

STN

SAS Technical News

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



SUMMER 2010

特集 01

SAS/STAT[®]の拡張点:
9.22のご紹介

SAS Academic News 08

- 事例紹介
- コラム「SAS四方山話」

Q&A 13

SASトレーニングのお知らせ 17

最新リリース情報 20

特集

SAS/STAT[®]の拡張点: 9.22のご紹介

SAS[®] 9.2がリリースされてから、Technical News(Summer 2009, Autumn 2009, Winter 2010)にてグラフ機能に関し、新たなプロシジャ、および拡張点をご紹介しました。より簡略、きれいに描画できるグラフのみではなく、分析に用いるプロシジャとの連携、ODS統計グラフ(ODS Statistical Graphics)機能の活用に関し、分散分析を例題として記述しました。また、SAS 9.2 SAS/STATの拡張点として、Bayesian分析に対するBAYESステートメントの追加、およびMCMCプロシジャの追加、逐次計画法に対するプロシジャの追加などをご紹介しました。

※Technical Newsのバックナンバーは以下にて参照できます。

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

SAS 9.2における最新のメンテナンス版SAS 9.2 Maintenance 3では、SAS/STATのリリースが9.22となり、特に線形モデルの推定後のプロセスに関し、多くの拡張がされています。その他にも、Poisson回帰モデルに対する正確検定、標本調査における重みを考慮した上でのCox回帰モデルに対するプロシジャ(評価版)など、機能が拡張されています。今号では、特に線形モデルの推定後のプロセスを中心に、主な拡張点をご紹介していきます。

※SAS 9.2以降では、プロダクトごとにリリース番号があります。詳細に関しては、以下のページにてご確認ください。

<http://support.sas.com/software/92/productnumbers.html>



Your Models Build up
Your
Next Analytics

1 線形モデルにおける拡張点

1.1 線形モデルとは

データ分析では、さまざまな手法、モデル名があります。回帰分析、分散分析、混合モデル、ロジスティック回帰モデル、Poisson回帰モデルなど、多くのものがあります。これらのうち、パラメータの線形式を用いたモデル式は、包括的に“線形モデル”と言及できます。

例えば、応答変数 Y 、2つの説明変数を $X1$ 、 $X2$ とする回帰モデルは、次のように記述できます。

モデル

$$Y = \beta_0 + \beta_1 X1 + \beta_2 X2 + \varepsilon$$

誤差項である ε が平均0、分散 σ^2 の正規分布と仮定した場合、以下のような記述もできます。

モデル

$Y \sim$ 正規分布(平均 μ , 分散 σ^2)

$$\mu = \beta_0 + \beta_1 X1 + \beta_2 X2$$

モデルに含まれるパラメータ、 β_0 、 β_1 、 β_2 の線形式となっています。また、二項分布を仮定する、ロジスティック回帰モデルは以下のように記述することができ、パラメータの線形式としてのモデルとなります。

モデル

$Y \sim$ 二項分布(試行数 n , 確率 p)

$$\text{logit}(p) = \beta_0 + \beta_1 X1 + \beta_2 X2$$

※ $\text{logit}(p)$ はロジット関数とも言及され、 $\text{logit}(p) = \log(p/(1-p))$ となります。

SASでは多くのプロシジャが、“線形モデル”に含まれる分析に対応しています。9.22では、これらのプロシジャにて指定できるステートメントが追加、またオプションなどが拡張されています。

1.2 モデルの推定

拡張されているステートメントは、主にモデル推定の情報を用いた、統計量、検定の算出部分となります。このため、最初にモデルの推定を行うことが必要となります。

線形モデルのプロシジャでは、MODELステートメントにて推定するモデル式を記述します。応答変数を等号の左側、説明変数を右側に記述し、‘/’の後にオプションを指定します。例えば、男女別の体重データが含まれているSASHELP.CLASSを用いた場合、以下のような記述になります。

例

```
PROC GLM DATA=sashelp.class;
  CLASS sex;
  MODEL weight = sex / solution;
RUN;QUIT
```



ここでは、説明変数として変数SEXのみとなっていますが、複数の説明変数を列記できます。また、この例のようにCLASSステートメントでは、カテゴリカル変数として扱う変数を明示します。MODELステートメントにおけるオプションの一例であるSOLUTIONは、パラメータ推定値の表示を行うものです。

この他、ランダム効果を含める場合にはRANDOMステートメント(MIXEDプロシジャ、GLIMMIXプロシジャなど)、反復効果を含める場合にはREPEATEDステートメント(GLMプロシジャ、GENMODプロシジャなど)、Bayesian分析に対応しているプロシジャではBAYESステートメントを用い、より複雑なモデルを指定し、推定ができます。

新たに9.22ではEFFECTステートメントが追加されており、スプライン関数、ラグ効果などをモデル式にも含めることができます。評価版の機能となりますが、以下の11プロシジャにて対応しています。

EFFECTステートメント(評価版)対応のプロシジャ

GLIMMIX	GLMSELECT	HPMIXED	LOGISTIC
ORTHOREG	PHREG	PLS	QUANTREG
ROBUSTREG	SURVEYLOGISTIC	SURVEYPHREG	

1.3 モデル推定後のプロセス

データ分析では、パラメータの線形式の推定、検定が必要となる 경우가多々あります。例えば、各グループ間における最小2乗平均(以下、LS平均)を算出し、比較することがあげられます。また、任意の(推定可能な)係数、対比係数を用いた、パラメータ線形式に対する推定、検定も考えられます。それぞれGLMプロシジャのLSMEANSステートメント、CONTRASTステートメントもしくはESTIMATEステートメントにて対応している内容となります。

以前のリリースでは、LSMEANSステートメント、CONTRASTステートメント、ESTIMATEステートメントは、一部のプロシジャのみで用いることができました。9.22では、より多くのプロシジャにおけるステートメントとして拡張されています。また、新たなステートメントとしてEFFECTPLOTステートメント、LSMESTIMATEステートメント、SLICEステートメントが追加されており、モデル推定後に対するプロセスが拡張されています。各プロシジャ、および対応しているステートメントに関しては、以下をご参照ください。

各プロシジャでのステートメントの対応

Procedure	CONTRAST	EFFECTPLOT	ESTIMATE	LSMEANS	LSMESTIMATE	SLICE	TEST
GENMOD	★	✓	★	★✓	✓	✓	
GLM	★		★	★			★
GLIMMIX	★		★	★	★✓	✓	
LOGISTIC	★	✓	✓	✓	✓	✓	★
MIXED	★		★	★	✓	✓	
ORTHOREG		✓	✓	✓	✓	✓	✓
PHREG	★		✓	✓	✓	✓	★
SURVEYLOGISTIC	★		✓	✓	✓	✓	★
SURVEYPHREG			✓	✓	✓	✓	✓
SURVEYREG	★		★	✓	✓	✓	✓

注)✓は新しくステートメントが追加、★は既存のステートメント、★✓はステートメントが更新されています。

1.4 ステートメントの概要

各ステートメントではさまざまなオプションを指定することができますが、具体的なプログラムの書き方に関しては、ドキュメントをご参照ください。ここでは、各ステートメントの概要のみ記述します。

● CONTRASTステートメント

パラメータの線形式、もしくは複数の線形式に対する検定を行うためのステートメント。推定が可能であるように、指定しているモデル式を考慮し、対比係数を指定する必要があります。

● EFFECTPLOTステートメント

さまざまなグラフ、効果プロットを作成するためのステートメント。グラフはODS統計グラフ機能にて作成され、視覚的に分析の結果を把握することに有用です。

● ESTIMATEステートメント

パラメータの線形式に対し、推定値、および検定を行うためのステートメント。CONTRASTステートメントと同様、推定が可能であるように、指定しているモデル式を考慮し、対比係数を指定する必要があります。

● LSMEANSステートメント

モデルに含まれるカテゴリカル変数の各水準に対する、LS平均を算出するためのステートメント。DIFFオプションを用いることによって、LS平均の差を算出でき、ADJUST=オプションにて多重性に対する調整の手法を指定できます。

※GLMプロシジャでは、DIFFオプションではなく、PDIFF、もしくはTDIFFオプションとなります。

● LSMESTIMATEステートメント

LS平均の線形式に対する推定、検定を行うためのステートメント。LSMEANSステートメントとは異なり、各LS平均の値は算出されません。また、ESTIMATEステートメントでは“モデルパラメータ”の線形式ですが、このステートメントでは“LS平均”の線形式を用いますので、より分かりやすい指定の方法となります。

● SLICEステートメント

カテゴリカル変数の2次以上の交互作用項に対し、1つの変数の水準を固定(スライス)した分析に対するステートメント。LSMEANSステートメントと同じオプションを用いることができます。

● TESTステートメント

指定した効果に対する検定を行うためのステートメント。検定のタイプ、また検定に用いる自由度を指定することができます。

各ステートメントに対するより詳細な内容に関しては、9.22のドキュメントにおける“Shared Concepts and Topics”の箇所をご参照ください。また、以下のページから同様のドキュメントを参照できます。

<http://support.sas.com/documentation/922/index.html>

2 モデル情報の保存(STOREステートメント)、活用(PLMプロシジャ)

モデルを推定した後のプロセス、ステートメントは、プロシジャの実行とともに記述する必要があります。LS平均を算出、表示したい場合には、プロシジャの実行時に、LSMEANSステートメントを記述します。また、プログラムを実行した後、他のステートメント、オプションを用いた結果が必要となった場合には、ステートメント、オプションを追加した上で、もう一度、プログラムを実行します。オブザベーション数が多い、モデル式が複雑である場合など、モデル推定に再度、多くの時間をかけてしまうことになります。

9.22では、推定されているモデル情報を呼び出し、統計量の算出、グラフの表示などを行うプロシジャ、PLMプロシジャが追加されています。モデル情報に基づく実行となりますので、再度の入力データセットの参照、線形モデルのプロシジャを実行することなく、モデル推定した後のプロセスを実行できます。

2.1 モデル情報の保存: STOREステートメント

PLMプロシジャは、他のプロシジャとは異なり、データセットを入力するのではなく、線形モデルのプロシジャで推定されたモデル情報を呼び出し、処理を実行します。このモデル情報、“アイテムストア”はSTOREステートメントにてバイナリファイルとして保存でき、以下の10プロシジャにて指定できます。

STOREステートメント対応のプロシジャ

GENMOD	GLIMMIX	GLM	LOGISTIC
MIXED	ORTHOREG	PHREG	SURVEYLOGISTIC
SURVEYPHREG	SURVEYREG		

例えば、SASUSERライブラリにSTUDYRESULTSという“アイテムストア”を作成する場合、以下のように記述します。

例

```
PROC GLM DATA=sashelp.class;
  CLASS sex;
  MODEL weight = height sex|age;
  STORE sasuser.StudyResults; /* アイテムストアに保存 */
RUN;QUIT;
```

※1 既存の“アイテムストア”名を指定した場合、モデル情報が更新されますので、注意が必要です。

※2 “アイテムストア”の指定は、libname.membrnameとなります。ライブラリ名(libname)が指定されていない場合、一時的なライブラリWORKがデフォルトとなります。“アイテムストア”を呼び出し、活用する観点から、永久ライブラリを用いることをおすすめします。

2.2 モデル推定後のプロセス: PLMプロシジャ

線形モデルのプロシジャにてモデル情報を“アイテムストア”として保存した後、PLMプロシジャで、LS平均の算出、グラフの作成などを実行します。“アイテムストア”を呼び出すため、PLMプロシジャではRESTORE=オプション(またはSOURCE=オプション)を指定します。例えば、前述の

プログラムにて作成した“アイテムストア”sasuser.StudyResultsを呼び出し、各性別のLS平均、およびLS平均の差を算出するためには、次のプログラムとなります。

例

```
PROC PLM RESTORE=sasuser.StudyResults; /* アイテムストアの呼び出し */
  LSMEANS sex / DIFF;
RUN;
```

2.3 ステートメントの概要: PLMプロシジャ

PLMプロシジャでは、10のステートメントがあり、大きく分けて2つに分けることができます。

2.3.1 推定、検定などに用いる7ステートメント

RESTORE=オプションで呼び出したモデル情報に基づき、グラフの作成、統計量の算出を行うためのステートメントとなります。

EFFECTPLOT	ESTIMATE	LSMEANS	LSMESTIMATE
SCORE	SLICE	TEST	

各ステートメントの概要に関しては、1.4節と同じとなりますが、CONTRASTステートメントに関しては、PLMプロシジャではサポートされていません。かわりに、ESTIMATEステートメントにて複数の対比を指定することができ、JOINTオプションを用いることでF検定の結果が求められます。

2.3.2 表示、フィルタリングなどに用いる3ステートメント

PLMプロシジャの表示に関し、以下の3つのステートメントがあります。

● FILTERステートメント

PLMプロシジャにて表示される結果に対し、フィルタリングを行うステートメント。LSMEANSステートメントなどで出力される統計量(検定統計量、p値など)に基づき、部分的に結果を表示するよう、フィルタリングすることができます。

● SHOWステートメント

アイテムストアに保存されているモデル情報を、表示するためのステートメント。モデル推定に用いたデータセット、モデル式、変数名、またプログラムなどを確認できます。

● WHEREステートメント

BY変数の水準にて、出力する結果を制限するためのステートメント。モデル情報にBY変数が含まれている場合、BY変数の値にて結果の表示を制限できます。

※BYステートメントにて指定している変数をBY変数と言及します。

2.4 例題

線形モデル対応のプロシジャにて拡張されたステートメント、またPLMプロシジャに関し、概要を記述しました。ここでは、プログラムの記述などを含めた例題をご紹介します。

2.4.1 例1: 共分散分析

SASHELPライブラリには、データセットCLASSがあり、各性別(SEX)における体重(WEIGHT)、身長(HEIGHT)、年齢(AGE)などの値が含まれています。例えば、身長、年齢、性別から体重を求めるような場合、最初に以下のようなプログラムを実行します。

例

```
PROC GLM DATA=sashelp.class;
  CLASS sex;
  MODEL weight = height sex|age;
  STORE sasuser.StudyResults;
RUN; QUIT;
```

※2.1節にても同じプログラムを例として記述しています。

新たな機能STOREステートメントを用い、モデル情報を“アイテムストア”として保存している点が、9.22より前のリリースとは大きく異なっています。次に、モデルの推定、GLMプロシジャを実行した後、性別によって体重に差があるかを検証する必要が生じたとします。以前は、再度、LSMEANSステートメントを加えた、GLMプロシジャを実行する必要がありました。しかしながら、新たなPLMプロシジャを用い、以下のプログラムを実行することによって、結果(図1)を求めることができます。

例

```
PROC PLM RESTORE=sasuser.StudyResults;
  LSMEANS sex / DIFF;
RUN;
```

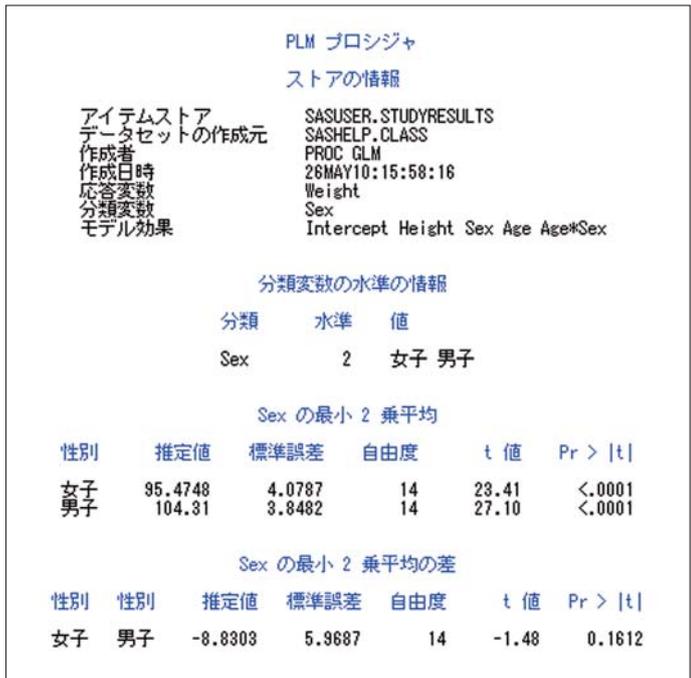


図1: 体重差に関する統計量1

※女子のLS平均:95.48ポンド=約43.31kg、男子のLS平均:104.31ポンド=約47.31kg、LS平均の差:-8.80ポンド= -約4kg

最初に保存されている“アイテムストア”の情報が出力されてから、各性別におけるLS平均、およびLS平均の差が出力されています。モデル式では、性別SEXの他、2つの連続変数HEIGHT、AGEが含まれており、LS平均を算出する際、データセットSASHELP.CLASSにおける算術平均の値(HEIGHT=62.34インチ(約158.3cm)、AGE=13.316歳)がそれぞれ用いられています。

変数HEIGHT、AGEの算術平均ではなく、任意の値におけるLS平均、および差を算出したい場合には、LSMEANSステートメントにてATオプションを追加します。(図2参照)

例

```
PROC PLM RESTORE=sasuser.StudyResults;
  LSMEANS sex / DIFF AT (height age)=(60 13);
RUN;
```

※HEIGHT=60インチ(約152.4cm)



図2: 体重差に対する統計量2

任意の身長、年齢において、性別によって体重に差があるかを上記にて検証することができます。

共分散分析では、この他、性別ごとに予測値を用いたグラフの表示が有用となります。GLMプロシジャのMODELステートメントでは、1つ、もしくは2つのカテゴリカル変数、1つの連続変数を指定している場合には、ODS統計グラフ機能を有効としておくことによって、グラフ(ANCOVAPLOT)が描画されます。しかしながら、この例題のように、2つの連続変数がモデル式に含まれている場合には、このグラフは作成されません。一方、PLMプロシジャのEFFECTPLOTステートメントを用い、以下の指定にて描画できます。(図3参照)

例

```
ODS GRAPHICS ON;
PROC PLM RESTORE=sasuser.StudyResults;
  EFFECTPLOT SliceFit;
RUN;
```

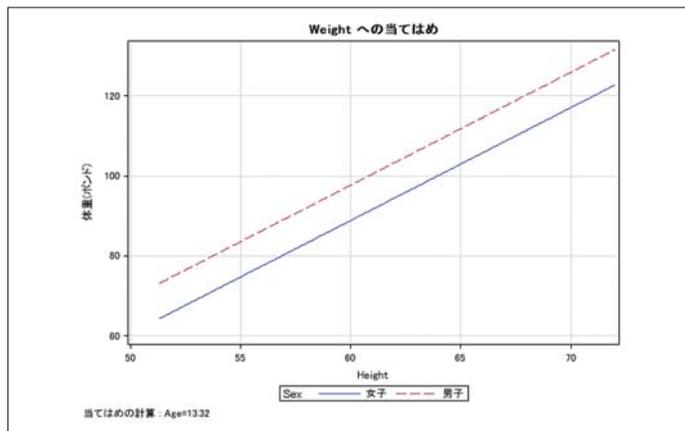


図3: 性別ごとのラインプロット

LSMEANSステートメントと同様、変数AGEの値としては算術平均を用いており、グラフの左下に表示されています。任意の値に変更するには、ATオプションを用います。

例

```
ODS GRAPHICS ON;
PROC PLM RESTORE=sasuser.StudyResults;
  EFFECTPLOT SliceFit / AT (age=13);
RUN;
```

この例では、LS平均の算出、グラフの表示を行いました。変数HEIGHT、AGEの値を変更しながら、統計量の算出、グラフの表示をPLMプロシジャにて実行していますが、入力データセットCLASSの参照、GLMプロシジャ実行でのモデル情報“アイテムストア”の作成は一度のみであることを今一度ご注目ください。また、グラフ表示に用いたEFFECTPLOTステートメントは、GLMプロシジャではサポートされていません。しかし、STOREステートメントにてモデル情報を“アイテムストア”として保存することで、PLMプロシジャでのEFFECTPLOTステートメントを用いることができます。PLMプロシジャを用いる、一つの利点となります。

2.4.2 例 2: スコアリング

分析を行う一つの目的として、推定したモデル情報を活用した、予測値の算出があります。簡単なケースとしては、身長から体重を予測することがあげられます。この場合、最初にCLASSデータセットのような体重、身長のオブザベーションからモデル式を推定し、そのモデルに身長値をあてはめ、体重の予測を行うこととなります。また、ある事象が生じるかの予測を必要とするケースがあります。このような状況に対し、ロジスティック回帰モデルがあり、LOGISTICプロシジャなどにてモデルの推定を行うことができます。

LOGISTICプロシジャにて推定したモデル情報を保存し、予測確率を算出するには、SAS 9以降ではOUTMODEL=オプション、INMODEL=オプションを用いることができます。最初にモデル推定した結果をOUTMODEL=オプションにてデータセットとして出力しておき、予測確率を算出するステップにてINMODEL=オプションにてデータセットを呼び出します。詳細に関しては、以下のFAQをご参照ください。

<http://www.sas.com/jp/service/technical/faq/list/body/stat089.html>

しかし、これらのオプションは他の線形モデルに関するプロシジャではサポートされていないため、プロシジャごとに異なる手法を用いる

必要がありました。この点、9.22では、プロシジャを実行するときに、STOREステートメントにてモデル情報を“アイテムストア”(例: sasuser.ScoreModel)として保存しておくことによって、以下のプログラムにて予測値の算出を行うことができます。

例

```
PROC PLM RESTORE=sasuser.ScoreModel;
  SCORE DATA=score_data out=score_result PREDICTED / ILINK;
RUN;
```

データセットSCORE_DATAのオブザベーションに対し、予測値(キーワードPREDICTED)を算出し、データセットSCORE_RESULTに出力する指定となっています。また、ロジスティック回帰モデルでは、ロジット変換などのリンク関数を用いていますので、その逆変換した値、予測確率を求めるため、ILINKオプションを併せて指定しています。また、すでに作成されている“アイテムストア”を用いる場合、モデル推定のプログラムコードを確認したいことがあります。この場合には、SHOWステートメントを用いた以下のプログラムを実行します。(図4参照)

例

```
PROC PLM RESTORE=sasuser.ScoreModel;
  SHOW Program;
RUN;
```

SAS プログラムの情報

```
proc logistic data=PastTasteTests;
  effect splines = spline(x1-x20/separate);
  model good=splines/selection=stepwise;
  store sasuser.ScoreModel;
run;
```

図4: プログラムの確認

多くのオブザベーション、変数を用いた入力データセット、また、複雑なモデル式などを用いている場合には、プロシジャの実行に多くの時間が必要となりますので、STOREステートメントにてモデル情報を保存しておくことが有用となります。

3 他の主な拡張点

線形モデルのプロシジャにおけるステートメントの拡張、また、新たに追加されたPLMプロシジャに関してご紹介しました。この他、さまざまなプロシジャにてステートメント、オプションなどが拡張されています。ここでは、主な拡張点についてご紹介いたします。

● CALISプロシジャ

共分散構造分析に対し、多くの拡張点を加えたTCALISプロシジャが評価版としてありました。9.22ではこれらの機能は、CALISプロシジャに含まれ、正規版として実行できます。詳細に関しては、次の文献をご参照ください。

Structural Equation Modeling and Path Analysis Using PROC TCALIS in SAS® 9.2

<http://www2.sas.com/proceedings/forum2008/384-2008.pdf>

※SAS/STAT 9.22にてTCALISプロシジャを実行した場合、以下のメッセージがログに表示されます。

WARNING: The capabilities in the experimental TCALIS procedure are included in the CALIS procedure of SAS/STAT 9.22. Please resubmit your statements to PROC CALIS

● FREQプロシジャ

Kendallのtau-b、Stuartのtau-c、SomersのD統計量に対し、正確な検定を行うためのオプションKENTB、STUTC、SMDCR、SMDRCがEXACTステートメントに追加されています。

層別2x2表に対し、TABLESステートメントにてGAILSIMONオプションが追加され、質的交互作用項の検証に用いるGail-Simon検定が行えます。また、Cochran-Mantel-Haenzel統計量を求めるCMHオプションに、MFサブオプションが追加されており、Mantel-Fleiss Criterion統計量を求めることができます。

● GENMODプロシジャ

カウントデータにおいて観測値として多くのゼロが含まれる場合、ゼロインフレーションモデルを用いることがあります。Poisson分布に加え、負の二項分布を考慮したZero-Inflated Negative Binomialモデルが追加されています。

例

```
PROC GENMOD DATA=one;
  MODEL y=x / DIST=ZINB;
  ZEROMODEL x;
RUN;
```

MODELステートメントにてDIST=ZINBオプションを指定し、ゼロインフレーションモデルにて用いる説明変数を記述します。

また、Poisson回帰モデル、ロジスティック回帰モデルにおいて、EXACTステートメントを用いることができ、パラメータに対する正確な検定を行うことができます。

● HPMIXEDプロシジャ

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、またオブザベーション数が非常に多い場合に対し、HPMIXEDプロシジャが評価版として追加されていました。9.22にて正規版のプロシジャとして実行できます。

● LIFETESTプロシジャ

生存関数の推定として、Breslow(METHOD=BRESLOW)とFlemming-Harrington(METHOD=FH)の手法が追加されています。

● MCMCプロシジャ

事後確率分布に基づき、応答変数のランダムサンプルを作成し、データセットに出力するPREDDISTステートメントが追加されています。

● PHREGプロシジャ

各時点におけるリスクセットを表示するため、オプションATRISKを指定できます。また、BAYESステートメントにおいて、Zellner g-priorが指定できる事前分布に追加されています。

● SURVEYPHREGプロシジャ (評価版)

標本調査における分析として、SURVEYにて始める複数のプロシジャがあります。新たに、Cox回帰モデルを推定するためのプロシジャとしてSURVEYPHREGプロシジャ(評価版)が加えられています。

● VARIOGRAMプロシジャ

2次元の空間データに対して(セミ)バリオグラムを算出し、新たに追加されているSTOREステートメントにて“アイテムストア”として保存できます。この“アイテムストア”は、クリギングに対応しているKRIGE2Dプロシジャ、2次元Gauss確率場におけるシミュレーションに用いるSIM2Dプロシジャで、RESTOREステートメントにて呼び出し、利用することができます。

4 参考文献

線形モデルにおけるステートメントの拡張、PLMプロシジャの追加などに関しては、米国にて開催されたSAS Global Forum 2010にて紹介されており、以下の文献にて詳細を参照できます。

The Next Generation: SAS/STAT® 9.22

<http://support.sas.com/resources/papers/proceedings10/264-2010.pdf>

Introducing PROC PLM and Postfitting Analysis for Very General Linear Models in SAS/STAT® 9.22

<http://support.sas.com/resources/papers/proceedings10/258-2010.pdf>

また、1.4節にてご紹介しましたが、SAS/STAT 9.22のドキュメントは以下のページにて参照することができます。SAS 9.2 Maintenance 3では、SAS/ETS®、SAS/OR®のリリースも9.22に更新されており、ドキュメントが公開されています。

<http://support.sas.com/documentation/922/index.html>

5 おわりに

SAS/STAT 9.22における拡張点として、主にモデル推定後のプロセスに関し、拡張されているステートメント、PLMプロシジャをご紹介しました。一度のモデル推定から、さまざまなアングルで結果を活用、検証する上で今号における内容が参考となれば幸いです。

SAS アカデミック・ニュース Academic News

今号から、事例紹介と四方山話の新連載がスタートします。事例紹介では、國學院大學経済学部および人間総合科学大学人間科学部兼任講師 高柳良太先生が大学での講義で行う分析ツールとして SAS® Enterprise Guide® を選ばれました。どのように講義で効果的に SAS Enterprise Guide を取り入れ、どのようにコンピュータ操作から分析の解釈まで学生達へ伝えるかなど苦悩と新たな発見が綴られています。四方山話では、大橋先生による「身近な教材で SAS を学ぶ〜 SAS ヘルプの使い方〜」を4回にわたりご紹介いたします。SAS ヘルプは、各ステートメントのシンタックスやオプション等を調べるときに使用します。SAS ヘルプをあえて教材として利用できる理由は、検索したプロシジャや関数の説明が英語で記述されていますが、数学的背景からの説明やプログラムの例題もあるため、あえて他の文献で調べることもなくヘルプから必要な情報を入手できるからです。



事例紹介

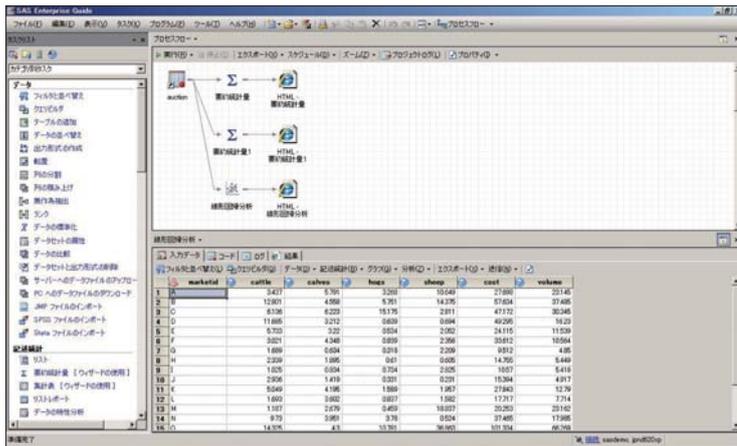
コラム
「SAS 四方山話」

事例紹介

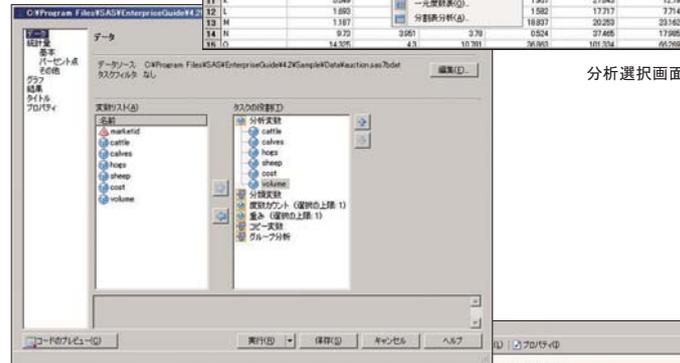
第1章 Enterprise Guide の特徴とメリット

高柳 良太

國學院大學 経済学部および人間総合科学大学 人間科学部 兼任講師



EGの全体画面



分析指定画面

EGとは

SAS Enterprise Guide(以下EGと略)は、SASにあってマウス操作で分析の指定が可能なアプリケーションです。プログラム作成なしに、マウス操作だけで分析を実行することができます。形式さえ整っていれば、Excelのデータをそのまま読み込んで分析することもできます。出力設定をHTMLにすれば、大抵の環境で見ることができます。

とここまで書いて、この文章を読んでいるみなさんがEGに対してどれくらい興味を持っていらっしゃるか、疑問があります。というのも、SAS使用歴の長い方ほどEGを使っている可能性は少ないように思うからです。

みなさんは、どのようにSASでデータ解析を行っていらっしゃるのでしょうか。やはり、SASを起動して、エディタ上で各プロシジャを用いてプログラムを作成しているのでしょうか。そういう方にとって、ご自分でデータを解析される範疇においては、新たに

EGを習得するメリットはあまりないかもしれません。今まで身につけたやり方で行うほうが、絶対に作業が早いからです。EGでできることはSASでもできますから、SASのコマンドをマスターしている方は、メリットは少ないといえるでしょう。

しかし、まったくSASに触れたことがない人や、データ解析が初めてという人にSASを使ったデータ解析を教える場合を考えてみてください。もちろん、自分がやってきたのと同じやり方を教えるほうが、教える側にしてみれば楽に決まっています。私もできればそうしたいです。ところが、それが許されない状況になっていることも、私と同様にみなさんもお感じになっているかもしれません。

最近のコンピュータ教育と大学生

個人的なことですが、私がSASに最初に触れたのは大学生の時でした。私の大学

では、SASは計算機センターの端末でしか使用できませんでした。そのため、授業で使う前に大学の計算機センターに利用者申請をして、IDをもらいパスワードを登録し、利用講習会に参加しなければなりません。それでも使い方がよくわからず、センター主催のFortranとSASの講習会に参加したりしてました。当時、PCはNECのPC98シリーズが幅を利かせ始めた頃でした。PCの性能は今よりも格段に低く、価格は今より比べものにならないくらい高く、国立大学に通う貧乏学生にはおいそれと手を出せないものでした。ましてやインターネット接続など、電話料金が定額接続ではなく従量接続しかない時代ですから、

分析選択画面

変数	分析	結果
marketid	単項分析	10549
cattle	単項分析	14376
calves	単項分析	2811
sheep	単項分析	2082
hogs	単項分析	2026
cost	単項分析	2206
volume	単項分析	2354

結果画面

MEANS プロシジャ

変数	平均	標準偏差	最小値	最大値	N
cattle	6.1562632	4.1999054	1.0250000	14.3250000	19
calves	3.8126842	1.9605874	0.6340000	9.0430000	19
sheep	2.8454211	3.9196619	0.3180000	15.1750000	19
hogs	5.9064211	5.0215219	0.2310000	35.9820000	19
cost	35.2934211	21.0294433	5.9120000	101.3340000	19
volume	18.7207895	14.5724846	4.8500000	66.2690000	19

教育学部の学部生には想像すらできない時代でした。

時は流れ、コンピュータやネットワークを取り巻く環境は激変しました。2006年からは高校で教科「情報」が必修科目となり、中学の技術の時間にもコンピュータの単元があります。私の息子の通う公立の小学校にも、当たり前のようにPC教室がある時代となりました。大学に進学する家庭なら、一家に1台のPCは当たり前、高校や大学の入学祝いにPCを買ってもらうのも全く珍しくないですし、価格的にもリーズナブルになってきました。

大学のコンピュータ環境も様変わりしました。今は入学と同時に大学のポータルサイトのIDが付与され、履修登録から成績管理、レポート提出までポータルサイトで行う大学も珍しくなくなってきました。ポータルサイトを使えば、家から履修登録やレポート提出もできます。休講情報もポータルサイトに表示され、設定さえすれば携帯に転送されるので、「せつかく1限にいったのに休講!」のような事態は過去の話になりました。今の大学生にとってコンピュータは欠かせないものとなり、大学入学までにコンピュータに触れている経験も飛躍的に増えています。

しかし、私たちが大学生だった頃と今の大学生の間で、コンピュータに対する決定的な考え方の違いがあります。今の多くの大学生にとって、コンピュータは計算機ではなくメディア端末なのです。自習用に解放されている大学のコンピュータ教室は、学生であふれています。もちろんレポートの作成や、ポータルサイトにアップされている授業資料の印刷といった学業利用の学生もいます。しかしそのほかに、動画サイトやコミュニティサイトの閲覧に勤しむ学生も大勢います。あまりのことに、帯域制限をかけて動画サイト等への接続を制限している大学もあると聞きます。

最近の大学生にとって、コンピュータは「他人が作ったものを見るもの」か「仲間と連絡を取るもの」といったメディアとしての使用が大半です。そしてそれは、GUI画面でマウスクリックを中心に行われます。従来の計算機としてデータの解析をしたりといった使い方を学生はわずかです。私の教えている大学でも、データ解析系やプログラミングの授業は年によっては受講者が1桁で、翌年度減コマの対象となってしまうたりします。一方、プレゼンテーションやWebコンテンツ作成といった講義は、受講可能数の数倍の希望があるほどの人気で、抽選に落ちた学生からの要望で

コマ数を増やしても、希望するすべての学生を収容することができません。

コマンドプロンプトってなんですか

私たちにとっては、コンピュータは計算機でした。もちろん、レポート作成もしましたが、ワードプロセッサではなくTeXで組版をしないと数式が混じった文章はレポートにできなかった時代ですから、今のGUI環境とは全く異なると言ってもいいでしょう。

何をやるにしても、コマンドを入力して実行結果を得るとというのが当たり前でした。今の学生は、コマンドプロンプトを見たこともなければ、Enterで実行できることも知りません。理工系の学生でも、PCでコマンド入力をさせることがあるのは、一部の大学や学科に限られるようです。理工系の大学生でも、コマンド画面をお目にかかったことがないというのはめずらしいことではありません。

コマンドを入力して結果を得る、私たちの時代には当たり前だったことが、もはや当たり前ではないのです。授業でExcelを使って簡単な関数を指導している際に、入力ただけでEnterキーを押さずにずっと待っている学生がよくいます。Enterキーを押して結果を得るという考えに至らないのです。そんな学生でも、動画サイトで違法動画を見たりすることはでき、あまつさえ自分のブログを開設していたりします。コンピュータは普及しましたが、それはインターネットに接続されたメディアとして普及したのであり、私たちが思っているような、もしくは学んできたような計算機としては普及しなかったのです。いい悪いという問題ではなく、多くの大学においてコンピュータ教育は計算機の教育ではなくメディア教育であり、コンピュータはレポートを含めプレゼンテーションツールなのです。

入門教育ツールとしてのEG

そのような経験を経てきた人にとって、コマンド入力を強いるSASを指導することは非常に抵抗の高いものです。教える方は、コマンド入力の経験がなく、コンピュータ操作を基本的にマウスクリックで行ってきた人に、コマンド入力やプログラミングを教えることは、コンピュータ経験のない人に初めて教える以上に大変な負担だということに気づかされます。文字入力以外はマウス操作がコンピュータオペレーションの中心という人に、コマンド操作を教えると「なんでこんな面倒なことしなきゃいけないの」「コマンドって面倒なだけじゃん」という非常に強い抵抗を訴えられることがあります。

マウス操作が中心だった人には、コンピュータとは「そういうもの」という概念ができあがっているので、コマンド入力はそれを壊していかないとけないのです。これが意外と大変です。もともと、これは私の教え方に問題があるという気もしますが。

コマンド入力の経験がない人がほとんどをしめるような現在の状況で、GUIでマウスオペレーションが可能なEGは、教える側の労力は非常に軽減されます。OfficeソフトライクのEGであれば、初めて使う人の抵抗は小さく、こちらとしても余分なコンピュータ操作の指導をすることなく、データ解析の本筋の説明に入ることができます。

私たちが自分のデータ解析用途に使用するのであれば、EGを使うメリットは少ないかもしれませんが、統計解析も解析ツールも初めてという人を相手にレクチャーをする場合、EGを使用するメリットは絶大です。少なくとも、コマンド入力に対する心理的な抵抗に遭うこともなく、最小限のSASの説明と、本来の目的であるデータ解析の内容について教えることに注力ができるからです。

EGを概観する

EGはSASのなかでもインタフェースがMicrosoft Officeなどのアプリケーションに近いものです。従って、「PCといえばネットとメール、少しワープロ使って、表計算はちょっとだけ」といったようなコンピュータ経験しかない人にも、かなり容易に分析技能を習得させることが可能です。

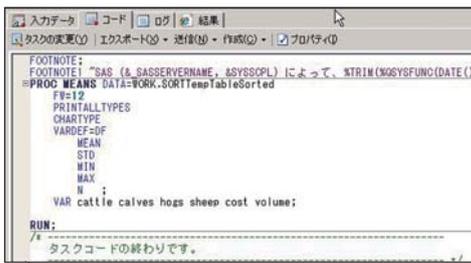
marketid	cattle	calve	cal	heifer	cow	volume
1	3,437	1,054	1,054	27,680	28,745	28,745
2	12,901	4,998	5,391	14,375	57,604	37,485
3	8,786	6,223	18,176	2,811	47,172	30,345
4	11,685	3,212	6,629	6,654	49,256	16,23
5	9,720	3,221	8,894	2,082	24,115	11,539
6	3,021	4,346	6,669	2,259	29,612	10,994
7	1,689	6,654	5,318	6,606	14,706	5,449
8	2,339	1,895	641	2,086	1,577	5,619
9	1,026	3,024	3,734	2,625	10,571	4,917
10	5,049	4,346	6,669	1,927	27,843	12,730
11	1,690	3,024	3,734	1,922	17,717	7,714
12	1,197	2,479	3,336	1,927	20,283	23,162
13	9,720	3,221	8,894	6,604	37,485	17,905
14	14,926	4,1	19,781	36,683	103,324	66,769

記述統計メニュー

marketid	cattle	calve	cal	heifer	cow	volume
1	3,437	1,054	1,054	27,680	28,745	28,745
2	12,901	4,998	5,391	14,375	57,604	37,485
3	8,786	6,223	18,176	2,811	47,172	30,345
4	11,685	3,212	6,629	6,654	49,256	16,23
5	9,720	3,221	8,894	2,082	24,115	11,539
6	3,021	4,346	6,669	2,259	29,612	10,994
7	1,689	6,654	5,318	6,606	14,706	5,449
8	2,339	1,895	641	2,086	1,577	5,619
9	1,026	3,024	3,734	2,625	10,571	4,917
10	5,049	4,346	6,669	1,927	27,843	12,730
11	1,690	3,024	3,734	1,922	17,717	7,714
12	1,197	2,479	3,336	1,927	20,283	23,162
13	9,720	3,221	8,894	6,604	37,485	17,905
14	14,926	4,1	19,781	36,683	103,324	66,769

分析メニュー

EGは、SASの全解析機能をカバーしてはいませんが、それでもかなりの部分をカバーしています。また、メニュー操作ではありますが、バックグラウンドではSASが動作しており、コマンドが自動生成されています。

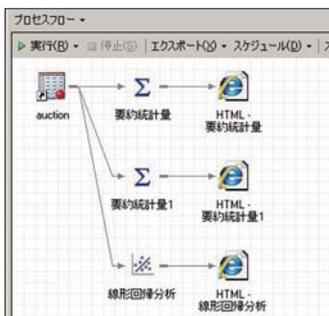


メニューから自動生成されたコード

導入にはEGを使用して、慣れてきた頃にコマンドの話をし、そこから従来のコマンドを学んでいくことももちろん可能です。分析の考え方に慣れてきたら、EGで自動生成されるコマンドプログラムに手を加えたり、そこから通常のSASの使用ということにつながってほしいのです。もちろん私たちの手持ちのSASプログラムも、EGで実行させることが可能です。見た目は違ってもSASですから、EGでも従来のデータやコマンドを使用することができます。何より、EGでSASをマスターしてもらえれば、SASという同じツールを使って作業をしてもらうことができるようになるのです。

マウス操作以外のEGのメリット

EGはご覧のように、マウス操作で分析を実行できます。それは大きなメリットです。そのほかの大きなメリットとして、プロセスが視覚化されるということがあります。実行プロセスがビジュアルに記録されるので、やり直しが楽です。これは慣れていない人に分析を再実行させる場合に絶大な威力を発揮します。



プロセスフロー

分析のやり直しは、意外と面倒なものです。もちろん、コンピュータはやり直しが得意ですが、慣れていない人に作業をやり直させるのは大変です。もちろんプログラムを組めば、やり直しは簡単です。手続きが残せる、それがプログラムの意義です。ただ、分析のやり直しを実行させるには、SASプログラムをかなり理解していないと、どれがその目的のプログラムなのか、どこを直すのかがわかりません。そこまで時間がとれる場合はいいですが、大学でも授業の科目が増え、データ解析の授業に通年を

費やせることはありません。そこでコマンド入力から教えていたら、分析に進まずに半期が終わってしまいます。業務でも、限られた人材を活用しなければならない場合、あまり悠長に時間をかけていられない事情があるでしょう。

EGの場合、プロセスが視覚化されるので、出た結果から遡って分析を再実行することが可能です。これが、通常のアンドウやダイアログリコールといった機能と違うのは、プロセス自体が保存されているので、保存してあるファイルを開いても、分析の再指定ができるということです。通常この手の機能は、実行メモリ上に保存されているので、ファイルを保存したり閉じたりすると、再実行が不可能になります。EGではプロセスが保存されるので、保存さえ正しくできれば、再実行が容易です。



プロセスフロー上での分析の再実行指定と、実行時のメッセージ

データ解析は、一回で済むことはまれだと思います。変数の組み合わせを変えるなどして、何度か分析を繰り返すこととなります。通常のマウスオペレーションのソフトウェアの場合、前述のように実行メモリに記録されるので、閉じてあるファイルを開いて、前回の続きからというような場合に、再指定が最初からになります。

マウスオペレーションの欠点としては、何となくクリックしても実行できてしまうので、どのように操作したのかを忘れやすいということがあります。EGの場合、クリックした結果が残っているので、そこから再度分析をすることもできますし、そして元の分析を残して、別個に新しい分析を残すこともできます。学生に、「前回の続きから」ということで講義を始めると、「何やったっけ?」ということが非常に多いです。そのような学生が多くても、保存さえちゃんと行われていれば、実行プロセスが残っているので、講義を進める上で非常に便利です。

HTML出力・Excelデータの使用

出力だけをエクスポートして、レポートを作成するのが容易です。SASやEGがインストールされていないPCでもOKです。

変数	平均	標準偏差	最小値	最大値	N
cattle	6.1562632	4.1899654	1.0250000	14.3250000	19
calves	3.8126842	1.9061874	0.6340000	9.0430000	19
hogs	2.8454211	3.9199619	0.3180000	15.1750000	19
sheep	5.5064211	9.0129419	0.2310000	36.9630000	19
cost	35.9342111	21.6984433	8.6200000	101.3340000	19
volume	18.7207995	14.8748446	4.8500000	66.2690000	19

HTML出力(再)

データも、形式さえ正しければExcelで作ったものをそのまま使用できるので、学生が自宅のEGのない環境でもデータファイルを作成することができます。



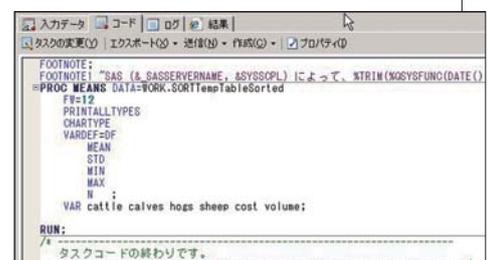
Excelファイルを読み込んだ状態

ほとんどの学生はWordやExcelの入ったPCを持っているはずなので、解析作業は大学で、データの作成やレポートの作成は自宅で行うということもできます。私が学生の頃のように、自宅にコンピュータ環境がないので、データ入力からなから全部大学の計算機センターで行わなくてはならないということもないので、学生にとっても好都合です。ちなみに、私の通っていた大学の計算機センターは17時までしか開いておらず、連続使用は2時間までだったので、非常に苦労したものです。そんな苦労もないわけです。

解析やレポート作成が、分析ソフトだけで完結するわけにはいきません。各ソフトとの連携を考えたときに、一般的に普及している環境というのは無視できません。Excelデータをそのまま利用でき、出力がHTMLにできるEGは教育利用には好都合です。

プログラミングへのつながり

そうはいつても、やはりプログラムは便利です。繰り返しの分析が大量にある場合は、特に威力を発揮します。そのようなときに、実行されているプログラムを保存して使用することができます。もちろんそこからプログラミングを教えることも可能です。



コード

前述のように、私が学生の頃はコマンド入力をしなければ、PCを使用することは不可能でした。何かをしようとするなら、プログラミングをしなければ何もできませんでした。コンピュータを使うことはプログラミングと同じだったのです。

今の人たちには、そんなことは想像もできません。コマンド入力にはミスタッチや間違いが付き物で、そのためにエラーメッセージがあるのですが、今の学生たちはエラーが出ようものなら大騒ぎです。「先生バグった！」などとあちこちで騒ぎます。余談ですが、学生たちはよくPCが予期せぬ動きをすると「バグった」と言って大騒ぎをします。元々のプログラミングのバグ(虫)という意味からすると間違っているようにも思いますし、百歩譲ってその言い方だとしても、バグっているのはその学生の操作手順なのですけど。

プログラミングの場合、どこでエラーが

出ているのかを理解させるのが至難の業です。そうこうしているうちに、講義時間などすぐに終わります。データ解析を教える前に、プログラミングやデバックの考え方を教えなくてはなりません、それをしていてはデータ解析を教える時間がなくなってしまいます。

EGは、SASですからバックグラウンドでSASコマンドが実行されています。もちろん、そのコマンドを表示させ、そのコマンドを実行することができます。

定型処理が多くなってくれば、さすがにマウス操作は面倒です。一括で一気に処理したくなります。それは、マウス操作によるオペレーションしか経験したことがない人でも感じます。そうなったときがコマンドの出番です。

EGでは実行されるコマンドを自動生成します。実行できるものですから、スペルミスやオプション指定の間違ひがありません。

まずはそこからです。そしてそこを少しずつ改造していったSASコマンドを理解させることも可能です。

ここまでで到達するのは、半期の講義では無理かもしれません。しかし研究室の学生や院生、社員なら時間はあるでしょう。そうすることで、私たちが使っているのと同じツールを使って分析する仲間を増やしていくことができるのです。そうすれば、研究や業務に広がりが出ます。

見た目の違うSASに、指導する立場から見ると抵抗があるかもしれませんが、指導のしやすさは絶大です。誰かにSASを使った分析をレクチャーしなければならないときに、是非EGを使ってみることをおすすめします。これから数回にわたって、EGで分析を行う場合のオペレーションや、EGで人に教える際のポイントなどについて書いていきたいと思います。どうぞよろしくお願ひします。

コラム「SAS四方山話」

大橋 渉
現：ヤンセンファーマ株式会社 研究開発本部 臨床統計部 統計解析グループ マネージャー 医学博士
前：東京医科歯科大学 医歯学総合研究科大学院

第10回 身近な教材でSASを学ぶ ～ヘルプの使い方(基礎編)～

新規SASユーザーの皆様

皆様の中には、この4月からSASを使い始めた(使わなければならなかった)方もいらっしゃるのではないのでしょうか。SAS社のトレーニングに出席された方、先生方や先輩方による指導を受けられた方、サルにも分かるSAS講座^{*1)}で勉強された方、いずれにしてもさまざまな形で、初めてSASとご対面された方も多かったのではないのでしょうか？

*1)筆者が新興医学出版社で連載させていただいておりました、「サルにも分かるSAS講座(月刊モダンフィジシャン)」は、この度オーム社様より、「統計を知らない人のためのSAS入門」として出版させていただくことになりました(この原稿を書いている段階では発売日未定ですが、近日発売であることは確かです)。

SASの場合はどうしても、プログラミングやSASを用いる分野(金融や生物統計など)のイメージの強さも手伝ってでしょうか、どうやら初心者にとっては「難しい」という声が強いです。実際にSAS社主催のトレーニングコースのうち、最も初心者向けである「SAS入門コース」でさえも「難しい」という声がある程ですので、我々SASに携わる一人としては、そろそろSASの初心者用教材の在り方について

考えなければならない時かもしれません。指導者の皆様にとってはごく当たり前でも、初心者の皆様にとっては非常に敷居が高く感じられてしまう事柄は結構あるということも、実はよくある話なのです(筆者も未だに上級者とは言い難いのですが…)。いきなり専門分野ならではのプログラムの洗礼を浴びせるのではなく、まずはSASに慣れましょうよ…ということで、たとえばEnterprise Guide(EG)などを有効活用してSASの雰囲気を理解していただくのも良いのではとも思います。しかしながら、一部では「プログラムしてこそSASだ!」「SASの能力を最大限に発揮できるのはプログラムだ!」「EGなんてクリックじゃないか!」等々…プログラム至上主義的な文化も根強く残っているようです。それらのご意見も確かに一理あるとは思いますが、いきなりのプログラム攻撃で、もし初心者の皆様がSASを嫌いになり、本来の目的である分析や解析に到達できなくなるのはいかなるものでしょうか? いくつかの言葉を用いるのであれば、「生物統計はやりたいけれど、SASが難しくてねえ」などという言葉が、願わくば出てこないようになればいいなあ…などと、筆者も指導時にはいつも思う次第です。

身近な教材とは? ～SASヘルプ～

そこで早速ではございますが、SASに携わる一人として教材の在り方について考えてみました。SASは特に、各分野における専門的手法に特化した使い方をされやすいソフトウェアではあるのですが、初心者にとっての基本的な部分は共通であると考えます。筆者は基本的に医薬系のユーザーではありますが、こちらのコラムではマーケット向けの内容をずっと書かせていただいております。統計の基本的手法に関わる部分は各分野とも共通です…ということで、全ての分野の皆様が共通に持っている教材…すなわちSASのヘルプについて紹介させていただきたいと存じます。



図1: ヘルプクリック時のメニュー

SASソフトウェア入門ガイド

まずは、「SASソフトウェア入門ガイド」を是非とも一度は見ただけならばと存じます。初心者の皆様にご覧いただくところ、「紹介されたことも見に行かなかった」という回答が余りにも多かったため、念のため簡単に紹介させていただきます。全くの初心者の方は、図2のプルダウンメニューから「はじめてSASプログラムを行う(クイックスタートガイド)」を選択して、Goを押して下さい。

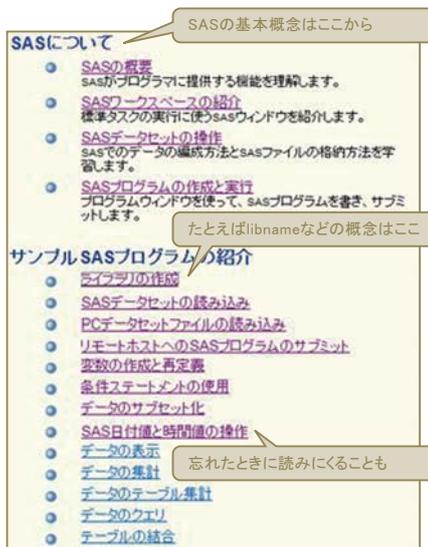


図2: 入門ガイドの主な項目

まず「SASについて」ですが、ここでは重要な2つのステップ(DATA、PROCEDURE)について、および作業場であるエディタ、ログ、アウトプット、エクスプローラのそれぞれの役割について説明がなされています。筆者もこれを見直すことによって、何故プログラムの実行ごとにログを見る必要があるのかということ初めて理解した気がします。「サンプルSASプログラムの紹介」では、ライブラリの概念やデータセットの読み込み方、日付、簡単なグラフなどの機能が紹介されています。例えば、筆者がなかなか理解できなかったlibnameの概念や、プログラムはテキストファイルでコピーアンドペースト可能であることなど、今となっては当たり前と思えるような事柄は全てこちらに記載されています。筆者は初心者時代に見ることは無かったのですが、SASを使い始めて数年後に見ても、実は結構得られる情報が多かった程でした。以上の項目ですが、SAS 9.1.3の段階では日本語で記載されていますが、9.2以降は現段階では英語記載です。バージョンアップの前に、日本語バージョンを教育用に(見直し用)に出力しておくのもお勧めです。

SASヘルプとドキュメント

一般的なソフトウェアにおけるヘルプはこの部分になります。これは先ほどの入門ガイドとは異なり、SAS 9.1.3でも現段階で英語表記になっております(それゆえあまり用いられないという説もありますが…)。

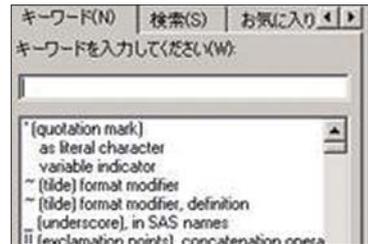


図3: キーワードの場合

図3のキーワードの場合は、記号、アルファベットの順に、特に何も検索語を入れなくても最初から用語が並んでおりますので、検索したいことが何となく頭にある場合や、興味本位でどんな用語があるのかを眺めたいときなどに使いやすいかと思えます*2)。

*2) 上級者の方でも新たな発見があることは間違いございません。筆者も「なるほど、こんなProcedureがあったのか」と感心してしまっただけは数知れませんが、たまに眺めているのも良いかと。



図4: 「検索」の場合

図4の検索の場合は、キーワードを与えなければ何も現れませんので、どちらかと言えば明確に調べたいことがある場合に用いるのがよろしいかと思えます。ところが、少しでも検索用語を含んでいる項目が全てリストアップされてしまいますので、少しずつキーワードを増やしながら絞り込んでいく必要があります。まずは「PROCEDURE」と入力し、具体的なプロシジャ名「FREQ」を入力します。これによりFREQプロシジャに関する記述が表示されます。その中からさらにプロシジャの構文(Syntax)、例題(Examples)、詳細(Details)と追加していくことで目的に応じてヘルプから必要な情報を徐々に絞り込みます。検索を繰り返しているうちに見えてくる部分もあると思われませんが、筆者が検索したものはおおよそ次のような感じです。

- ・SAS関数: Language References (関数の使い方の紹介)
- ・The ○○ Procedure (プロシジャの検索)
- ・Overview (○○Procedureとは何か…について解説をしてくれております)
- ・Getting Started (文字通り「始めてみよう」です。簡単な例題と共に解説が掲載されておりますので、特にプログラム部分はエディタにコピーアンドペーストして実行してみましょう!)
- ・Syntax (文法です。プログラム時に迷った時などは、こちらを参照してみましょう。各項目をクリックすると、さらに詳細な解説が出てきます)

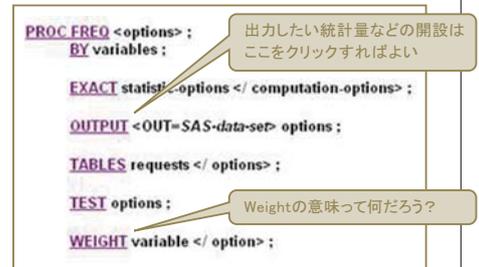


図5: Syntax画面の項目(Freq Procedureの場合)

- ・Details (たとえば統計的な手法について、数学的背景から説明などがなされております。所々例題を交えながらの解説がありますので、こちらもエディタにコピーアンドペーストして実行してみましょう)
- ・Examples (事例集です。○○Procedureではこんなことができるのか…と、改めて感心させられてしまうところでもあります。こちらはコピーアンドペーストして実行するためにあると言っても良いでしょう! かなりの頻度で用います)

ヘルプは全てのSASユーザーの皆様が共通に用いることができる教材であり、インストールさえしてしまえばいつでも使うことができます。Exampleの中には、ほとんど書き換え無しで日常的に業務に用いることが可能なプログラムも数多く含まれ、筆者的には極めて有用なものも多いと思っております。英語の壁の高さを指摘される方もいらっしゃると思いますが、実行結果は全て日本語で表示されます。初心者の皆様にとっては、実はこうすること自体が勉強になり、数多いSASの機能を理解することにもつながっていくのではないかと思えるわけです。もちろん、それなりに使いこなしてこられた皆様にとっても、新たな発見や知識の再確認があることは間違いございません。

というわけで、次号はもう少しヘルプの各論に迫ってみたいと思います。

Q&A



- 二つの日付の間の週数(7日間)をカウントする方法
- WORKライブラリフォルダを複数指定する方法
- ルックアップ・テーブルを利用したデータマッチング
- EBCDIC符号付き十進数の読み込みについて
- 特定のプロダクトがライセンスされているか確認する方法
- 複数クラスターへの分類
- 先頭空白を含む可変長データをそのまま出力したい
- 変数名を一括で接頭辞付きの形にリネームしたい

Q ある特定の期間に対し、7日間を一週として計算したいと思います。INTCK関数では、週が変わる度にカウントが増えてしまうため、実際には週の数が余分に増えてしまいます。なにか他に良い方法はありますか？

A SAS 9.2より、INTCK関数の拡張点として、第4引数に‘C’を指定することによって7日間が経過するたびにカウントすることができます。下記に例を示しますので、ご参照ください。

例

```

/* SAS9.2 以降 */
DATA work.test;
  INFILE DATALINES;
  date1='01APR2010'd;          /* 基準となる年月日 */
  INPUT date2 : DATE9.;       /* 比較する年月日 */
  /* 7日間が経過する度にカウントを増やす */
  count1=INTCK('WEEK',date1,date2,'C');
  /* 週が変わる度にカウントを増やす */
  count2=INTCK('WEEK',date1,date2);
  FORMAT date1 date2 YYMMDD10.;
DATALINES;
03APR2010
04APR2010
07APR2010
08APR2010
;
RUN;
```

なお、第4引数の指定がない場合は、週が変わるたびにカウントが増えます。

INTCK関数の詳細については、米国SAS Institute Inc.が公開しているLanguage Referenceを参照してください。

<http://support.sas.com/documentation/cdl/en/lrdict/62618/HTML/default/a000212868.htm>

Q WORKライブラリに任意のフォルダを複数指定したいのですが、何か良い方法はありますか？

A SAS 9.2より前のバージョンでは、WORKライブラリフォルダに一つのフォルダのみWORKライブラリとして指定できました。SAS 9.2からは、WORKライブラリに任意のフォルダを複数指定できるように機能が拡張され、効率よくディスクを使用できるようになりました。

【 構文 】

-WORK 'WORKフォルダ定義ファイル'

WORKフォルダ定義ファイルに、WORKライブラリに任意のフォルダを指定します。WORKフォルダ定義ファイルの記述方法は次の通りです。

```

フォルダパス1
フォルダパス2
フォルダパス3
METHOD = RANDOM または SPACE
```

METHODに指定できるパラメータは次の通りです。デフォルト値はRANDOMとなります。

RANDOM : I/Oの負荷分散をするようフォルダを使用します。
SPACE : 空き容量の多いフォルダから順に使用します。

例えば、次のようなフォルダをWORKライブラリとして使用し、空き容量が最も多いフォルダをWORKライブラリとして使用する場合、

```

C:¥TEMP
D:¥TEMP
E:¥TEMP
```

作成するWORKフォルダ定義ファイルは次のようになります。ここでは、ファイル名をWORKFOLDER.TXTとし、Cドライブの直下に配置したと仮定します。

WORKFOLDER.TXTの内容

```

C:¥TEMP
D:¥TEMP
E:¥TEMP
METHOD = SPACE
```

SASV9.CFGファイル内の「-WORK」システムオプションを次のように変更します。

-WORK 'C:¥WORKFOLDER.TXT'

詳細に関しましては、次のオンラインヘルプをご参考にいただければと思います。

[WORK System Option: UNIX]

<http://support.sas.com/documentation/cdl/en/hostunx/61879/HTML/default/o-w1.htm>

[WORK System Option: Windows]

<http://support.sas.com/documentation/cdl/en/hostwin/61924/HTML/default/win-sysop-work.htm>

Q

非常に小さなデータセットに格納されている値をキーに、他のデータセットに格納されている値を取得しています。効率よく処理する方法を教えてください。

A

SAS 9.1以降から、SASにてハッシュオブジェクトが利用可能となりました。これにより、パフォーマンスの向上が期待できます。

例

```
DATA small;
  INPUT id:$8. name:$3. att:$8.;
DATALINES;
1      AAA 5550001
100    BBB 5550002
1000   CCC 5550003
10000  DDD 5550004
20000  EEE 5550005
300000 FFF 5550006
700000 GGG 5550007
800000 HHH 5550008
900000 III 5550009
;
RUN;

DATA large;
  ARRAY a(*) a1-a10;
  DO i=1000000 to 1 by -1;
    id=LEFT(PUT(i,8.));
    DO j=1 to 10;
      a(j)=RANUNI(123);
    END;
  OUTPUT;
END;
DROP i j;
RUN;

DATA match;
  LENGTH id $8 name $3 att $8;
  IF _N_ = 1 THEN DO;
    /* ハッシュオブジェクトの作成 */
    DECLARE HASH h(DATASET: "work.small");
    h.DEFINEKEY('id'); /* キー変数の定義 */
    h.DEFINEDATA('name','att'); /* データ変数の定義 */
    h.DEFINEDONE(); /* 初期化完了 */
    CALL MISSING(id,name,att); /* 初期値の指定 */
  END;
  SET large;
  rc = h.FIND(); /* 検索 */
  IF (rc=0) THEN OUTPUT;
RUN;
```

米国SAS社にて、関連するドキュメントを公開させていただいております。あわせてご確認ください。

<http://support.sas.com/rnd/base/datastep/dot/hash-getting-started.pdf>

Q

ある変数の値を\$EBCDICw.入力形式で読み込むと符号が正しく認識されず正しい値として読み込めません。どのようにすれば値を認識させられますか。

例

データ : 'F1F0F0F0D0'x
期待する結果 : -10000
出力された内容 : 1000}

A

もとのデータはz/OSなどMVS系OSなどで作成されたゾーン十進数での符号付きの値のようです。

ゾーン十進型の数値を読み込むには文字を読み込む\$EBCDICw.入力形式ではなく、数値として読み込むための入力形式となるS370FZDw.入力形式、あるいはその数値の表現方法によって、S370FZDで始まるいくつかの入力形式などを使い分ける必要があります。

今回の値であれば、S370FZDw.入力形式、もしくはS370FZDLw.入力形式をご利用いただければ、意図した数値になると存じます。

詳細につきましては、次のマニュアルにてご確認願います。

[SAS® 9.2 Language Reference: Dictionary]

=> [Dictionary of Language Elements]

=> [Informats]

=> [Working with Packed Decimal and Zoned Decimal Data]

※どのような形式にどの入力形式が有用であるかの表も記載されております。

プログラム例

```
DATA _NULL_;
  a = 'F1F0F0F0D0'x;
  x = INPUT(a,$EBCDIC5.);
  y = INPUT(a,S370FZDL5.);
  PUT x= y= ;
RUN;
```

出力例

```
13 DATA _NULL_;
14 a = 'F1F0F0F0D0'x;
15 x = INPUT(a,$EBCDIC5.);
16 y = INPUT(a,S370FZDL5.);
17 PUT x= y= ;
18 RUN;

x=1000} y=-10000
```

Q ライセンスされている全てのプロダクトを確認するには、
PROC SETINIT;
RUN;

を実行し、確認することができますが、特定のプロダクトのみライセンスされているか確認することはできますか。

A %SYSPRODマクロを使用しますと、特定のプロダクトがライセンスされているか確認することができます。
戻り値はそれぞれ下記の通りになります。

1 : ライセンスされている
0 : ライセンスされていない
-1 : 指定したプロダクト名が無効

例

```
%sysprod (GRAPH);
```

Q CLUSTERプロシジャにて階層的なクラスター分析を実行しています。この結果から複数のクラスターに分類し、各クラスターに属する対象を特定する方法はありますか。

A CLUSTERプロシジャを実行する際、OUTTREE=オプションにてデンドログラム(Dendrogram)、樹形図を作成するためのデータセットを作成します。その後、TREEプロシジャを用い、NCLUSERS=オプションにてクラスターの数を指定します。

例

```
PROC TREE DATA=tree NCLUSERS=5 OUT=out;  
  ID id_name;  
RUN;
```

※データセットTREEはCLUSTERプロシジャのOUTTREE=オプションにて作成しています。

上記にて作成されるデータセットOUTに、対象を示す変数(ID_NAME)、属するクラスターを示す変数(CLUSTER)が含まれます。

Q SASデータセットの文字変数のデータをファイルに出力する際に先頭に空白がある場合は、\$charw.フォーマットを使用することで空白を出力することが可能ですが、実際のデータが変数長よりも短い場合はデータの末尾に空白が出力されます。

先頭空白の5-6文字のデータを\$char8.フォーマットで出力した場合のイメージ:

```
-----  
□□□aaa□□  
□□bbb□□□  
□□□ccc□□  
-----
```

※□はスペースを表します

先頭空白を含む可変長のデータが複数ある場合に、該当の空白データを含んだ各々の変数の長さでデータを出力するにはどうしたらよいでしょうか。

A \$VARYINGw.フォーマットを使用することで変数データを可変長にてファイルに出力することが可能です。
\$VARYINGw.フォーマットでは変数データの変数長を含む変数名を指定します。

```
-----  
$VARYINGw. <変数長をデータとして含む変数名>;  
-----
```

\$VARYINGw.フォーマットで指定する変数はあらかじめ変数長を含むデータを用意するか、LENGTH関数を使用して該当の変数を作成します。以下にLENGTH関数を使用して\$VARYINGフォーマットを指定したサンプルを示します。

例

```
/* 元データ */  
DATA work.a;  
  INPUT no 1. data $CHAR8. ;  
CARDS;  
1   aaa  
2   bbb  
3   ccc  
;  
RUN;  
  
/* 出力するファイル名 */  
FILENAME test 'C:\a\test.csv';  
  
/* CSV形式でデータを出力します */  
DATA _NULL_ ;  
  FILE test DLM="," ;  
  SET work.a ;  
  /*****  
  PUT ステートメントにて NO 変数には「1.」フォーマット、data  
  変数には「$varying8.」フォーマットを適用する  
  *****/  
  varlen=LENGTH(data);  
  PUT no data $VARYING8. varlen;  
RUN;
```

Q

データセットの複数の変数名を一度にリネームする(例えば名前目の先頭に“a_”を付加する)ことは可能ですか？

A

以下のようなマクロを記述することで実現できます。

例

```
/* リネームマクロ
lib: ライブラリ名
dsn: データセット名
pre: 新しい変数名を作るための接頭辞
*/

%MACRO ren(lib,dsn,pre);

/* SQL プロシジャで dictionary.columns という特殊なテーブルを検索することにより、データセットに含まれる変数名を取得することができます。
*/

PROC SQL NOPRINT;

/* 変数名をスペース区切りで一連の文字列にし &names に格納します。*/
/* 変数名の個数を &cnt に格納します。*/

SELECT name, COUNT(name)
INTO :names SEPARATED BY ' ', :cnt
FROM dictionary.columns
WHERE libname=UPCASE("&lib") and memname=UPCASE("&dsn");
QUIT;

/* リネームの実行 */
DATA tmp;
SET tmp;
RENAME

/* 変数の個数分ループ */
%DO i=1 %TO &cnt;

/* 変数名のリストから変数名を1つずつ取り出します。*/
%LET name=%SCAN(&names,&i,' ');

/* 指定の接頭辞をつけてリネーム */
&name = &pre.&name
%END;
;
RUN;

%MEND ren;

/* サンプルデータ */
DATA tmp;
LENGTH sales 8 income $10 capital 8 mktcap $20;
RUN;

/* マクロの実行 */
%ren(work,tmp,a_)
```



SAS Training

SAS トレーニングのお知らせ



SAS Learning Session 2010開催のご案内

ご好評を頂いております「SAS Learning Session」を7月26日(月)~27日(火)に開催されます「SASユーザー総会」の翌日に開催する運びとなりました。

今回は、9月と12月に最新版英語書籍「Analysis of Observational Health Care Data Using SAS」を使用した特別コースを開催するにあたり、その前提となる観察研究データに対する統計解析の概要をご紹介します。この機会に是非ご参加ください。

[午前セッション]

「観察研究データ(observational data)の統計解析: Overview」

[日 程]

2010年7月28日(水) 9:30 ~ 12:30 (東京会場)

[価 格]

31,500円(税込) / ※割引・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・疫学データの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」

[講習内容]

ランダム化をとまなわない臨床・疫学研究データの統計解析では、群間の比較可能性が崩れているために生じる「交絡」と呼ばれる現象を調整することが常に要求される。交絡調整法としては、層別解析と回帰モデルに基づく方法の2つが古くから知られているが、近年、傾向スコア(propensity score)に基づく方法やいくつかの因果モデルに基づく方法も応用例が増えてきている。特に、治療や曝露が繰り返し行われる状況では、時間依存性交絡と呼ばれる現象が生じるため、通常の層別解析や回帰モデルでは交絡の調整ができないという問題が存在する。本セッションでは、以下の内容を講義する。

1. 因果効果とは
2. 傾向スコアとは
3. 時間依存性交絡とは
4. 構造ネスト平均モデルと周辺構造モデル

[テキスト]

当日資料を配布します。

[講 師]

東京大学大学院医学系研究科 准教授 松山 裕 氏

特別トレーニング・コースの開催のご案内

● 「SAS 9.2プラットフォーム管理: 拡張点」コース (2日間)

[日 程]

2010年7月26日(月) ~ 27日(火) 10:00 ~ 17:00 (東京会場)

[価 格]

115,500円(税込) / チケット捺印数: 2

[受講対象]

現在SAS 9.1.3プラットフォームの管理者で、今後SAS 9.2プラットフォームを管理する予定のある方

[前提知識]

SAS 9.1.3プラットフォームでの管理作業の経験が必要です。

[学習内容]

- ・新しいシステム管理機能
- ・新しいメタデータ管理
- ・新しいユーザー管理機能
- ・新しいWebアプリケーション管理機能

[コンテンツ]

- プラットフォームのレビュー
 - ・プラットフォーム概要
 - ・プラットフォーム管理者の役割の定義
 - ・SAS 9.2プラットフォーム管理の新機能の識別
 - 新しいシステム管理機能の識別
 - ・サーバー・コンフィギュレーションの変更点の識別
 - ・新しいメタデータ・サーバーの機能の調査
 - ・SAS®管理コンソールの新しいサーバー・モニタリングの機能の探索
 - ・新しいサーバーロギングの機能の探索
 - 新しいメタデータ管理の機能の識別
 - ・SASフォルダ構造の探索
 - ・メタデータ・セキュリティの拡張点の調査
 - ・メタデータの移動
 - ・メタデータの保守
 - 新しいユーザー管理機能の識別
 - ・メタデータの役割の調査
 - ・接続プロファイルの探索
 - ・IDと認証の管理
 - 新しいWebアプリケーション管理機能の識別
 - ・共通的なWeb管理機能の探索
 - ・Webインフラストラクチャ・プラットフォームの探索
 - ・SAS® Web Report Studio管理の拡張点の識別
 - ・SAS® Information Delivery Portal管理の変更点の識別
- ※ 内容は予告なく変更する場合があります。

●「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8月5日(木)～6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、

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による観察研究データの統計解析【前編】」コース（2日間）

[日 程]

2010年9月13日(月)～14日(火) 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。

【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。また、前編と後編に分かれていますので内容をお確かめください。

[受講対象]

SASシステムによる臨床・疫学研究データの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」

[テキスト]

本コースは、下記英語版の書籍を使用し、日本語で説明を行います。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。

『Analysis of Observational Health Care Data Using SAS』
(SAS Press)

【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。

[学習内容]

ランダム化をとまなわない臨床・疫学研究データの統計解析では、群間の比較可能性が崩れているために生じる「交絡」と呼ばれる現象を調整することが常に要求されます。交絡調整法としては、層別解析と回帰モデルに基づく方法の2つが古くから知られていますが、近年、傾向スコア(propensity score)に基づく方法やいくつかの因果モデルに基づく方法も応用例が増えてきています。上記テキストは、傾向スコアを用いた交絡調整法、周辺構造モデル(marginal structural model)や構造ネストモデル(structural nest model)といった因果モデルに基づく交絡調整法などについて、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。

[コンテンツ]

【前編にて学習：2010年9月13日-14日開催】

上記テキストの1章から4章までを学習する予定です。1章は観察研究についての概論、2章は傾向スコアによる層別解析、3章は傾向スコアによるマッチング解析、4章は二重ロバスト推定解析です。

【後編にて学習：2010年12月13日-14日開催】

上記テキストの5章、9章、10章を学習する予定です。5章は欠測データの解析、9章は周辺構造モデル、10章は構造ネストモデルによる解析です。

[書籍購入方法]

Amazon.co.jp(www.amazon.co.jp/)にてご購入ください。

●「SAS 9.2BI：拡張点」コース（2日間）

[日 程]

2010年8月26日(木)～27日(金) 10:00 ～ 17:00 (東京会場)

[価 格]

115,500円(税込)／チケット捺印数：2

[受講対象]

SAS 9.1.3でビジネス・インテリジェンス・アプリケーションの使用経験があり、SAS 9.2における新機能を習得されたいビジネス・アナリスト

[前提知識]

SAS 9.1.3ビジネス・インテリジェンス・アプリケーションを使用した経験が必要です。

[学習内容]

- ・プラットフォームの拡張点
- ・SASビジネス・インテリジェンス・アプリケーションの新機能

[コンテンツ]

- SAS 9.2の拡張点 インTRODクシヨソ
 - ・概要
 - ・共通のJob roleの紹介
 - ・コース・ロジスティックスの紹介
 - ・プラットフォームの拡張点
 - ・プロンプト・フレームワークの探索

- SAS® Add-In for Microsoft Office 4.2の拡張点
 - ・インターフェイスによる対話
 - ・役割ベースの機能の理解
 - ・サーバー接続とプロファイルの管理
 - ・結果とリストア・コンテンツのプレビュー
 - ・新しいSASタスクと拡張されたSASタスクの使用
- SAS Enterprise Guide 4.2の拡張点
 - ・ユーザー・インターフェイスによる対話
 - ・役割ベースの機能の理解
 - ・サーバー接続とプロファイルの管理
 - ・クエリビルダでの作業
 - ・新しいSASタスクと拡張されたSASタスクの使用
 - ・SASレポートでの作業
- SAS 9.2ストアプロセスの拡張点
 - ・SAS 9.2ストアプロセスの拡張点の理解
 - ・SASストアプロセスWebアプリケーションでの作業
 - ・SAS Enterprise Guide 4.2を使用したストアプロセスの構築
 - ・SAS管理コンソール9.2を使用したストアプロセスのメタデータとの対話
- SAS® Information Map Studio 4.2の拡張点
 - ・ユーザー・インターフェイスによる対話
 - ・サーバー接続とプロファイルの管理
 - ・データアイテムでの作業
 - ・フィルタとプロンプトでの作業
- SAS Web Report Studio 4.2の拡張点
 - ・ユーザー・インターフェイスによる対話
 - ・役割ベースの機能の理解
 - ・レポートの編集と表示
 - ・レポートの印刷
- SAS Information Delivery Portal 4.2の拡張点
 - ・ユーザー・インターフェイスによる対話
 - ・ナビゲーション・ポートレットの変更点の理解
- SAS® BI Dashboard 4.2の拡張点
 - ・ユーザー・インターフェイスによる対話
 - ・データソースでの作業
 - ・インジケータでの作業
- SAS 9.2 OLAP Serverの拡張点
 - ・SAS® OLAP Cube Studioのインターフェイスとの対話
 - ・キューブの構築
 - ・キューブの保守
 - ・キューブの表示

●「SAS 9.2 SAS/STAT拡張点」コース (1日間)

[日 程]

2010年9月13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

【ご注意】本コースは、英語版のテキストを使用し、日本語で説明を行います。

[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II」

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]

・次のプロシジャでODS統計グラフを作成します:

PROC UNIVARIATE(評価版)、PROC TTEST、PROC GLM、PROC REG、PROC LOGISTIC、PROC GENMOD(自己学習)、およびPROC FREQ

・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

●「SAS 9.2 Data Integration: 拡張点」コース (1日間)

[日 程]

2010年9月29日(水) 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 9.1.3でSAS® Data Integration Studioの使用経験があり、SAS 9.2における新機能を習得されたいデータ・インテグレーション開発者

[前提知識]

SAS 9.1.3でSAS Data Integration Studioを使用した経験が必要です。

[学習内容]

以下のSAS Data Integration Studioの新機能の利用方法を学習します。

- ・新しいインターフェイスでの作業
- ・ジョブエディタでの作業
- ・新しい変換での作業

[コンテンツ]

- SAS Data Integration Studioの基本機能の探索
 - ・プラットフォームの探索
 - ・SAS Data Integration Studioの基本的な変更点
 - ・変更管理の基本的な変更点
- ジョブエディタの機能
 - ・一般的な変更点
 - ・ジョブエディタでの作業
 - ・SQL結合変換での作業
 - ・ジョブの連鎖

- ・ジョブ・パフォーマンスの統計量
- ・ジョブに対するメタデータ・レポート

■ SASコード変換

- ・SASコード変換での作業
- ・新しい変換ウィザードの利用

■ マッピングとプロパゲート機能

- ・新しいマッピング機能の探索
- ・新しいプロパゲート機能の探索

■ データクオリティの拡張点

- ・DataFlux® Integration Server イントロダクション
- ・DataFlux ISジョブ変換での作業
- ・DataFlux ISサービス変換での作業

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
Linux(Intel) 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>

発行: SAS Institute Japan株式会社



SUMMER 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



THE
POWER
TO KNOW.

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©2010, SAS Institute Inc. All rights reserved.