを習得可能と考え、 彼らにSASプログラムに挑んでもらうことに 1.ました。



予想していなかったのは、彼らがSASプロ グラムを私の想像以上に早く実行できたこ と、そして競うように次のプログラムを教え て欲しいと言ってきたことでした。彼らは、 過去に指導してきたゼミ生達より、パソコン 操作にかなり習熟していたと思います。 これは"パソコン操作"という点に限って、高 校や大学初年次の情報教育に成果があっ たことを示唆しているのかもしれません。

彼らが学んだのはBase SASとSAS/ STAT®に含まれるごく一部のプロシジャで 基礎的なものばかりでした。しかし、プログ ラミングや統計学・数学に関して特に訓練 を受けていない彼らが、ゼミ指導教員の私 が決して簡単でないと説明したプログラム ベースのSASを約1ヶ月弱である程度使用 できたことは事実でした。

加えて、因子分析の基礎まで学んだ段階 で、習得した分析をSAS Enterprise Guide (以下、EGと略称)でも試してもらいました。 プログラムを記述することなくマウス操作で 分析を完結できるEGの操作性は彼らの分 析のスピードを、デフォルトでHTML形式と なるEGのアウトプットは(Power Point等 で)彼らが分析結果を整理するスピードを 格段に向上させました。

現在、彼らはEGを主に使用するソフトウェ アとして選択しています。これは彼らが直面 してきた分析がEGで十分なものがほとん どだったことが主な理由です。SASプログラ ムが記述できれば、ソフトウェアが何をして いるのかある程度わかると思います。一定 程度SASを理解した上で、必要に応じて EGを選択するという彼らのやり方は理に 適っているのではないかと考えます。

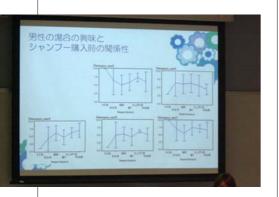
そして、彼らへの指導の成果は私の予想 を大きく超えるものでした。毎年、ゼミナー ルでデータ分析をどの程度やってもらうか

は、私にとって悩ましい問題でしたが、2名 への指導を通じて、ある程度丁寧に指導 すれば、現在、本学商学部でSASやEGを 使用することは可能であると感じました。

# 入口としてのEG

2名への指導を経て、昨年11月頃、現3年 ゼミ生全員にEGを利用してもらいました。 EGを採用した主な理由はその操作性と インターフェースにあります。現在までのと ころ、ゼミナールではEGの使用に関して大 きな問題は生じていません。他のゼミ生達 に先行した2名は意識的にも無意識的にも 他のゼミ生達を引っ張り、他のゼミ生達も ほとんど抵抗なくEGを受け入れました。

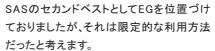
ゼミ生達へは、最初からあまり多くを要求 せず、例えば2年次には全員が(線形)回帰 分析の操作だけ習得できるように手法を 限定しました。もし、得られた回帰係数では 解釈が困難な場合、説明変数と被説明変 数の散布図やグラフを作成して変数間の 関係を視覚的に捉えること、つまりデータ 分析の基礎から行なうことを勧めました。



EGでは散布図やグラフの作成後、色指 定、データラベル表示、散布図のプロット記 号や回帰直線のあてはめ、三次元散布図 の回転等、マウス操作だけでかなり自由に カスタマイズできます。単にSASプログラム を通すためにEGを使うのでなく、プレゼン テーション資料等にEGのアウトプットを可 能な限り利用できるよう、これら機能をゼミ では重視しています。

他方、自分が期待する分析結果を導くに は、私が教えた分析手法だけでは十分で ないと気づき、(荒っぽいやり方ですが)ほ ぼ自力で分散分析のグループ分析等を行 なったゼミ生もいます。加えて、3名はEG の機能を使いデータを整える等の加工 を行ない、データマイニングソフトの SAS Enterprise Miner™(以下、EMと略称)で

試すところに至って います。現在はEM の決定木分析機能 で、探索的なデータ 分析、新たな知見 の発見にEMを試し てもらっています。 EGはデータ分析に も有益ですが、上手 にデータハンドリン グできればEMへ接 続できるという側面 を有しています。最 初、私はあくまでも プログラムベースの



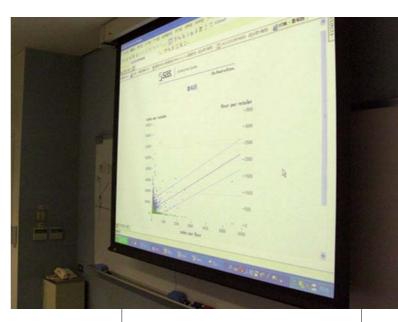
ゼミ生達は私の思惑を良い意味で裏切 り、直面するデータ、分析手法、そしてアウト プットの形等に応じ、必要なソフトウェアを 使い分けようとしています。EGはその入口 の役割を果たしたと思います。

# "データを読み解く目"の養成

データ分析に関する到達目標として私が 設定した「データをある程度自由に分析で きること」は、EGの活用である程度クリアし たと考えております。しかし、「マーケティン グやビジネスへのインプリケーションを一 定程度導出できること」「その結論を他人 が理解できるように自分の言葉で伝えられ ること」に関しては苦戦し続けています。 目の前のアウトプットを理解することに精一 杯なゼミ生が少なくありません。

適切な分析手法を選択できるか否か、 分析結果を丁寧に解釈できるか否かと いう点で、現在、ゼミ生の差が明らかになり はじめています。マーケティングやビジネ ス、さらに統計学等を学んでいるか(学ぼう としているか)否かが、この差を生んでいる ようです。アウトプットだけを一見すると分析 はできているように見えます。しかし、適切 な分析手法を選択しているか否か、分析結 果を解釈できているか否かは別な問題で す。特に、EMを使おうとするような学生達 は、私が学生の時とは比較にならないくら い、統計学やデータ分析について学ぶ必 要があるはずです。

昨年、まだ2年生だったあるゼミ生が「こ んな結果が出るのは(常識で考えて)おか



しい。これはデータが良くないです!」と強く 私へ主張しました。説明変数の変数選択 が不適切な重回帰分析結果でしたが、学 内調査の便宜サンプリングによるデータで したので、大きく外れた解釈ではなかった かもしれません。このゼミ生には、変数選 択とサンプリングに関して説明し、データの 何が"良くない"のか考えてほしいと指導し ました。

現在、このゼミ生は丁寧に分析結果を読 むことができる一人になっています。しかし、 ある意味でナイーブな(本学商学部2年生 による)この解釈は、"データを読み解く目" を養うことが難しいと私に教えてくれま した。データ分析に入門した彼・彼女らを 今後どのように導いていくか、分析結果を 解釈するために必要な知識をどのように つけてもらえば良いかは、現在の私が抱え ている大きな課題となっております。

"敷居が高い""玄人向け"といったイメー ジが強いSASの敷居をEGは大きく下げ、 本学商学部でもデータ分析の入り口として 有益であるとわかってきました。EGは本年 度商学部で私が担当するマーケティング・ リサーチの講義でも使用予定です。次号で は、同講義での活用等についてもお伝えし たいと思います。



大橋 渉

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

# コラム「SAS四方山話」

# 第7回 アンケート四方山話(5)「アンケートデータの解析②」

# (1) p値依存症?

では今回は「回帰と相関」のお話…の前 に、前回お話させていただきました、解析 結果の一般化(=いわゆる検定)について、 少しお話させて下さい。検定の考え方につ きましては前回を参照して頂きたい次第で すが、この連載に限らず、講義などでも多く の皆様より「検定の考え方を理解するのが 難しい」という声をよく耳にします。

あくまで筆者の推測ですが、その理解が 難しいがゆえに、検定結果の解釈は「p値 が0.05を下回ればよい」と機械的に処理さ れるようになった部分も大きいのではない でしょうか? 「p値の大小だけに(特に0.05 を下回るか否かに)しか着目しないのはけ しからん!」とはよく言ったもので、これは統 計学の先生方のみならず、通称バンクー バー・グループによる「生物・医学論文の統 一投稿規程」でも主張されている有名なお 言葉です。しかしながら、そのような人々の 思いとは裏腹に、p値は一部の人々の間で 「0.05を下回れば差がある」「小さければ小 さいほどよい」「小さくした者勝ち?」などと、 伝言ゲームよろしくその存在感は大きくなっ てしまいました。さらには神格化され、統計 学における全知全能の神となり、これさえ 表示してあれば全てOKの免罪符になって しまったのです。

ここからは昔筆者が関わったある自治体 における調査のお話です。筆者は結果のプ レゼンテーションの場で、主任研究者の先 生から「どうしてp値を出さないんだ!」と怒鳴 られてしまったのでした。筆者の怠慢かそれ とも…皆様もご一緒にお考え下さいませ。

調 査 名: A市高齢者調査

調査対象: A市に居住する70歳以上のお年寄り全て

(約18 000名)

調 查 期間 平成20年3月1日~平成20年5月31日

調査方法:郵送法。返答なき場合に督促、訪問面接

法を併用する

調査の目的: A市のお年寄りの生活、行動様式などを

調査することにより、A市における高齢者 福祉予算の効果的配分のための基礎資

調査項目:年齢・性別・居住地域、その他具体的項目

前記の調査概要にもありますが、調査目 的はこの自治体における予算の効果的配 分のための基礎資料を得ることです。結果 はこの自治体のみに効果的に用いること ができれば良いのであって、たとえば全国 や他の自治体に結果を反映することを考 える必要はありません。さらに、居住する全 ての70歳以上のお年寄りに尋ねている訳 ですから、検定により推定や推測をする必 要はありません。つまり、今回は知りたい対 象の全てを調べているのですから、単純に 数えた結果こそが、知りたい結果そのもの なのです。

さて、当の主任研究者は最後まで納得が 行かない様子でした。研究の結論にp値を 表示しないのは不安であるということを言 い続けており、風の噂によれば(?)自分た ちで独自に検定を行なったとも聞きます。 「何かの集計をしました→とにかくp値→ 0.05より上か下か?」…集計したらとにかく p値! 求めないと不安! 実はこんな事 例は決して少なくはありません。p値が強く なりすぎてしまったゆえの出来事だと思え ませんか? もしも皆様の大学における統 計学の指導が、「ソフトウェアのココを押して 出力のソコを見る」※に終始しているとした ら…要注意かも知れません。

※SASの場合はp値の出力が多過ぎて、どのp値を 参照すればよいかわからないというお話もありま すが…ここでも統計の知識が必要であるという ことです

# (2) 回帰と相関

やっと本題に入ります。さて、BMIとコレス テロール、景気と企業における雇用者数、 気温とビールの売上…等々、世の中には相 互に関係がありそうなものが多いです。詳 細は例によって専門書にお任せしますが、 これら(x,v)の関係の強さを数字によって示 したものが相関係数であり、通常はrによっ て示されます。rの値は必ず-1≦r≦1となり、 1もしくは-1に近ければ近いほど相互に関 係が強く、0に近いほど相互に関係が弱い ということです。仮に(x,v)の両方が計量 データであれば、Pearsonの積率相関係 数を用いることになり、通常「相関係数」と いえばこちらのことになります。また、もしも 一方でもカテゴリデータ(1.よい 2.ふつう 3.わるいなど)が含まれるのであれば、 Spearmanの順位相関係数を用いる必要 があります。

相関関数

$$r = \frac{\sum (x_i - \bar{x}) (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 (y_i - \bar{y})^2}}$$

 $\bar{X}$ :変数Xの平均値  $X_i$ :変数Xの各値  $y_i$ :変数yの各値 y:変数yの平均値

一方で回帰とは、一方の変数の値を基準 として、もう一方の値を決定付けることです。 例えば、y=2x+3の場合は、xが決定すれば yも自然と決定するので、xの値がyの値を 決定付けているということになります。図1 のような、中学校時代に学習したいわゆる 「一次関数」はxの値からyの値を100%予測 することが可能でした。ところが実際の データは決して直線で予測できるほど単純 ではなく、直線的にxの値からyの値を予測 できることはまずありません。

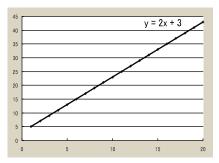


図1 単純な回帰直線(y=2x+3)

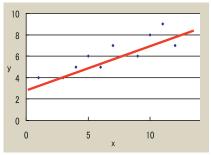


図2 実際の回帰直線

2変数の関係という意味では相関と回帰 は共通には違いありません。ところが、相関 は2つの変数の関連の強さを見るものであ り、回帰は、一方の変数の値からもう一方 の値を導くことで、明らかに異なるもので す。相関はrの絶対値が小さければ「関連 が小さい」「関連が見られない」と結論付け るものであり、回帰は、予測のための式を 立ててその当てはまりを見るものです。図1 のように(x,y)が一直線になってくれれば予 測の精度は100%となり、これ以上ないほ どの予測式になるのですが、実際には図2 程度でもかなり上出来な方です。なかなか 直線に乗らない(x,y)を強引に直線に乗っ けたとしても、予測の制度が悪ければ使い 物にならないということになります。

# (3) 相関係数と因果関係

基本的に2つの数量化された事柄があ れば相関係数を求めることは可能です。片 方、または両方とも順序カテゴリデータで あろうとも、はい/いいえ、あり/なしの2値 データであろうとも、とにかく何らかの相関 係数を求めること自体は可能です。よくマス コミが得意とする、巨人の順位と経済効 果、アニメファンに肥満者が多い等々、明ら かなこじつけに見えるようなものでも、とに かく求めるだけであれば可能なのです。

ですが、そもそも相関関係を見たい理由 は、そこに何らかの因果関係があるかどう かを見たいということであり、相関係数を求 めること自体が目的ではありません。因果 関係が成立するためには相関係数以外の 判断要素(時間性、 合理性等)が多数 存在するため、相関 係数が高いからと いってそこに何らか の因果関係が成立 してわけではない のです。有名な「血 圧と給料」も相関係 数自体は高いので

しょうが、そこには一番必要と思われる「合 理性」が存在しないのは言うまでもありま せん。血圧と給料の相関係数は高いかも 知れませんが、血圧が上がったら必ずしも 給料が上がるわけではありませんし、そこ には「年齢」という要素が含まれていること は容易に理解できます。

マーケティング担当の方などは、人が 気付かない事柄の間に何らかの関連を見 つけ、それを商品開発のHintにすることを 業務とされている方も多いと思われます。 アンケートの結果などを数量化して相関係 数を求めてみると…意外な組み合わせの 相関係数が高いということを発見すること もあると思われます。その中には勿論、 世紀の大発見(?)に匹敵するようなものも 無いとは申しませんが、今一度「因果関係 の有無」には気をつけましょう!

#### (4)SASによる相関係数の求め方

では、SASにより相関係数を求めてみま しょう。出色の詳細などはSASの専門書に お任せするとして、まずは散布図を書く癖

散布図

/\*散布図の作成\*/ proc gplot data=data1; plot y\*x; /\*y軸、x軸にそれぞれ示したい変数\*/

run; 散布図

proc corr data=data1 outp=wk1 outs=wk2 pearson spearman; /\*求めたい相関係数の種類は?\*/

var x y; run;

> を付けましょう! 慣れてくると、おおよその 散布図の形で相関の有無は何となく(?)見 えるようになります。なお、美しくグラフを書 く方法は、筆者の苦手とするところですので 割愛させて頂きます(笑)。

相関係数を求めるプロシジャのcorrとは correlation(相関)の略であり、ここでは Pearson およびSpearmanの両方について 相関係数を求めています。ここで何も設定 しなければPearsonの相関係数のみが算 出され、Spearmanとだけ記載した場合は Spearmanの順位相関係数のみが算出さ れます。outpはPearsonの相関係数とp値、 outsはSpearmanの順位相関係数とp値 を、それぞれSASデータセットwk1および wk2に出力するためのオプションを示して います。僅か数行ではありますが、2組の数 値の相関係数はこれだけで求めることが できます。

次回、回帰の続きをもちまして「アンケート 四方山話」はいったん着地させて頂きます。

# 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



# I. イントロダクション

#### (1) 山形大学の説明

山形大学は、6学部を持つ県内唯一の総 合大学です。

県都山形市には人文学部、地域教育文 化学部、理学部のある小白川キャンパスと 医学部のある飯田キャンパスがあります。

工学部は愛と義に生きた名将「直江兼 続」で有名な上杉藩ゆかりの城下町米沢 に、農学部は米どころ庄内平野の鶴岡にそ れぞれキャンパスがあります。

このように山形大学は4つのキャンパスに 分散しています。4キャンパスとも自然に恵 まれ、研究や教育にとても適した環境です。

情報の中枢となる学術情報基盤センター (平成21年4月現在)は小白川キャンパスに あり、全ての学生と教職員は距離的制約を 受けず等しく利用できるように、センターに はセントラルシステム(メールサーバ、ファイ ルサブシステム)があり、小白川キャンパスと 他のキャンパスは、高速ネットワーク回線で

結ばれています。

統計処理ソフトウェアを母体とするSAS システムは、利用者の多い学部がある2つ のキャンパスに専用システムを配置してい

1つ目のシステムは、医学部のある飯田 キャンパスにSAS専用のUNIXサーバが設 置されており、UNIX版SASが導入されてい ます。2つ目のシステムは農学部のある鶴岡 キャンパスにあり、Windows版SASが導入さ れています。

我々が担当する飯田キャンパスには、医学部、関連する附属施設、附属病院、遺伝子実験施設があり、構成員は、学生が870名、大学院生が200名、教職員が920名、合計約2,000名です。

飯田キャンパス内のパソコンルームは、 120台を設置した視聴覚教室と20台ほど を設置した計算機室の2つがあります。

計算機室は随時利用できるように部屋の前に入室管理システムの磁気カードリーダがあって、各個人が所有する磁気カードで入室します。室内には、オンデマンド型カラープリンタ・フラットベッド型スキャナー(ネットワーク対応)・B0版カラープリンタ・モノクロプリンタ・パソコン20台が設置され、24時間365日利用できるようになっています。

視聴覚教室と計算機室は基礎研究棟と臨床研究棟の間の建物の2階にあり、キャンパス内の他の各建物間も2階部分が渡り廊下で結ばれているので、パソコンルームは利用者にとって利便性が良くなっています。講義が終了するやいなや学生がパソコンルームになだれ込んでレポートの作成等に精を出す姿をよく目にします。ここ数年のパソコンルームの利用者の数は、多いときには1日で150名を超えることがあります。

我々は、このような利用状況の中で時おり発生するトラブルに対処したり、個人情報に対する対応を行なっていますが、対応件数や対応要件が以前より増えてきたようです。

その他には、インターネットに関する問い合わせやネットワーク障害の対応、各種サービス、例えばポスタープレゼンテーション用の大判印刷を行う利用者の方に、印刷する上での注意点やテクニックをアドバイスしています。

また、研究者や大学院生に対してSASの利用に関するコンサルテーションを実施しています。

# (2) これまでの統計処理ソフト ウエアSASの利用経緯について

当センターは、当初山形大学情報処理 センターとして昭和62年(1987年)に発足 し、1989年4月より最初に導入したSASシ ステムはPC版SAS6.03で、ライセンス数は 10本であり、プロダクト(解析機能)は、 Base SAS、SAS/STAT、SAS/IML®、 SAS/AF®、SAS/FSP®、SAS/GRAPH® で開始しました。

PC版SASとは、MS-DOSから起動する SASシステムです。パソコンルームの数台 のパソコンにインストールしたものを利用し て貰っていました。

翌年の1990年4月にはETSが追加されて時系列分析ができるようになりました。1991年11月からはバージョン6.04が運用されました。

余談になりますが、このころSAS JAPAN が制作したSAS教育用ビデオ「生存時間解析の基礎から生存関数の比較まで」(東京大学医学部附属病院中央医療情報部の大橋靖雄先生が解説)を購入して、生存時間解析をしたい希望者にビデオテープの貸し出しもしていました。

1991年11月にUNIX版SAS6.07Jを導入して、ネットワークを経由してリモートで利用できるようになったことで利用者数は1998年に60名となりピークとなりました。

その後SASのバージョンが、6.12から8.2 へ、8.2から9.1.3へとバージョンアップして いくうちに、SASの利用者が減少しました。

第1にバージョン8.2になった時にVTエミュレーション端末でSASのプログラムを作成したり実行処理ができなくなったこと、第2にSAS言語を知らなくてもメニューから項目を選択でするだけで統計解析ができたSAS/ASSIST®が削られてしてしまったことが最大の理由のように思われます。

この頃、利用者数の減少は全学の情報処理コンピュータシステムの更改時にSASを導入してもらうことができなくなることにつながるという危機感がありました。そこで、2005年度の中期計画には、SASを存続させるために「今後予定されているシステム更新後も統計処理パッケージが利用できるように努力する。」という命題を掲げてSASソフトウェアの継続的な導入を要望しつづけました。こうして辛うじて現システムにおいてもSASソフトウェアが導入されました。

# Ⅱ. 開催までの経緯

# (1) そもそもセミナーを開催 しようと考えた理由

このようなSAS利用の危機的状況において、講習会を開催したら参加してもらえるか聞いてみたところかなりの希望者が

いたことと、平成20年度にGlobal COE (Centers of Excellence)「以下、GCOEと略称」における拠点のプログラムが採択されたことがきっかけでセミナーの開催を計画することになりました。

採択された山形大学のGCOEにおける拠点のプログラムは、「分子疫学の国際教育研究ネットワークの構築」というテーマなので、多くの若手研究者がSASを必要としていると推測され、大学院委員長などの意見も同様であったからです。

講習会と関連することを拠点形成の目的から抜粋すると、「分子疫学は、従来の臨床疫学研究に『ヒトゲノム計画』の成果を取り入れた新しい学問分野であり、病気の原因・病態・多様性に関与する分子を遺伝子レベル・蛋白質レベルで明らかにし、これらの成果を活用して『テーラーメイドを療』や『ゲノム創薬』の確立に貢献が必った。出発点には深い臨床的洞察が計学、遺伝子解析技術、バイオインフォマティクス、生化学、分子生物学、細胞生物学、生理学、病理学など、多種多様な学問分野である。」とされています。

個々の研究者においては統計処理・解析を行なうための何かしらの統計ソフトウェアが使われているかもしれませんが、統計処理ソフトウェアは学内の研究者間で統一化する必要性を感じました。SASは世界的に見て統計解析ソフトウェアの標準・スタンダードであって、他にスタンダードと称されるSPSSと比べても遜色はありませんし、各方面の研究分野においても利用できるプロダクト(解析機能)を有していてサポートも充実していることから是非推薦したいと考えました。

次の冬号では、セミナー開催にあたって の障壁や開催後の受講者の反響等に ついて、紹介いたします。

# STN

# Q&A

- ERRORメッセージ、WARNINGメッセージ取得
- DI Studioで多数のプロセスを配置したジョブが 開けない場合の対応
- ライブラリ単位でデータセットの内容を比較する方法
- WARNINGの制御

- CONTENTSプロシジャの結果を変数名の アルファベット順、数字順にソートしたい
- PNG形式でのグラフ出力
- オッズ比の算出





プログラム実行時、ログに出力されるERRORメッセージ、WARNINGメッセージを取得したいのですが、取得するためのシステム・オプション、もしくは自動マクロ変数など、何

か容易な方法はありますか?

A

SAS 9.2から新しくERRORメッセージ、WARNINGメッセージを格納する自動マクロ変数SYSERRORTEXT、SYSWARNINGTEXTが追加されました。これらの自動

マクロ変数を参照する方法があります。

取得したメッセージを確認するためのサンプル・プログラムを記述いたします。

#### サンプル・プログラム

DAT sample; SET sashelp.clas; RUN;

%PUT &SYSWARNINGTEXT; %PUT &SYSERRORTEXT;

# ログの出力

1 DAT sample;

14

**WARNING 14-169: DAT** はシンボル **DATA** のスペルミスであると 判断しました。

2 SET sashelp.clas;

ERROR: ファイル SASHELP.CLAS.DATA は存在しません。

3 RUN;

NOTE: エラーが発生したため、このステップの処理を中止しました。 WARNING: データセット WORK.SAMPLE は未完成です。このステップ は、0 オブザベーション、0 変数で停止しました。

NOTE: DATA ステートメント 処理 (合計処理時間):

処理時間

0.06 秒

CPU 時間

0.04 秒

5 %PUT "&SYSWARNINGTEXT";

" データセット WORK.SAMPLE は未完成です。このステップは、 0 オブザベーション、0 変数で停止しました・

%PUT "&SYSERRORTEXT";

" ファイル SASHELP.CLAS.DATA は存在しません・



4

SAS Data Integration Studioのプロセス・デザイナにて ジョブを作成しましたが、処理内容を増やしたところ、ジョ ブが開けなくなってしまいました。どのようにすればジョブ

を開くことができるでしょうか?

A

SAS Data Integration Studio(以下、DI Studioと略称) にて多数のプロセスを配置し、複雑なフローを作成した 場合、クライアントPCにて使用しているJavaの設定で

Overflowが発生し、ジョブが開けなくなる場合があります。

このような場合には、Javaの[-Xss]オプションの数値を変更して、 スタックサイズを調整してみてください。

DI Studioでの設定は次のファイルにて可能です。

!DI Studioのインストールフォルダ\etlstudio.ini

※標準では C:\Program Files\SAS\SASETLStudio\9.1

#### 指定例

変更前:(一部抜粋)

CommandLineArgs=-Xmx1024m -Djava.system.class.loader =com.sas.app...

# 変更後:(一部抜粋)

CommandLineArgs=-Xmx1024m -Xss512k -Djava.system.class.loader=com...

...



あるライブラリ内に存在する複数のデータセットを、別ライブラリにコピーしました。コピーが正しく行なわれていることを確認するために、データセットの内容を比較したいの

ですが、何かよい方法はありますか?



COMPAREプロシジャにて、データセットの内容を比較することができます。しかしながら、COMPAREプロシジャではライブラリ単位にて、データセットの内容を比

較することはできません。そこで、データセットの一覧を SASHELP.VMEMBERから取得し、CALL EXECUTEルーチンにて COMPAREプロシジャを実行する方法が考えられます。

以下のサンプル・プログラムは、MYLIBライブラリとWORKライブラリ中のデータセットを比較しています。

# サンプル・プログラム

/\* サンプルデータの作成 \*/

DATA mylib.class work.class; SET sashelp.class; RUN:

DATA mylib.prdsale work.prdsale;
SET sashelp.prdsale;

```
RUN;

DATA _NULL_;

/* データセット一覧取得 */

SET sashelp.vmember(KEEP=memname libname memtype
WHERE=(LIBNAME=UPCASE("mylib")
AND MEMTYPE="DATA"));
/* COMPAREプロシジャの実行 */

CALL EXECUTE("PROC COMPARE BASE=mylib." || memname);
CALL EXECUTE(" COMPARE=work." || memname);
CALL EXECUTE(";");
CALL EXECUTE("RUN;");
RUN;
```



DROP、KEEP、RENAMEステートメント使用時、指定した 変数が存在しない場合にWARNINGが発生しますが、このWARNINGを表示しない方法はありますか?



DKRICOND、DKROCOND オプションを指定すると、該当の WARNING ステートメントの制御を行なうことができます。下記の値が指定可能です。

DKRICOND DKROCOND=ERROR | WARN | WARNING | NOWARN | NOWARNING

#### **ERROR**

DROP、KEEP、RENAME ステートメントに指定された変数が存在しない場合、ERROR メッセージを出力します。

WARN | WARNING WARNING を出力します。

NOWARN | NOWARNING WARNING を表示しません。

#### 例

```
OPTIONS DKRICOND = NOWARN;
                               /* テストデータ作成 */
DATA in:
 A=123;
RUN;
                           /* 存在しない変数bの指定 */
DATA out;
 SET in(DROP=b):
RUN:
OPTIONS DKROCOND = NOWARN;
                           /* 存在しない変数bの指定 */
DATA test:
A=123;
DROP b:
RUN;
結果)
1
    OPTIONS DKRICOND=NOWARN:
                               /* テストデータ作成 */
2
3
    DATA in:
      A=123;
```

```
RUN:
NOTE: データセット WORK.IN は 1 オブザベーション、1 変数です。
NOTE: DATA ステートメント 処理 (合計処理時間):
      処理時間
                      0.01 秒
      CPU 時間
                      0.01 秒
                        /* 存在しない変数bの指定*/
6
    DATA out;
8
      SET in(DROP=b);
NOTE: データセット WORK.IN から 1 オブザベーションを読み込みまし
NOTE: データセット WORK.OUT は 1 オブザベーション、 1 変数で
す。
NOTE: DATA ステートメント 処理 (合計処理時間):
      処理時間
                      0.00 秒
                      0.00 秒
      CPU 時間
10
11
    OPTIONS DKROCOND=NOWARN:
12
                        /* 存在しない変数bの指定*/
13
    DATA test:
14
      A=123;
      DROP b;
15
16
    RUN:
NOTE: データセット WORK.TEST は 1 オブザベーション、1 変数です。
NOTE: DATA ステートメント 処理 (合計処理時間):
      処理時間
                      0.00 秒
      CPU 時間
                      0.00 秒
```



CONTENTSプロシジャを使用すると変数名の結果リストが変数名の大文字、小文字のアルファベット順でソートされます。そのため、同じような名前の変数でも大文字、小

文字の違いで並べて表示されず、目的の変数を見つけることが困難です。変数名を大文字、小文字に関係なくアルファベット順にソートし、しかも変数名に数字が含まれる場合も数字順にソートしたいのですが、これを容易に実現する方法はないでしょうか?



CONTENTSプロシジャのORDER=IGNORECASEオプションを使用することで実現可能です。ORDER=IGNORECASEオプションは変数名を大文字、小文字に関わらず、アル

ファベット順にソートします。変数名に数字が含まれる場合、アルファベット順にソートした後に数字順にソートします。 以下はサンプル・プログラムと出力結果になります。

# サンプル・プログラム

# 出力結果

変数と	上属性の昇順!	ノスト			
#	変数	タイプ	長さ		
4	amount1	数値	8		
5	Amount2	数値	8		
6	aMount3	数值	8		
1	Id1	数值	8		
2	id2	数值	8		
3	ID3	数值	8		
10	returns1	数值	8		
11	Returns2	数值	8		
12	returns3	>> III	8		
7	Sales1	数值	8		
8	sales2	数值	8		
9	Sales3	数值	8		
13	x1	数值	8		
14	x2	数值	8		
15	X10	数值	8		

上記のORDER=IGNORECASE以外にも、SAS 9.2よりCONTENTS プロシジャのORDER=オプションに新しいオプションが追加されました。

ORDER=COLLATE:変数名を大文字、小文字のアルファベット順にソートします。

ORDER=CASECOLLATE:変数名を大文字、小文字にかかわらず、アルファベット順にソートします。変数名に数字が含まれる場合、アルファベット順にソートし、数字の順番は無視されます。

#### ORDER=COLLATEでの出力例

変数。	と属性の昇順リスト	
#	変数タイ	プ 長さ
5	Amount2 数值	<b>8</b>
3	ID3 数值	<b>8</b>
1	Id1 数值	8
11	Returns2 数值	8
7	Sales1 数值	8
9	Sales3 数值	8
15	X10 数值	8
6	aMount3 数值	8
4	amount1 数值	8
2	id2 数值	8
10	returns1 数值	<u> </u>
12	returns3 数值	8
8	sales2 数值	
13	x1 数值	8
14	x2 数值	<u> </u>

## ORDER=CASECOLLATEでの出力例

	変数と属性の昇順リスト						
	#	変数	タイプ	長さ			
	1	amount1	数値	8			
	5	Amount2	数值	8			
	6	aMount3	数值	8			
	1	Id1	数值	8			
	2	id2	数值	8			
- 1							

3	ID3	数値	8		
10	returns1	数値	8		
11	Returns2	数値	8		
12	returns3	数値	8		
7	Sales1	数値	8		
8	sales2	数値	8		
9	Sales3	数値	8		
13	x1	数値	8		
15	X10	数値	8		
14	x2	数值	8		

PNG形式にてグラフイメージをファイルに出力する際、 XPIXELS、YPIXELSオプションを用いて、解像度を高く設 定しています。以前のリリースでは有効でしたが、SAS 9.2 では有効となっていないようです。SAS 9.2にて、解像度の設定を変更

し、PNG形式にてグラフを出力するにはどのように指定すればいいの

A

でしょうか?

SAS 9.2ではより多くのフォント、カラーをサポートするため、PNGデバイスの設定が変更されています。この変更に伴い、GOPTIONSステートメントのXPIXELS、

YPIXELSオプションは有効とはなりません。

以前のリリースにおけるPNGデバイスと同じ設定は、SAS 9.2では ZPNGとなりますので、DEVICE=ZPNGの指定に変更していただくことにより、XPIXELS、YPIXELSオプションが有効となります。 この点に関しては、以下のUsage Noteに記載があります。

# Usage Note

http://support.sas.com/kb/33/678.html



LOGISTICプロシジャではオッズ比がデフォルトに算出されますが、交互作用項に含まれる主効果に対しては、値が表示されません。オプションなどの指定にて、表示させる

ことができますか?



交互作用項、および交互作用項に含まれている効果に 対するオッズ比を算出、表示するためのオプションはありません。しかしながら、ロジスティック回帰モデルは、

SAS 9.1.3より追加されているGLIMMIXプロシジャにて推定することができます。このプロシジャでは、ODDSRATIOオプションが含まれており、交互作用項に含まれている主効果のオッズ比を算出することができます。ODDSRATIOオプションは、PROC GLIMMIXステートメント、MODELステートメント、LSMEANSステートメントにて指定することができます。

#### 例

```
PROC GLIMMIX DATA=one ODDSRATIO;

CLASS treat gender;

MODEL resp = treat|gender / DIST=BINOMIAL

LINK=LOGIT;

RUN;
```

# **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による臨床試験データの統計解析【後編】」 コース (2日間)

[日 程]

2009年11月19日(木)~20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

SASシステムによる臨床試験統計解析に従事している方(上級者) 「前提知識〕

下記の3コースを受講済みか、同程度の知識のある方

- 「医薬向け分散分析」
- ・「医薬向けカテゴリカルデータ解析1」
- ・「医薬向けSASによる生存時間解析」

#### [学習内容]

使用するテキストでは、新医薬品の臨床試験において頻繁に使用される統計解析について、その理論的側面よりもSASによる実際の解析手順、結果の解釈を主眼として書かれています。

【後編にて学習:2009年11月19日-20日開催】

3章では、安全性・診断データにおける基準範囲の推定と比較について学習します。第4章では、群逐次デザインにおける解析(繰り返し有意差検定、エラー消費関数、確率的打ち切り検定)を種々のSASマクロ・プログラムを用いて学習します。

[コンテンツ]

【後編にて学習:2009年11月19日-20日開催】

3. Analysis of Safety and Diagnostic Data

4.Interim Data Monitoring

「テキスト」

本コースは、英語版の書籍『Analysis of Clinical Trials Using SAS: A Practical Guide』(SAS Press)を使用し、日本語で説明を行ないます。また、テキストの内容を簡単にまとめた補助資料を用意します。 【注意】受講申込者は、トレーニング申込とは別に事前にご購入ください。(上記書籍を既にお持ちの方は除きます。)購入方法は、下記『書籍購入方法』をご参照ください。

[書籍購入方法]

Amazon.co.jp(http://www.amazon.co.jp/)でご購入ください。

#### ● 「医学研究におけるベイズ統計学」コース (1日間)

[日 程]

2009年11月27日(金) 10:00 ~ 17:00 (東京会場)

[価格]

73,500円(税込)/※チケットのお取り扱いはありません。 「受護対象]

ベイズ統計を臨床研究・臨床試験において用いる方

[前提知識]

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を聞いたことがある

#### [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。

[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
- 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
- 事例紹介
  - ・ベイズ流試験デザインとデータ解析の事例紹介

#### ● 「SAS® 9.2 SAS/STAT 拡張点」 コース (1日間)

「日 程」

2009年11月30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 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PPPLOTを使用します。
- ・PROC TTESTでTOST同等性検定を行ないます。
- ・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ROCCONTRAST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 プロットを行ないます。
- ・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します。

# ●「SASによる臨床試験データの統計解析結果に関するバリ デーション」コース (1日間)

[日 程]

2009年12月1日(火) 10:00 ~ 17:00 (東京会場)

「価格

73,500円(税込)/※チケットのお取り扱いはありません。

#### [受講対象]

SASシステムによる臨床試験統計解析またはDM業務に従事している方。本コースは、バリデーションに関するポリシーを紹介するコースです。臨床開発経験1年未満のSASプログラマの方から、SASプログラマを管理する方まで、幅広いレベルの方々を対象としています。

#### [前提知識]

特に必要ありません。

#### [学習内容]

本コースでは、FDA(米国食品医薬品局:Food and Drug Administration) の要求事項をもとに、米国における「バリデーション」に関する概念、およびSASプログラマに期待されるポリシーなどについて、英語版書籍『Validating Clinical Trial Data Reporting with SAS』に沿って紹介します。また、SASを使用して臨床試験より得られたデータを解析するにあたり、認識しておくべきバリデーションの考え方を理解することを目指します。

※本コースは、プログラミングのテクニックなどについて説明するものではありません。 「コンテンツ〕

Chapter 1: 新医薬品の申請に関する概要

Chapter 2: バリデーションとは

Chapter 3: 文書化について

Chapter 4: バリデーションのための一般的な方法

Chapter 5: データを受け渡しする際の注意点

Chapter 6: 一般的なデータの形式

Chapter 7: 解析結果のバリデーション

#### [テキスト]

本コースは、英語版の書籍『Validating Clinical Trial Data Reporting

with SAS』(SAS Press)を使用し、日本語で説明を行ないます。また、 テキストの内容を簡単にまとめた補助資料を用意します。

【注意】受講申込者は、トレーニング申込とは別に事前にご購入ください。(上記書籍を既にお持ちの方は除きます。)購入方法は、下記『書籍購入方法』をご参照ください。

[書籍購入方法]

Amazon.co.jp(http://www.amazon.co.jp/)でご購入ください。

# 新規トレーニング・コースの開催のご案内

#### ● 「SAS® Enterprise Miner™ 分析の応用」 コース (3日間)

[日 程]

2009年12月2日(水)~4日(金) 10:00 ~ 17:00 (東京会場)

[価格

173,250円(税込)/チケット捺印数:3

[受講対象]

SAS Enterprise Minerを用いて分析を行ないたい方

[前提知識]

下記の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」

さらに下記のコースを受講済みか、SAS Enterprise Minerを使用した 分析経験のある方

■「SAS Enterprise Miner 操作入門」

#### [学習内容]

パターン発見(セグメンテーション、アソシエーション、およびシーケンス分析)や予測モデリング(決定木、回帰、およびニューラル・ネットワーク)といったデータ・マイニングにおける分析手法を、SAS Enterprise Minerを利用して実現する方法を紹介します。

[コンテンツ]

- はじめに
  - ·SAS Enterprise Miner入門
- データへのアクセスと評価
  - ・SAS Enterprise Miner 5.3におけるプロジェクトの定義
  - ・データソースの定義
  - ・データソースの検証
- 決定木による予測モデリング
  - ・モデリング用のデータソースの定義
  - ・モデル開発を目的としたデータの分割
  - ・Tree Desktop Applicationを使った決定木の成長
  - ・ツリーノードの実行
  - ・ツリーノードのオプションの使用法
  - ・予測モデリングによる結果データの理解
- 回帰による予測モデリング
  - ・線形回帰とロジスティック回帰
  - ・回帰ノードの使用法
  - ・補完ノードによる欠損値の補完
  - ・置換ノードによるデータ値の置換
  - •変数選択
  - ・回帰モデリングの出力の理解

- ニューラル・ネットワークや他のモデリング・ツールによる予測モ デリング
  - ・ニューラル・ネットワーク・モデル (多層パーセプトロンモデル) の概論
  - ・ニューラル・ネットワーク・ノードの利用
  - ・自動ニューラル・ノードを用いたモデル選択(自己学習)
  - ・SAS Enterprise Miner 5.3 の他のモデリング・ツール (自己学習)
- モデルの評価
  - ・事前確率ベクトルの定義

SASトレーニングのお知らせ

- ・モデル開発における事前確率ベクトルの効果
- ・モデル選択基準の変更
- 利益行列の定義
- ・モデルの比較ノードを用いたモデルの比較
- ・モデルのアセスメント統計量
- モデルの適用
  - ・スコア・データセットの定義
  - ・スコアノードによるデータセットのスコアリング
  - ・SASコードノードを用いたスコアリング済みデータセットのエク スポート
  - ・スコアコードの生成と利用
- 専門的なトピック
  - •アンサンブル・ノードによるモデルの統合
  - ・変数の選択
  - ・ツリーノードによる質的な入力水準の併合

- ・複雑なモデルに対するツリーノードを利用した代理表現
- パターン発見の概論
  - ・データのクラスタリングとセグメンテーション
  - ・変数の変換ノードの利用
  - ・クラスターノードを用いたk-means法によるクラスター分析
  - ・アソシエーション分析とシーケンス分析
- ケーススタディ(自己学習)
  - 銀行業務のセグメンテーションに関するケーススタディ
  - ・Webサイト利用の関連性に関するケーススタディ
  - ・クレジット・リスクに関するケーススタディ
  - ・大学入学者のマネジメントに関するケーススタディ

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

# 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

IBM版(OS/390,z/OS) SAS 9.1.3 / 9.2



#### SAS Technical News入手

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

SAS 9.1.3 / 9.2

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:03-6434-3680 FAX:03-6434-3681



Linux(Intel) 版

POWER

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

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@2009, SAS Institute Inc.All rights reserved