STN
SAS Technical News

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

WINTER 2013

特集 …… 01
SAS® Add-in for Microsoft Office 5.1のご紹介

SAS Academic News …… 09
- 随想「マーケティングとデータ解析」
- コラム「SAS／JMPとの歩み」

Q&A …… 17
最新リリース情報 …… 20

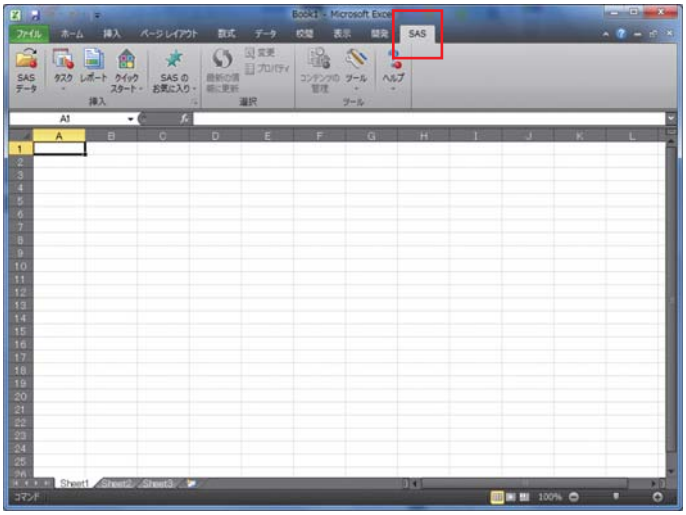
SAS® Add-in for Microsoft Office 5.1

2012年2月にSAS® Add-in for Microsoft Office 5.1がリリースされました。SAS® Add-in for Microsoft Officeとは、Microsoft OfficeにSASの機能を追加するアドイン製品です。特にSASの知識をお持ちでないビジネスユーザーの方でも、慣れ親しんだOfficeのインターフェースを介してSASの強力な分析・レポート機能を利用することができるようになります。従来よりも大規模なデータの分析や迅速なレポーティング、高度なデータ連携が求められる場合はSAS® Add-in for Microsoft Office 5.1を活用いただくことをご検討ください。本特集ではSAS® Add-in for Microsoft Officeの基本機能である「データソースの閲覧・編集」「タスクによるデータ分析」「ストアドプロセスの実行」と、5.1の新機能である「クイックスタートツール」についてご紹介いたします。**1：SAS® Add-in for Microsoft Office 5.1 製品概要　2：SAS® Add-in for Microsoft Office 5.1 主要機能　3：SAS® Add-in for Microsoft Office 5.1 の機能拡張／新機能「クイックスタートツール」**

1 SAS® Add-in for Microsoft Office 5.1 製品概要

1.1 SAS® Add-in for Microsoft Officeとは

SAS® Add-in for Microsoft Office(以後、AMOと略)は、Microsoft OfficeにSASの機能を追加するアドイン製品です。アドインの対象はMicrosoft Word、Microsoft Excel、Microsoft PowerPoint、Microsoft Outlookです。アドインを追加するとOfficeのリボンに[SAS]というタブとメニューボタンが追加されます。これらメニューボタンからSASの拡張機能を呼び出すことができます。例えば、ExcelにAMOを追加した場合、次のようなタブが追加されます。



その他のOfficeでも同様に、[SAS]タブ配下のメニューボタンからSASの機能を呼び出すことができます。

1.2 SAS® Add-in for Microsoft Office 5.1を使い始めるには

AMO5.1はSAS® 9.3 TS1M1以降で次のパッケージに含まれます。

- SAS® BI Server
- SAS® Enterprise BI Server
- SAS® Enterprise Miner
- SAS® Office Analytics

これらパッケージはサーバーマシンとクライアントマシンで構成されます。AMO5.1はクライアントマシンのMicrosoft Officeに対するアドインとしてインストールします。サポートされるOfficeのバージョンはMicrosoft Office 2007とMicrosoft Office2010(32bit版または、64bit版)です。また、サポートされるオペレーティング・システムは以下(32bit版または、64bit版)となります。

- Microsoft Windows XP Professional(ServicePack3を適用)
- Microsoft Windows Vista(ServicePack1を適用)
- Microsoft Widows Server 2003(ServicePack2を適用)
- Microsoft Widows Server 2008
- Microsoft Widows Server 2008 R2
- Microsoft Windows 7

AMO5.1の詳細なシステム必要条件は、次のページからご確認いただけます。

SAS® Add-in for Microsoft Office 5.1システム必要条件
http://www.sas.com/offices/asiapacific/japan/service/documentation/installcenter/msofficeint/5.1/sreq.pdf

SAS®
Add-in for
Microsoft Office 5.1

1.3 SAS® Add-in for Microsoft Office 5.1の製品構成

AMOは内部的にSASサーバーと連携して動作します。AMOで行われた操作はSASサーバー側で処理されます。サーバーのシステム資源を活用する構成であるため、ファイルサイズの大きいデータの加工や複雑なデータ分析など負荷が高い処理を行う事が可能になります。処理の実行結果はSASサーバーからOfficeにさまざまなデータ形式で返されます。
ただ、AMOの利用者はこのような内部的な動きは意識することなく、Officeの拡張メニューを使うような感覚でAMOを使うことができます。

2 SAS® Add-in for Microsoft Office 5.1 主要機能

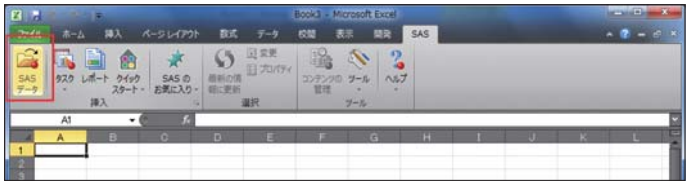
AMOは、大きく分けて次の3つの機能を提供します。本章ではこれらについてご説明します。

・データの参照と編集
・タスクによるデータ分析
・ストアドプロセスの実行

2.1 データの参照と編集

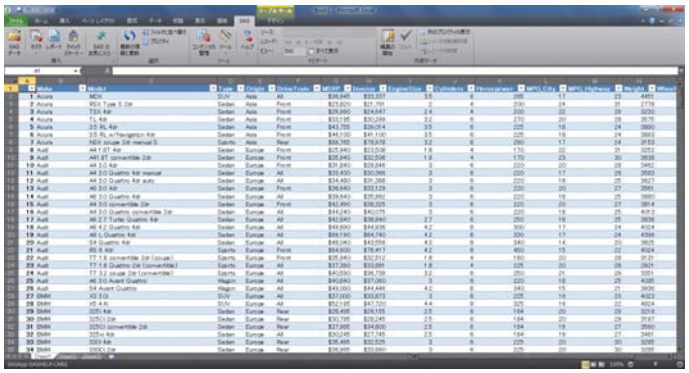
データの参照と編集はMicrosoft Excelでのみ利用可能な機能です。SASデータソースをExcel上のワークシート、もしくは、ピボットテーブルとして開いて扱うことを可能とします。SASデータソースとは、SASサーバーからアクセス可能なデータベースファイル一般を指します。具体的には、SASデータセット、Excelファイル、Accessデータベースファイル、SAS/ACCESSプロダクトから接続可能な各種リレーショナルデータベース（RDB）のテーブルなどが挙げられます。AMOを使うことで、これらデータベースファイルをあたかもExcelワークシートのように参照、編集することが可能となります。

データソースを開くには、[SAS]タブから[SASデータ]ボタンを選択します。



次に、SASデータの表示ウィンドウからデータソースや表示方法を選択します。[フィルタと並べ替え]による高度な絞込みや、表示レコード数の制御、また、ワークシートかピボットテーブルとして開くオプションを選択できます。

ワークシートとして開いた場合、次のような形でデータがExcelワークシートにロードされます。ロードされたデータは通常のExcelワークシートとして扱うことができます。当然、セルを編集する操作やExcelファイルとして保存する操作も可能です。



データソースをワークシートにロードした後、[SAS]タブの[編集の開始]ボタンを押すとデータソースの編集モードに入ります。

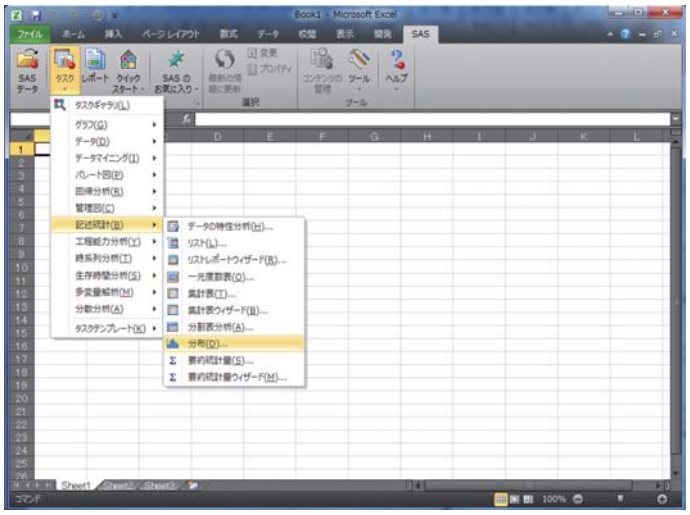
編集モードでワークシートを編集すると、編集内容がロード元のデータソースに反映されます。つまり、SASプログラミングやデータベース問い合わせ言語に不慣れなユーザーでも、使い慣れたExcelのインターフェースでSASデータセットやRDBのテーブルを編集できるということです。

2.2 タスクによるデータ分析

AMOではタスクという機能を使ってデータ分析、レポート作成を行います。タスクとは、主にマウスのクリックやドラッグ&ドロップで操作可能なメニュー画面です。入力データと処理対象の列を選択した後、任意でパラメータやオプションを指定することで、多彩な分析レポートをExcel上に出力することができます。AMO5.1では、次の通り、合計80のタスクが用意されています。SASプログラミングの知識がないユーザーでもこれらタスクを使うことで、あらゆる規模・形式のデータを自由に分析・レポート作成をすることを可能とします。

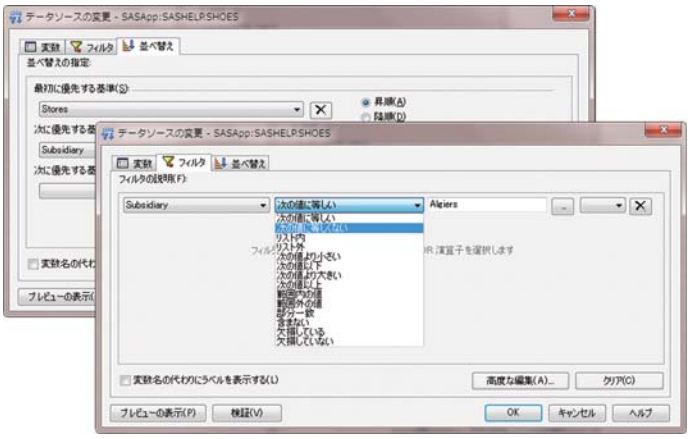
グラフ		
自動作成チャート	ODS統計グラフの表示	タイルグラフ
ドーナツチャート	バブルプロット	マップチャート
レーダーチャート	円グラフ	円グラフウィザード
曲面プロット	散布図	散布図行列
折れ線グラフ	折れ線グラフウィザード	等高線プロット
箱ひげ図	棒グラフ	棒グラフウィザード
棒-折れ線グラフ	領域プロット	
データ		
クイック統計量	SASServerへのコピー	データセットの属性
データの比較	データの標準化	データの並べ替え
ランク	転置	無作為抽出
列の積み上げ	列の分割	
パレート図		
パレート図		
回帰分析		
ロジスティック回帰分析	一般化線形モデル	線形回帰分析
非線形回帰分析		
管理図		
cチャート	npチャート	pチャート
uチャート	個々の測定チャート	箱ひげ図
平均と範囲のチャート	平均と標準偏差のチャート	
記述統計		
データの特性分析	リスト	リストレポートウィザード
一元度数表	集計表	集計表ウィザード
分割表分析	分布	要約統計量
要約統計量ウィザード		
工程能力分析		
CDFプロット	PPプロット	QQプロット
ヒストグラム	確率プロット	
時系列分析		
ARIMAモデリングと時系列予測	パネルデータの回帰分析	基本的な時系列予測
時系列データの加工	時系列データの作成	自己回帰誤差付き回帰分析
多変量解析		
クラスター分析	因子分析	主成分分析
正準相関分析	相関分析	判別分析
分散分析		
t検定	ノンパラメトリックな一元配置分散分析	
一元配置分散分析	混合モデル	線形モデル
生存時間分析		
ノンパラメトリック法による生命表分析		

今回は例として、分布タスクをご紹介します。分布タスクを使用すると、数値変数の分布や要約情報を格納した表を作成することができます。加えて、ヒストグラム、確率プロット、QQプロット、箱ひげ図などのプロットを作成することができます。タスクを実行するには、[SAS]タブ配下の[タスク]ボタンを押します。分布タスクの場合は[SAS]→[タスク]→[集計表]→[分布]を選択します。

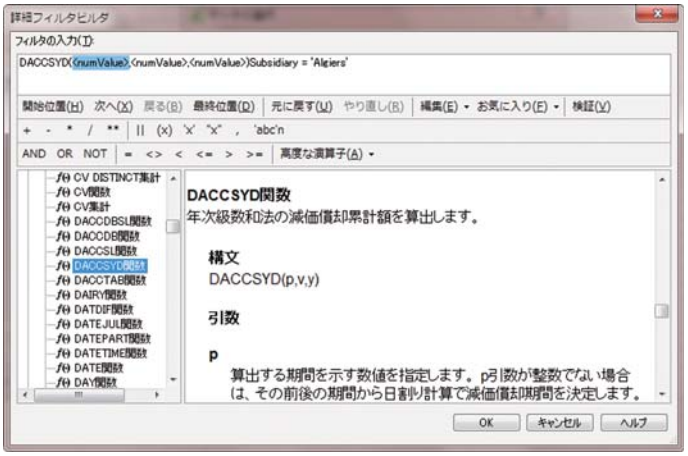


多くのタスクでは、初めに[データの選択]画面が表示されます。ここでは入力データソースと結果レポートの出力場所を指定します。以下の例では、データソースにサンプルデータセットSASHELP.CARSを選択し、出力先範囲を既存ワークシートの「列A行1」としています。

上の画面で[フィルタと並べ替え]ボタンを押すと、[データソースの変更]画面が表示されます。この画面では入力データソースの行数をフィルタで絞り込みしたり、列で並べ替えることができます。

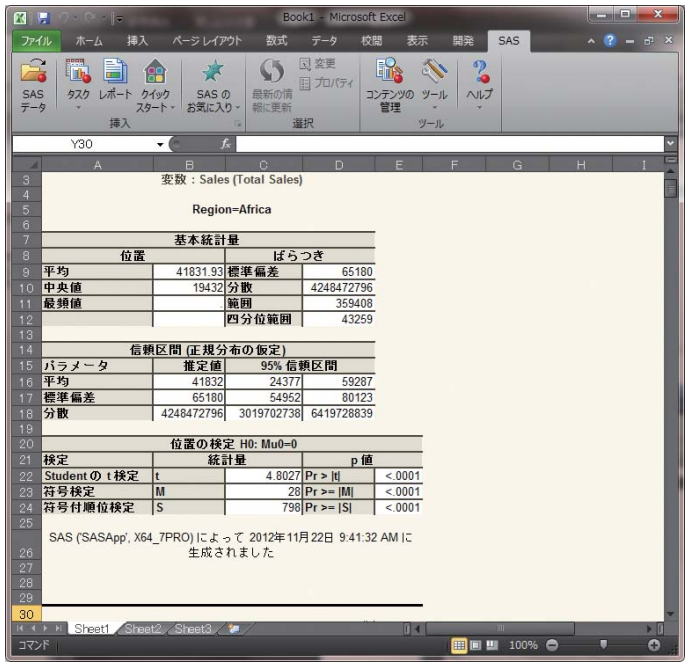


より複雑なデータの絞込みを行う必要がある場合は、[高度な編集]ボタンを押して[詳細フィルタビルダ]ウィンドウを開きます。ここでは論理演算子やSAS関数／SQL集計関数を使った絞込みの条件式を手動で指定できます。利用可能なSAS関数／SQL集計関数の一覧は左下のパネルに表示されます。また、右下のパネルに各関数の構文と使い方に関する説明が日本語で表示されます。



入力データソースの指定が終わると、タスクのメインウィンドウが表示されます。この画面では、データソースの変数に対して役割を割り当てます。例えば今回の場合、地域ごとの売上額の分布を確認するには[分析変数]に“TotalSales”、[グループ変数]に“Region”を指定します。

実行ボタンを押すと、SASサーバー側で処理が行われ、結果レポートがワークシートに出力されます。



AMOによる分析結果はSASレポート、HTML、もしくはCSVデータ形式でExcelのワークシートに出力することができます。AMOがインストールされている環境では、[SAS]→[最新の情報に更新]を押すことで、最新のデータソースにアクセスしてレポートを更新することができます。この機能を用いることで、一度作成したレポートを常に最新の状態に保つことができます。

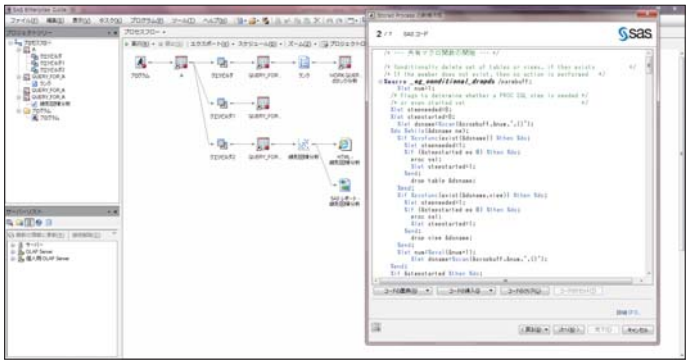
また、レポートを右クリックし、[プロパティ]→[表示]を選択することで、レポートの表示スタイルを変更できます。[最新の情報に更新]に

チェックを入れた状態でスタイルを変更し、[OK]ボタンを押すことで瞬時にスタイル変更を反映させることができます。

2.3 ストアドプロセスの実行

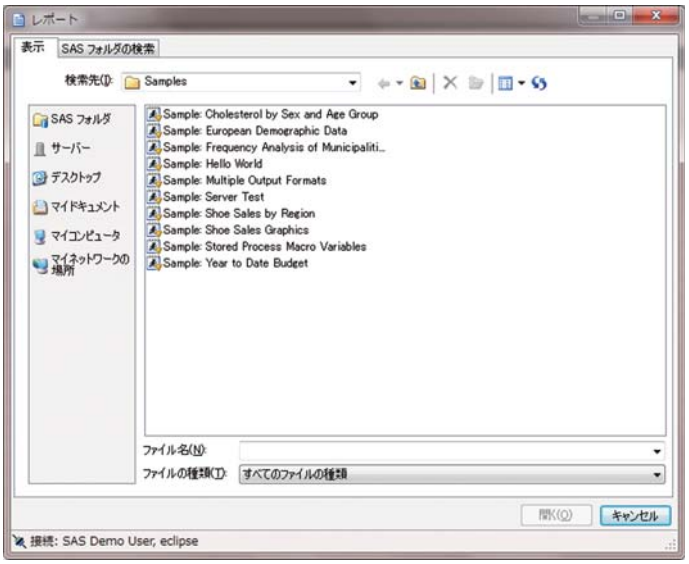
ストアドプロセスとは、SAS® Enterprise Guide® やSAS® Data Integration Studioを含むアプリケーションで作成し、SASサーバー上で管理され、クライアント・アプリケーションからの要求に応じて実行されるSASプログラムを指します。これらアプリケーションは、複数のデータ加工処理や分析処理を接続させた複雑な処理を作成することができます。SASプログラムをサーバー上で一元管理することにより、プログラムの保守性・一貫性を高めます。

例えばEnterprise Guide 5.1の場合、次のような画面でストアドプロセスを作成します。

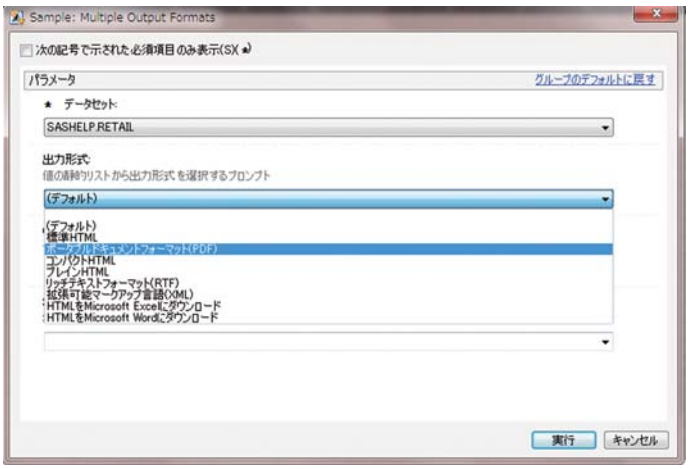


AMOでは、上述のアプリケーションで開発されたストアドプロセスを実行し、レポートを作成することができます。ストアドプロセスを実行するには、[SAS]→[レポート]を選択します。

[レポート]ボタンを押すと、SASサーバー上で管理されているストアドプロセスの一覧が表示されます。



ストアドプロセスに含まれるSASプログラムコードそのものは高度で複雑でも、それを実行するユーザーは特別そのことを意識する必要はありません。ストアドプロセス実行に必要なパラメータを入力して[OK]ボタンを押すことで、結果を得ることができます。例えば上のリストにあるサンプルストアドプロセス「Sample:Multiple Output Formats」を選択した場合、次のような入力画面が表示されます。



このような入力項目をプロンプトと呼びます。プロンプトで入力したパラメータによってストアドプロセスの実行結果が変化します。今回の例では、ストアドプロセスの処理対象データセットと出力データ形式を実行ユーザーが選べる形となっています。

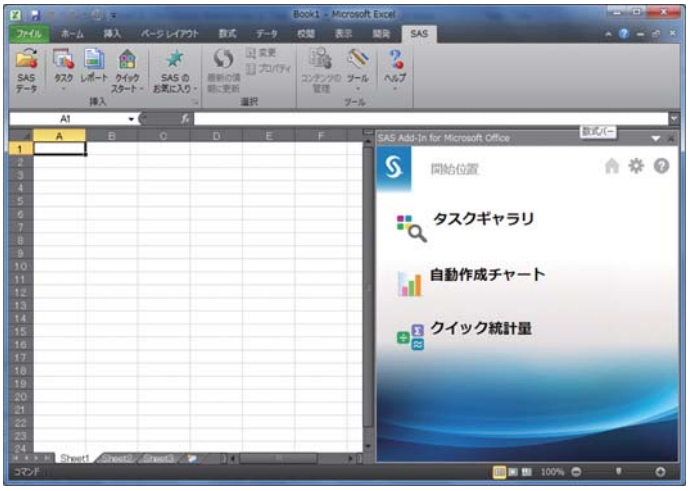
AMOのタスクでは実現できない分析では実現できない分析も、事前に処理の雛形が定義されているストアドプロセスを呼び出すことで、簡単な操作で実現することができます。

3 SAS® Add-in for Microsoft Office 5.1の機能拡張

AMO5.1では次の機能が拡張されました。

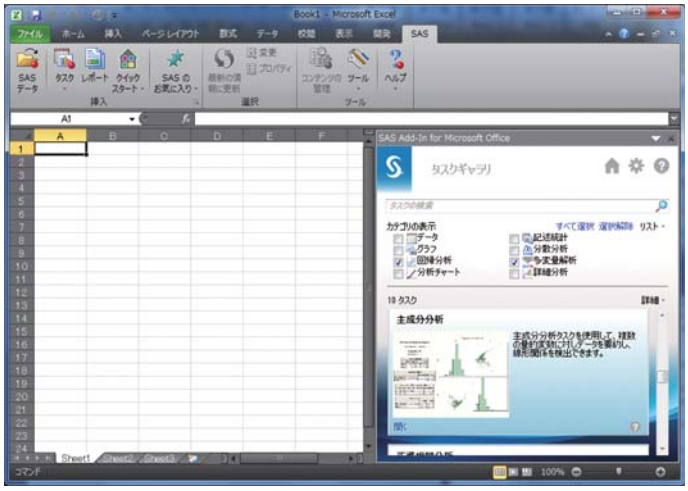
・クイックスタートツールの実装
・64-bit版Microsoft Officeのサポート
・Microsoft Office Communicatorに統合
・オートメーションインターフェイス(VBAのAPI)の拡張
・Microsoft Excelチャートを作成する機能を追加
・グラフの新しいデフォルト出力形式を提供
・ODS Graphics Editorで作成したグラフの表示

特に大きな拡張点として「クイックスタートツールの実装」が挙げられます。クイックスタートツールとは、「タスクギャラリ」「自動作成チャート」「クイック統計量」という三つの便利機能を包括した新しいインターフェースです。このクイックスタートツールを使用することにより、従来のAMOで作成するレポートより、データの特性を容易に把握可能とし、また、目的の出力結果に辿り着けるため、タイムリーに求められる定型レポートをより簡単に作成できます。
クイックスタートツールを起動するには、[SAS]→[クイックスタートツール]を選択します。以下のように大きなパネルが画面の右半分に表示されます。
本章ではこのクイックスタートツールに含まれる三つの便利機能についてご紹介します。

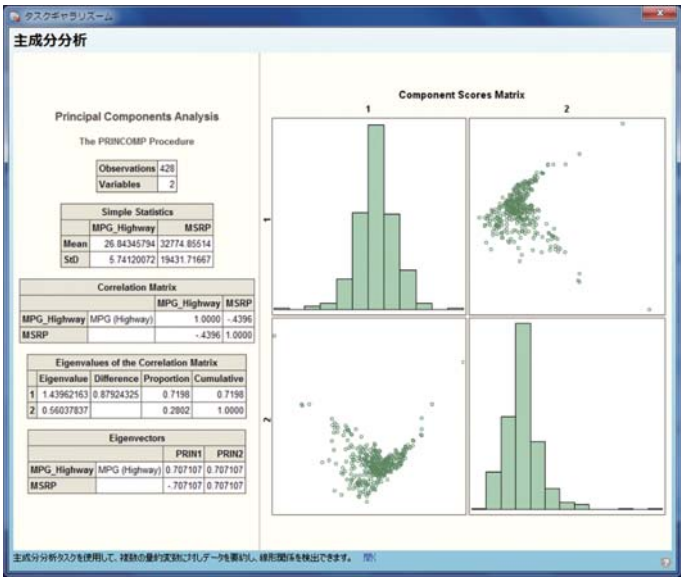


3.1 タスクギャラリ

タスクギャラリを選択すると、タスクの一覧が機能カテゴリ別に表示されます。タスクビューボタンを操作することで幾つかの表示方法が選択可能です。例えば、表示を「詳細」に設定すると、次の画面のようにタスク名、実行結果のスクリーンショット、タスクの簡単な説明が一覧で表示されます。



タスク名の下にある「開く」ボタンを押すとタスクの画面が開き、分析を開始できます。一方、実行結果のスクリーンショットをクリックすると、次のように実行結果のスクリーンショットが拡大で表示されます。

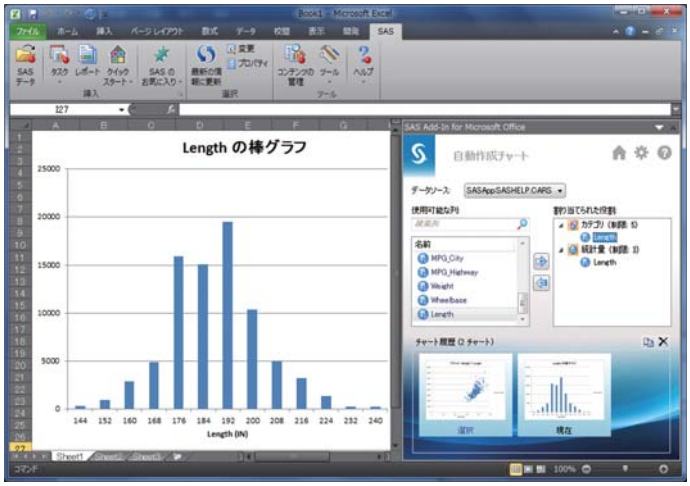


タスクギャラリを使うことで、AMOのタスクに不慣れな方でも、各タスクの用途や実行結果の概要が視覚的に表示されることにより、AMOのメニューから目的のタスクを探す手間がなく、処理したい分析や得たいグラフによりスピーディにアクセスして分析作業を進めることができます。

3.2 自動作成チャート

自動作成を使うと、利用者が役割変数に割り当てた変数に応じて最適なチャートが自動作成されます。役割変数は「カテゴリ」と「統計量」があります。チャート作成にあたって、少なくとも、1つの列をカテゴリ役割と統計量役割にそれぞれ割り当てるか、2つの列を統計量役割に割り当てる必要があります。割り当てが完了すると、最も適した形と組み合わせのチャートが自動的に作成され、ワークシートに挿入されます。一度作成したチャートはパネル下部の「チャート履歴」に表示され、画像をクリックすることで再表示できます。
自動作成チャートを用いることで、未知の組合せに対しても最適な

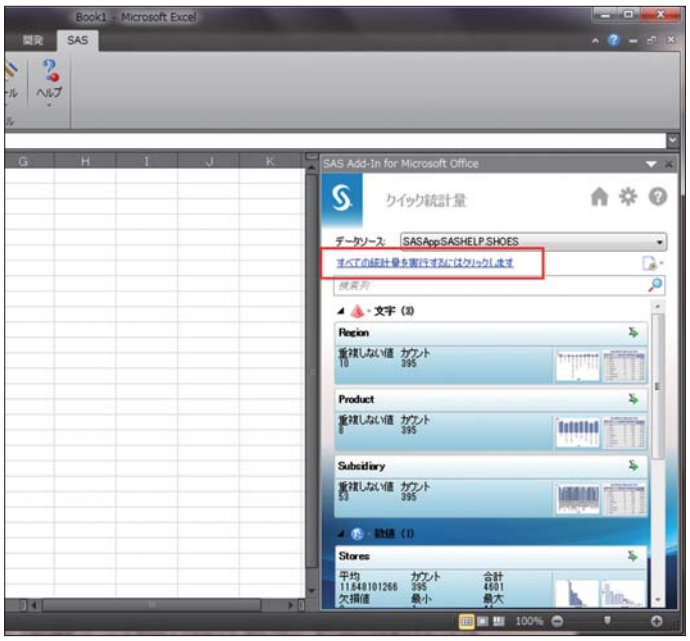
チャートが作成されるため、予想していなかったデータの有益な特性も得ることができます。また、通常のチャート作成タスクは、設定項目を詳細に指定することで、ようやく目的のチャートを作成することができますが、自動生成チャートは固定の画面からさまざまな列の組合せを素早く指定できるため、逐一個別のグラフタスクを起動して作業するよりも効率的です。



3.3 クイック統計量

クイック統計量パネルを使うと、データソースに含まれる各列について基本統計量と基本グラフを素早く確認することができます。変数は文字、数値、日時、通貨という4つのグループに分類され、それぞれのグループに適した基本統計量と基本グラフが表示されます。

すべての変数に対する基本統計量を求めたい場合は、[全ての統計量を実行するにはクリックします]リンクをクリックします。



また列名が部分一致する列のみを表示するには、上部の検索フォームに文字列を入力します。

クイック統計量を使用することで自動作成チャートと同様に、通常のタスクによる詳細な設定を行うことなく、AMO がデータを判断して基本統計量を算出するため余計な手間を省くことが可能です。また、設定値を簡易的に変更することができるため、データ別によるシミュレーションを簡易的に素早く行うことが可能となり、データに適した分析手法を判断できます。

4 おわりに

駆け足での説明ではございますが、SAS® Add-in for Microsoft Office 5.1 の紹介をいたしました。是非とも最新版のご利用の参考にしていただければと存じます。

SAS アカデミック・ニュース
SAS Academic News

第3回、朝野先生による「マーケティングとデータ解析」では、統計プログラムとどのように付きあうか、実際にSAS® Enterprise Guide® 5.1を使用して直感的にGUIを操作し統計解析を行うという内容です。また新村先生のコラムでは、よりシステムインテグレータであり、且つ分析者としての確固たる実績を残していく様子が書かれています。
なお、長い間掲載しておりました「SASアカデミック・ニュース」は、次号より弊社マーケティング部門から発刊するAcademic News(タイトル仮)に連載が移行することになりました。長い間ご愛読いただきまして誠にありがとうございました。

随想
「マーケティングとデータ解析」

コラム
「SAS／JMPとの歩み」

随想 「マーケティングとデータ解析」

朝野 熙彦
多摩大学大学院客員教授

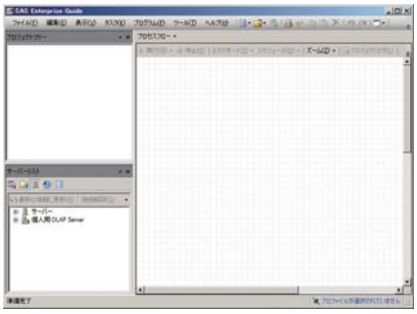
第３回 統計プログラムとのおつきあい

因子分析はマーケティングの分野でポピュラーに利用されている分析法です。そこで今回は因子分析を一例にして統計プログラムとのかかわりについて随想を述べたいと思います。データ解析者として統計プログラムとどうつきあってきたか、という話しです。

1. 初めてのEG

私はプログラミングが苦手なので、メニューをクリックするだけでデータ解析ができてしまうソフトは重宝な道具です。そこで最近話題のSAS Enterprise Guide 5.1（以下EGと略称します）が、本当にGUIに優れたソフトなのかどうかを試してみましょう。

自分自身のPC操作がおぼつかないのですが、EGを起動すると次のような画面が出てきました。以下、ともかくマニュアルを何も見ずに操作してみたいと思います。



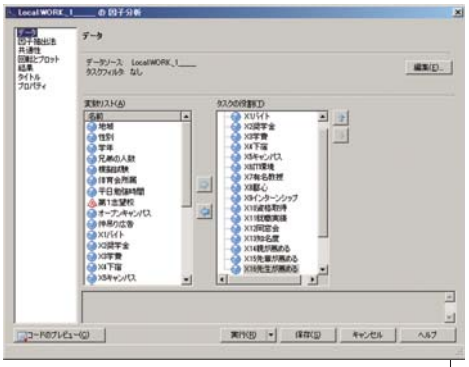
EGの基本画面

プロセスフローという真っ白い画面が出てくるので一瞬途方にくれますが、たぶん最初にデータを入力するのだろうと見当をつけて「ファイル」をクリックしますと「データのインポート」というメニューがありました。それを選んでエクセルのデータをインポートすることにしました。

今回の分析データは、2006年8月に私のゼミ生がインターネットで調査した「大学受験生へのアンケート」です。高校生500人が回答した調査で、第1志望校に選んだ大学を16個の変数についてそれぞれ、「5.とてもそう思う、4.そう思う、3.どちらでもない、2.そう思わない、1.全くそう思わない」の5段階尺度で評定してもらったデータから出来ています。

この調査は、18歳人口が減少するなかで定員割れの大学が出てきたという時代背景から、志願者の獲得がこれからの大学の重要なマーケティング課題になってくるだろう、という問題意識から企画されました。そこで大学のポテンシャルユーザーである大学受験生の志望理由を知って対策を検討しよう、という意図から自主調査したものです。

さて500行16列のデータ行列を変数名付きでSASに読み込みました。日本語の変数名がそのまま分析に利用できるのには驚きました。変数名が英数字でなければならないとしたら、二重手間の入力作業が必要になります。次に分析⇒多変量解析とメニューをたどると因子分析が見つかりました。



因子分析のメニュー画面

変数リストのウィンドウから、右のタスクの役割のウィンドウに16個の分析変数を移すのはごく自然な流れだといえましょう。さて次は何をすればよいのか迷いましたが、メニュー画面の左の白い枠の中に、因子抽出法などのメニューが並んでいるのが見つかりましたので、順に一つずつクリックして、オプションを選んで【実行】ボタンを押しますと、因子分析が実行できて一段落になりました。

因子の抽出法では主成分解、共通性の初期値としてはSMCを選び、因子の数は固有値が1以上で打ち切り、直交バリマックス回転、因子得点を出力するようにチェックしました（註）。起動してからここまでの所要時間が30分でした。慣れれば5分でできる作業でしょう。

初めてのEGはなかなか快適で、操作の流れが人間の思考のプロセスにそって出来ていることに感心しました。ただし、自分が選んだ因子の抽出法はprincipal factor methodのはずなので、主成分解

ではなく主因子法と書くのが正しい。因子分析についてわが国で定評のある専門書を確かめますと、

1) 柳井 晴夫他「因子分析ーその理論と方法」朝倉書店の 51-53 頁
2) 芝 祐順「因子分析法」東京大学出版社の第 3 章（80-115 頁）
3) 芝・渡部・石塚編著「統計用語辞典」新曜社、107 頁

1)～3) いずれも因子と分析変数の共分散（因子負荷量）の平方和を最大にする因子抽出法を主因子法だと明記しています。芝先生、柳井先生はこの分野の日本の権威です。また因子分析の古典の1つであるHarman,H.H."Modern Factor Analysis, Second Edition."The University of Chicago Press.1968,pp.137-143、においてもPrincipal- Factor Methodとprincipal component analysis（主成分分析）が異なることが書かれています。

因子も主成分も同じだ、というのは雑な翻訳だと思います。

2. 受験生が気にする志望校の条件は

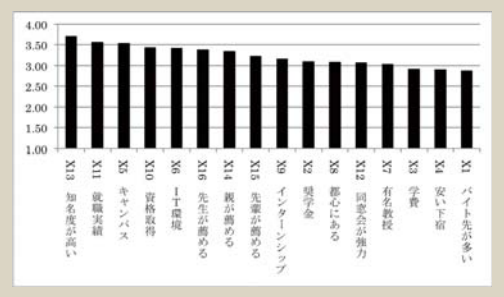


図1：5段階尺度の平均値

EGでは入力データの平均と標準偏差、および相関行列にもチェックを入れておきました。そういう基礎的な情報も因子分析と一緒に出力されます。500人の受験生の第1志望校の該当度を平均値が大きい順に並べたのが図1です。有名校で就職実績がよいこと、キャンパスがきれいなことが上位に来ています。いかにも実利本位の理由です。なお私のゼミ生は「教育熱心な教授がいる」とか「4年間しっかり学べる」、というような変数を始めから質問項目に入れていませんでした。残念！

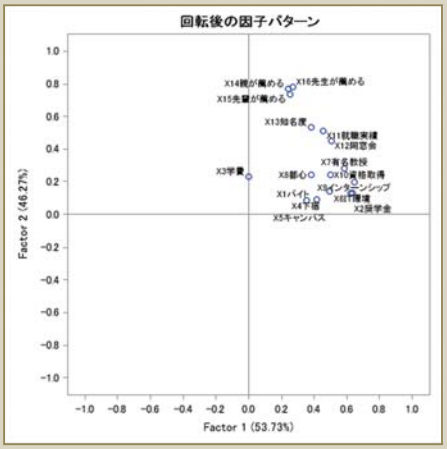


図2：因子分析の結果

	共通性	SMC
X16 先生が薦める	0.684	0.642
X14 親が薦める	0.646	0.604
X15 先輩が薦める	0.604	0.582
X11 就職実績	0.472	0.535
X12 同窓会が強力	0.457	0.466
X9 インターンシップ	0.454	0.418
X13 知名度が高い	0.428	0.494
X7 有名教授	0.421	0.446
X6 ＩＴ環境	0.417	0.475
X2 奨学金	0.407	0.403
X10 資格取得	0.306	0.323
X5 キャンパス	0.266	0.392
X8 都心にある	0.206	0.348
X1 バイト先が多い	0.182	0.269
X4 安い下宿	0.131	0.267
X3 学費	0.053	0.137

表1：共通性とSMC

さて16の変数には相関がありますので、その背後には少数の潜在変数が働いているのかもしれません。その潜在変数を因子といいます。分析変数の因子負荷量を2因子の空間にプロットしたのが図2です。因子負荷量というのは、私が選んだオプションの場合ですと、分析変数と因子との相関係数を意味します。

第1因子は有利な制度に関する因子で、第2因子は他者の推奨の因子のように理解できます。一般の消費財でいう品質因子と口コミ因子に相当します。さて2次元空間を見ると先生の推奨がこの空間の原点から遠くに位置し、学費が原点に近く位置しています。これが何を意味するかといいますと、それは各分析変数が因子空間によって説明できる程度を表しています。因子空間で説明できる程度は「共通性」という指標で表わされます。共通性は0以上1以下の値をとります。この表でSMCというのは、この共通性を推定するための初期値であって、最終的な共通性とは一致しませんがそれで構いません（表1）。

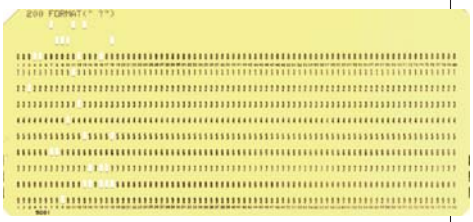
EGを使ってみて凄いと思ったのは、以上の出力結果がそのままWindows上でコピー＆ペーストできることです。特に表1のような数表をダイレクトにエクセルに貼り付けて大きい順に並び替えができるのはとても便利だと思いました。

3. 統計プログラムとの出会い

前節で見てきたように、今日は生まれて初めてさわったソフトでもいきなり使えてしまう時代になりました。一昔し前ならまさに驚天動地のはずです。時代をさかのぼってみましょう。

大型コンピュータを使っていた時代

個人的な回想をさせてください。学生だった1960年代に私は授業で初めて汎用コンピュータに出会いました。電子計算機実習という科目でした。その時、実習問題に出たのが分散と相関係数を計算する課題でした。利用したのはFORTRANというプログラム言語で、データとプログラムをそれぞれ次のような80桁のIBMカードに穿孔してコンピュータに読み込んでバッチで処理するシステムでした。カードに穴を空けることをパンチと呼びます。



当時のパンチカード

当時のプログラムの入力には、プログラムを一行ごとにこのパンチカード専用機でキーボードから打ち込んでいました。1000行からなるプログラムでは、パンチカードを1000枚、打ち込みが必要な時代でした。

さて、分散と相関係数はわずか数十行で済む簡単なプログラミングなのですが、完成まで何日もかかりました。クーラーの効いた電子計算機室に自分がパンチしたカードを持ちこみますが、直ちに自分の番は回ってきません。自分のジョブの実行までぼんやりと順番を待ち、そのあげくエラーメッセージが出て、またプログラムを修正するためにパンチルームに引き返すという日々でした。

自分には根本的に注意深さが欠けているのでプログラミングには向いていないことを悟った授業でした。

理論面については因子分析の授業があって、そこではセントロイド法という今では消えてしまった歴史的計算法を習いました。この解法は重心法とも呼ばれ因子負荷量を手計算するための方法です。因子負荷量の絶対値の和を最大化することを狙ったアルゴリズムでしたが、数学的にその最大化が保証されないという欠点がありました。その後、コンピュータで固有値・固有ベクトルを求めることが可能になったので利用者がいなくなったのです。

セントロイド法でも紙と鉛筆で計算するのは大変なので、「手回し計算機」を使って解いていたのです。私の先輩にあたる草創期の方々は何カ月も手回し計算機を回し続けて因子分析をしていたそうで、その努力には頭が下がります。EGなら1秒もかからない計算だったでしょう。

外部のデータセンターに通った時代

随想第1回で自己紹介しましたように、大学卒業後はマーケティング・リサーチの会社に就職してデータ解析まで自分でしなければならない羽目になりました。自分でプログラミングする能力はなく、また小さな会社ですから当然、社内には電子計算機はありません。そこで外部のデータセンターにデータを持ち込んで計算を依頼することになりました。1970年代はマーケティング・リサーチの世界で多変量解析が普及してきた時代です。特に因子分析とクラスター分析に人気がありました。

そのせいか、統計パッケージなるものが当時はコンピュータのメーカー単位で開発されていました。その後、SASをはじめ有力な汎用統計ソフトに収斂していくことになります。しかしいずれにしても大型コンピュータが設置してあるところにデータを持参してコンピュータの操作員に計算を依頼する、という意味では学生時代と同じことを会社でも繰り返していたことになります。依頼が済んだらただぼんやりと時間待ちをしていました。自分が働いている時間よりも待合室でぼんやりしていた時間の方が長かったように思います。

ダウンサイジングの時代

その後メインフレームから次第にデスクトップPCへと計算環境が変わり、もともと大型コンピュータ用に作られたプログラムが、かつての残滓を引きずりながらPC用のソフトに移植されていったことは皆さんもご存じでしょう。残滓の一例として、メモリの制約のために共用できるアルゴリズムを中心に分析プログラムを組み込んでしまったという弊害がありました。因子分析と主成分分析は分析モデルが異なるにもかかわらず、固有値を解くサブルーチンが共用できるという理由で、同じプログラム内で処理してしまうという大混乱のソフトもみられます。

さて、一人1台のPCの時代になり、またPCの演算速度が飛躍的にアップした結果何が起きたかというと、「ぼんやりした待ち時間」が無くなってしまいました。データ解析の生産性が上がってどんどん仕事ができるようになった、というプラス面はあります。しかし一方でぼんやりする時間がなくなったというマイナス面も出てきました。

人間はぼーっとしている時間に自分のデータ解析のミスに気づいたり、マーケティングの問題解決がひらめくことがあります。傍目からは無駄にみえる待ち時間にも創造的な意義があるのではないかと思います。

4. 統計プログラムとのおつきあい

その後、某大学に移ってから情報処理教育を副科目で担当させられました。役目ですので、学生にはこれからの社会では情報処理が重要だとか、プログラミングも勉強したほうがいいなどと自分のことは棚にあげて指導しました。そうした情報処理教育にも意義はありましょうが、データ解析はPCやソフトウェアの操作を習得すれば済むものではないと思います。

ユーザーとプログラムのつきあい方は、大きくは次の3つのパターンに分けられるでしょう。

① 自分でCやBASICのような言語を使ってプログラミング
② エンドユーザー言語を使ってプログラミング
③ 完成品のパッケージプログラムを使う

①は自分でゼロからコーディングするやり方です。今日のようにデータ解析の環境が豊かになっている時代では、一般のユーザーが①を選ぶのはあまりにも迂遠ではないかと思います。

②はSAS/IMLのようなエンドユーザー言語を使ってコーディングする方法です。大きなサブルーチンが1つのコマンドに収まっているので、プログラミングが簡潔になります。たとえばFORTRANで数百行を要した固有ベクトルの計算がeigvec( )のわずか一語で済みます。とても素晴らしいと歓迎する反面、学生時代にその数百行をパンチした自分としては内心忸怩たるものがあります。

③はたとえばSASを使うというつきあい方です。各分析に対応したプロシジャが用意されています。それでもSASの文法を覚えてコマンドを入力すること自体が面倒だという人はEGを使えばよいのです。EGとSASの関係は、前者が後者のユーザーインターフェースだと理解すればよいでしょう。つまりEGのバックグラウンドでは普通どおりにSASが動いていて、SASのプロシジャが何であるかを知りたければそれを見ることもできるのです。

私のようにプログラムが苦手であれば③のつきあい方が一番よいでしょう。できるだけ楽にデータ解析をすることは望ましいことだと思います。しかし、そうは言っても一つ注意があります。それは、統計プログラムを安易に信用してはならない、ということです。

世の中には多数の統計プログラムが存在します。SASのように信頼できるソフトばかりではなく、中には計算や出力のロジックに誤りがあるソフトもあります。また不適切なデフォルトを設定しているためにユーザーの誤用を促進しているソフトもあります。同じ名前の分析法ならばどれも同じだろう、と信じてはならないのです。つまりソフトの真贋を見分け、統計プログラムを正しく使いこなすには、多少は理論的な理解も必要になるのです。これはPC操作に習熟するだけの情報処理教育とは異なる、より本質的な問題です。

（註）
マーケティング・リサーチの実務で発生する様々なトラブルについて対策を示した本をまとめました。今回述べた因子分析と主成分分析の概念の違いについても詳述しています。
朝野熙彦「マーケティング・リサーチ～プロになるための7つのヒント」講談社、2012年11月刊
この本については、下記の「新刊書籍のご案内」をご参照ください。

略歴：
千葉大文理学部卒業後市場調査会社に就職、埼玉大大学院修了、千葉大・筑波大講師、専修大・都立大・首都大教授を経て多摩大学大学院客員教授。学習院マネジメントスクール講師、日本マーケティング・サイエンス学会論文誌編集委員、日本行動計量学会理事。
主な著書に『アンケート調査入門』東京図書（編著）、『最新マーケティング・サイエンスの基礎』講談社、『Rによるマーケティング・シミュレーション』同友館、『入門共分散構造分析の実際』講談社、『魅力工学の実践』海文堂出版、『入門多変量解析の実際第2版』講談社、『新製品開発』朝倉書店（朝野熙彦・山中正彦著）などがある。

新刊書籍のご案内

新刊書籍をご紹介します。

マーケティング・リサーチ　プロになるための7つのヒント

●著者：朝野熙彦　●出版社：講談社　●サイズ：A5判　●ページ数：191ページ
●価格：¥2,940（税込）　●発売日：2012年11月20日　●ISBN：978-4-06-157303-1
●URL：http://www.bookclub.kodansha.co.jp/bc2_bc/search_view.jsp?b=1573039&x=B

今のやりかたで大丈夫？
序説＋7つの章に散りばめられた、数えきれないヒント、知恵、ときどき裏ワザ
こうすればできる、こう考えれば間違わない。

主な内容
目次より
●第1章　マーケティング・リサーチ序説
●第2章　消費者を理解する
●第3章　コンセプトを作成する
●第4章　製品をテストする
●第5章　価格戦略を決める
●第6章　稀少セグメントを発見する
●第7章　時代のトレンドをとらえる
●第8章　社内ノルムをアップデートする

まえがきより
リサーチとは枯れた技術の寄せ集めではありません。
近年産業界で関心を集めているインサイト、ビジネス・エスノグラフィー、ビッグデータ、そしてベイズ統計などもマーケティング・リサーチとかかわりがあります。そうした新しい概念や理論にも目配りしつつ、長年培ってきたマーケティング・リサーチの技法を今日的な視点で見直す、いわば棚卸のようなテキストが必要ではないかと考えて本書を上梓することにしました。

コラム「SAS／JMPとの歩み」

新村 秀一
成蹊大学 経済学部教授 理学博士

第4回 SASとJMPで一生の研究を完成 ー学位取得までー

前号のつづきです。

15. 判別分析の深い闇を切り開く

(1) 判別分析の抱える問題

1936年にFisherは、p個の説明変数の線形結合 $f(x)=a_0+a_1X_1+\cdots+a_pX_p$ から「**2群の分散比最大化基準**」でFisherの線形判別関数（LDF）の世界を切り開きました。その後さまざまな最適化基準で種々の判別関数が提案され、パターン認識、医学診断、経済の各種格付けや企業の倒産あるいは優良企業と不良企業の判別、最近ではゲノム判別へと応用されています。しかし判別分析は「**f(x)>0なら群1、f(x)<0なら群2と判別するという単純な判別規則**」に隠れて、大きな問題が未解決のまま放置されてきました。そして多くの統計家の中でも誤解と混乱を生んでいます。

1) f(x)＝0のケースをどちらに判別するかは未解決です。統計ソフトでSASだけがこの問題を意識して、f(x)＝0の扱いをユーザーが指定できます。一方JMPでは、ロジスティック回帰で2＊2の分割表に判別結果をまとめる際、どの群を陽性と指定するかを問い、判別境界上のケースを全て陽性に判別しているようです。

2)「2群が多次元正規分布し分散共分散が等しいといういわゆるFisherの仮説」を置けば、多次元正規分布の対数尤度からFisherが導出したLDFと同じものが求まります。この定式化があまりに見事なため、この仮説を満たさないデータにLDFを適用してはいけないという異論を唱える人も多くいます。しかし、「2群の分散共分散行列が等しいという仮説を満たさない」場合は、2次判別関数（QDF）を先人は開発しました。このことは、先人はFisherの仮説を満たさないデータがあることを知っていたことを示します。また近年医学や経済分野でロジスティック回帰が用いられています。判別結果がLDFやQDFに比べて良いことが多いためですが、判別成績がよいのは「Fisherの仮説」を考えず、データに合わせてオッズ比の対数尤度を求めているからです[50]。

3) 判別境界を動かせば得られた判別結果より良いものが得られることが多くあります。結果があまりにも悪い場合には「正規性からの乖離」と言ってあきらめてきました。また判別境界を動かしてより良い判別結果を求める研究をしている若手研究者も多くいます。すでに述べたように、判別分析で事前確率やリスクを導入し判別境界を自由に動かす手法が既に与えられています。そこでそれを評価するためにROC曲線の利用を提案しました。

4) 田口玄一先生は、正規分布に現れるマハラノビスの距離を用いて、正常状態の分散共分散行列から異常を考える1クラス判別分析を提案されています。私が考えた正常群と異常群の「地球モデル」と同じく、ある種の判別問題にFisherの2群判別の概念を持ち込む危うさを示しています。

以上の全ての点が、私が1998年から2010年まで行ってきた「最適線形判別関数[51]」の研究で解決しました。また新しく次の問題点が解決しました。詳しくは解説書[52]と論文などを参考にしてください。

5) 筆者の研究で初めて分かったことですが、既存のLDF、QDF、ソフトマージン最大化SVMは、**線形分離可能なデータを認識できない**ことです。また、変数選択法の統計量はMNM=0の最小次元を基準にして考えると一定の傾向を示しません。ロジスティック回帰は、回帰係数の推定が不安定になり誤分類数が0になれば、線形分離可能なデータを認識したことが筆者の開発した最適線形判別関数（改定IP-OLDF）から分かります。しかし、最小次元のMNM=0の空間を必ず求める保証はありません。ハードマージン最大化SVMは、線形分離可能なデータを識別できますが、MNM=0以外のデータに適用できない問題点があります。

6) 判別関数の誤分類数と判別係数は推測統計学とは無縁であり、この関係が分かりませんでした。筆者の研究で、**判別係数は定数項が正と0と負の3つの異なった構造**をもち、MNMが最少な**最適凸体**の内点を判別係数とする最適線形判別関数を考えれば判別分析の抱える問題が解決できることが分かりました。

7) LDFやQDFは、試験の合否判定という自明な線形分離可能なデータを認識できず、LDFでは誤分類確率が0.34、QDFでは0.9以上になることが分かりました。

(2) 2次元における3個のケースの判別例

最適線形判別関数の概略を、次の2個の説明変数をもつ3ケースで説明します。

群1；$x_1=(-1/18, -1/12)$,
群2；$x_2=(-1, 1/2)$,　$x_3=(1/9, -1/3)$

さらに記号y_iは群1の場合1とし、群2の場合−1とすることで、$y_i*f(x)>0$であれば正しく判別され、$y_i*f(x)<0$であれば誤判別されたとすることで不等号の向きを統一できます。

$$y_i*f(x)=ax'+1,\quad y_i*g_i(b)=x_i\,b'+1$$

判別係数aとxは交換可能であり、xの値を係数とする次の3つの線形式を考えます。

$$H1=y_1*g_1(b)=-(1/18)\times b_1-(1/12)\times b_2+1=0,$$
$$H2=y_2*g_2(b)=-b_1+(1/2)\times b_2+1=0,$$
$$H3=y_3*g_3(b)=(1/9)\times b_1-(1/3)\times b_2+1=0$$

これらの線形式は**図1**に示すとおり、判別係数を表す2次元平面$b=(b_1, b_2)$を7個の凸体に分割します。Hi上の点を判別係数b

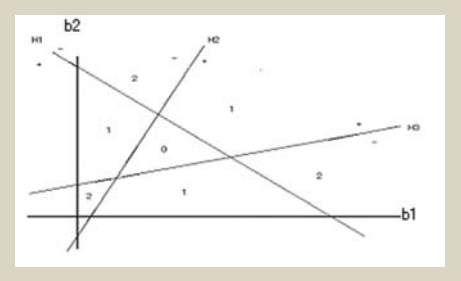


図1：判別係数の空間と誤分類数の関係／
3件のケースによる判別係数空間の最適凸多様体

とすると、ケースx_iが“判定保留群”に属し、その他のケースは$y_i*g_i(b)>0$であれば判別係数bを用いた$y_i*f(x)=bx'+1>0$はx_iを正しく判別し、$y_i*g_i(b)<0$であればx_iを誤判別します。すなわち、各凸体の内点で$y_i*g_i(b)<0$になる図に書き込んだ数が誤分類数で、しかも内点を判別係数に用いれば判別超平面上にケースがくることを避けられます。そして、最小の誤分類数をもつ凸体を**最適凸体**と呼ぶことにすると、MNM基準による最適線形判別関数はこの最適凸体の内点に対応した判別関数になります。

定数項を1に固定して考えましたが、一般に正の実数cに固定すると、図1の各軸との切片はc倍になり、相似変換されただけで本質的な構造は変わりません。またこのcが0に収束していけば切片も0に収束し、3個の超平面は原点を通る第2の構造になります。またcが負であれば、正の場合と異なったc＝－1と相似な構造になります。すなわち、判別係数と誤分類数は**3つの異なった構造で説明できます**。

(3) IP-OLDFと改定IP-OLDFと改定IPLP-OLDFの定式化

上の例題を一般化すると、次の整数計画法を用いたIP-OLDFになります。y_iは群1の場合は1で群2の場合は-1、x_iは説明変数のデータ、bは判別係数、e_iはx_iが正しく判別される場合は0に誤分類される場合は1に対応する1/0の整数値です。誤分類されるケースの制約式は、$y_i\times(x_i'b+1)>=0$から$y_i\times(x_i'b+1)>=-1$に変わることで制約式が満たされます。目的関数はこのe_iが1になる誤分類数の和を最小化します。ただしこのモデルはデータで表わされる配置行列が一般位置にある場合は最適凸体の頂点を正しく求めますが、一般位置にない場合は正しい最適凸体の頂点を求めないこともあることが分かりました。一方、e_iを非負の実数にしたものをLP-OLDFといいます。数理計画法の判別分析研究で行われてきたL1ノルム判別分析の一種です[52]。ただし、誤分類されるケースの判別超平面上からの距離の和を最小化しても、現実的に意味がないのでこの種の研究成果は利用されてきませんでした。

$$\text{MIN} = \Sigma e_i;$$
$$y_i\times(x_i'b+1)>=-e_i;\ i=1,\ldots,n \quad \cdots\cdots ①$$

改定IP-OLDFは、パターン認識で古くから考えられているマージン概念を取り入れて式②のように定式化します。マージンの反対側にきて誤分類される全てのケースx_iの制約式は、$y_i\times(x_i'b+1)>=1$から$y_i\times(x_i'b+1)>=-9999$に変わり、$y_i\times(x_i'b+1)=-9999$というSVに引き寄せられます。これで判別超平面上にケースは含まれなくなるので、正しい最適凸体の内点が得られます。

$$\text{MIN} = \Sigma e_i;$$
$$y_i\times(x_i'b+1)>=1-10000\times e_i; \quad \cdots\cdots ②$$

改定LP-OLDFは、式②のe_iを0/1の整数変数から非負の決定変数に変えただけです。

改定IP-OLDFは整数計画法を用いているので計算時間がかかります。そこで最初に改定LP-OLDFを適用し、e_i=0になる正しく判別されたケースを第2段階で0に固定します。そして第1段階で1と誤判別されたものだけに改定IP-OLDFを適用することで、MNMの近似値が高速に求められます。これによって**100重交差検証法の適用**が可能になりました。

(4) MNMの有用性

MNMに関して次の利点があります。

1) MNMの単調減少性：p変数モデルのMNMをMNM_pと表すことにします。このモデルに1変数追加した$(p+1)$変数モデルの$MNM_{(p+1)}$は、必ず減少します（$MNM_p \geqq MNM_{(p+1)}$）。この証明は簡単です。追加した説明変数の判別係数を0とすれば、その$(p+1)$変数モデルの誤分類数はMNM_pになります。よって、$(p+1)$変数モデルの$MNM_{(p+1)}$はMNM_pより小さいか等しくなります。これで、変数増加法で選ばれるモデルのMNMも単調に減少することになります。

MNMの単調減少性から「線形分離可能な最小次元のデータ空間」の発見ができます。すなわち、$MNM_p=0$であれば、このp変数を含む全ての判別モデルのMNMは0になります。

2) 正規性からの乖離を示す尺度：データがFisherの仮説を満たせば、得られた誤分類数はMNMに等しくなります。すなわち、LDFの誤分類数がMNMから大きいほど、データが正規性からの乖離していることが分かります。

3) MNMによる判別手法の評価：MNMで他の判別手法の誤分類数を単回帰分析し評価できます。

16. 人生をかけた研究

1996年に大学に移って良かったことの一つに、研究業績として評価が低いので嫌う研究者が多いですが、年2回経済学部論集が出され自分の論文を投稿できることがあります。大学に赴任した1996年以降ウイーン滞在の1回を除いて全ての号に寄稿しています。学会誌と異なり、査読者とのやり取りの時間が限りなく少ないので、タイムリーに成果を発表し、自分にとっても研究履歴が分かります。最適線形判別関数の研究は、空理空論でなく、多くの実証研究に裏付けられているので、タイムリーに記録を紀要に残せたことはありがたいことです。

16-1 博士号の研究テーマの決定

日本医科大学の三宅章彦先生とは、Fisherの仮説を認めたうえで母集団と標本の誤分類確率に関する研究[41]を行いました。また誤分類数最小化（Minimum Number of Misclassifications, MNM）基準による最適線形判別関数を、ヒューリスティックなアプローチで行いました[53]。IBMの汎用機でアンダーテーブル研究として行ったため体系的な研究ができず、CPDデータを用いた6変数の分析結果だけです。また、SASで正規乱数を作って評価データとして分析しましたがそれほど良い結果ではありませんでした。そこで研究は

それ以上継続せず、終わりになりました。

1995年の秋に両国の住商情報システム(株)で私が大会長になり、日本計算機統計学会のシンポジュームを開きました。1996年には成蹊大学の教員になってからすぐに、岡山大学の垂水先生から「うちで学位をとりませんか」といわれていたが回答していませんでした。1997年9月に東北大学で行動計量学会がありました。夕食を食べに街に出ると、垂水先生にあって再度すすめられました。「大学に移って間もないので、岡山まで通うことはできないのですが?」というと「論文博士で考えています」ということで、数日考えて学位申請することにしました。研究テーマを夜考えていて、「数理計画法は関数の最大/最小値を求める学問である」から、誤分類数最小化基準による最適線形判別関数が定式化できないはずがないと思い考え始めると、すぐに定式化を思いつきました。すでにLinus教授のテキスト(実践数理計画法(朝倉、1992))の翻訳で回帰分析の数理計画法による定式化は知っていたので、それほどハードルは高くありませんでした。もし私が普通の研究者のように、既存の論文を探していたら1980年代から線形計画法を使ったL1ノルム基準による判別関数の論文が数多く出ていたので、整数計画法を用いた最適線形判別関数などの研究は行わなかったでしょう。実はその事実を知ったのは、2005年ごろLinus 教授から届いた先行研究の論文リストからです。

自分の人生の勉強のテーマの統計と数理計画法はソフトウエアを使って独学でマスターしていましたが、それらが連携していませんでした。日本では統計と数理計画法は数理工学科などで教えられていますが、同級生であっても専攻が違うと研究では永遠の平行線になります。それがこれでようやく自分の中でもつながることになりました。

16-2 博士号取得まで (1998年~2000年3月) [54-60]

(1) FisherのアイリスデータとCPDデータによる実証研究

Fisherのアイリスデータは、FisherがLDFの検証に用いたため、こう呼ばれていますが、実際はEdgarが集めたデータです。Fisherの仮説も誰が考えたのかわかりません。折につけ、「ご存知の方は教えてください」と言っていますがいまだに分かりません。そしてLDFをFisherの仮説を満たさないデータに適用してはいけないという方もいます。「判別したい2群が多次元正規分布か否かを示す良い統計量はないので、どう判定するのでしょうか。」このように本人の同意を得ないでFisherの名前を付けるのは、Fisherも望んでいないと思うがいかがでしょうか? 本データは、相関係数を散布図で事前にチェックしないと間違った判断をするという統計入門の良い教材になります。IP-OLDFとLDFとQDFで、15個(=2^4-1)の判別モデルを分析し15個の誤分類数を評価しました。2群の分散共分散行列が等しくないときLDFに代わってQDFを用いることを判別理論で勧めていますが、その場合でもQDFの誤分類数が多いものが出てくるのでこのガイドラインは間違っていることが分かりました。

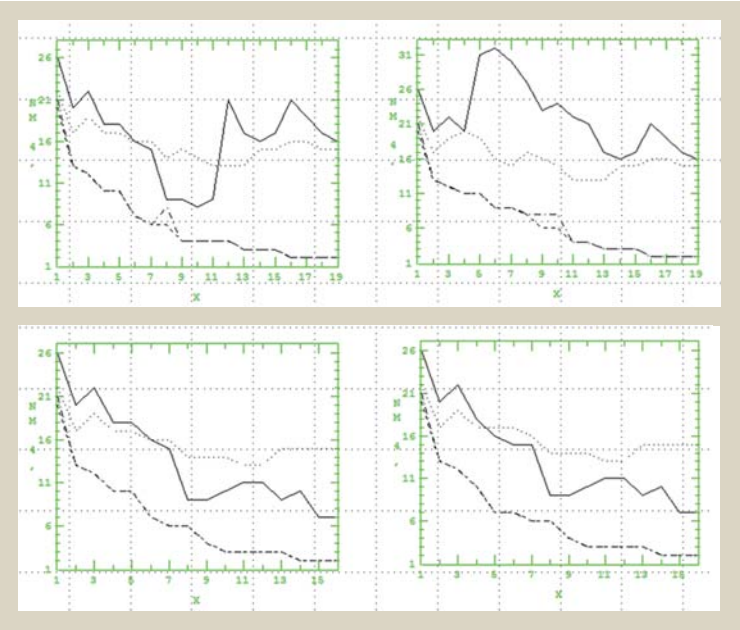


図2:CPDデータの誤分類数(上左:19変数の増加法、上右:減少法、下左:16変数増加法、下右:16変数減少法)

CPDデータは17個の計測値と、2組の2変数の差で作られた2変数の計19変数あります。作成された2変数が判別に重要であることを示すことを目的としていますが、VIFで多重共線性があることが分かります。19変数と多重共線性を解消した16変数で変数増加法と変数減少法を行い40個の判別モデルに限定して考えます。この成果はSCSのIBM上ですでにSASのRSQUAREを使って分析していたものを再利用し、QDFとIP-OLDFとLP-OLDFの誤分類数を付け加えました。**図2の上**は、19変数のフルモデルに対して左は変数増加法、右は変数減少法の誤分類数をプロットしました。実線はQDFを表します。変数増加法では10変数まで誤分類数は減少し、それ以降増えるのは多重共線性に関係する変数がモデルに入ってくるからです。点線はLDFで、誤分類数が13以下に減らないことが分かります。証明したわけではありませんが変数選択法で選ばれた説明力のある有力な変数で作られた分散共分散である程度の構造が決まってしまうのではないかと考えています。これに対してIP-OLDFとLP-OLDFは8変数でLP-OLDFが悪くなっていますが、単調に減少しています。これからMNMの単調減少性という重要な事実が分かります。右の変数減少法では、QDFは6変数まで誤分類数が増加し、多重共線性に関係する変数が掃き出されると減少しています。それ以外は変数増加法とほぼ同じ傾向を示します。

図2の下は、16変数のモデルの誤分類数のプロットで、QDFは多重共線性がなくなったのでどちらもほぼ減少傾向を示します。その他は19変数とほぼ同じ傾向です。以上からQDFは多重共線性に対して問題があることが分かります。このような大きな瑕疵をもつ手法は、利用すべきではないということです。

図3は、QDF(1点鎖線)、LDF(直線)、LP-OLDF(点線)、MNM(下の直線)の誤分類数を目的変数にしてMNMで単回帰したものです。X軸はMNMの値なので大きくなると説明変数が少ないことを表します。

これによってQDFは一般的にMNMより12例（誤分類確率で0.05）悪くなります。LDFはフルモデルでは12例ほど悪いですが、説明変数が少なくなると2例まで少なくなります。

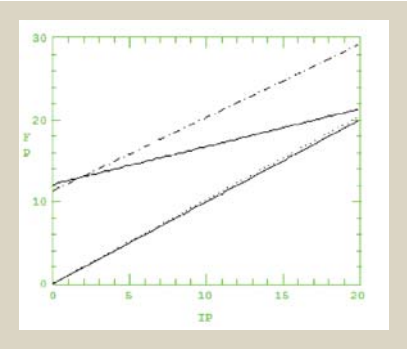


図3：各判別関数の誤分類数をMNMで単回帰
（QDF：1点鎖線、LDF：直線、LP-OLDF：点線、MNM：下の直線）

（2）2変量正規分布による教師データと評価データの検証

IP-OLDFとLP-OLDFに関する理論と、アヤメとCPDデータの実証研究を途中成果として垂水先生に送ったところ、乱数実験も付け加えるようコメントが来ました。統計研究では、乱数データによって研究内容の正当性を証明するのが常套手段です。しかし、「現実のデータはFisherの仮説を満たすものが少ない」という仮説にたって研究を進める予定であったので、種々の特徴のある実データで検証しようと考えていました。しかし、教師データ（内部標本）で判別関数を作成し、分析に用いていない評価データ（外部標本）で最終的に評価する必要があります。実データを教師データとして良い結果を得ましたが、評価データに関する見通しがありませんでした。そこでSpeakeasyという数学ソフトでXの標準偏差が2で、Yが1の2変量正規乱数を400個作成しました。最初の100件は群1の教師データ、次の100件を群1の評価データとし、次の200件を群2の教師データと評価データとしました。

そして群1の平均を原点に固定し、0度、30度、45度、60度、90度で回転した5組の標本を行列の積で作成しました。次に群2のXの値には0から8までの整数i、Yに0、2、4の整数jを加えて平行移動しました。これらの5*9*3個の組合せの教師と評価データを作成して、検証を行うことにしました。群1で回転を0度、群2でY=0に固定した8組の教師データと評価データのペアだけがFisherの仮説を満たします。それ以外はお互いの長軸が平行でないため、XとYの分散が異なり、Fisherの仮説の「2群の分散共分散が等しい」を満たさない状態を作ることができます。

教師データで各誤分類数をMNMで単回帰分析し「IP-OLDF＜QDF＜LDF＜LP-OLDF」という傾向にあることが分かります。評価データでは「QDF＜IP-OLDF＜LDF＜ LP-OLDF」という結果になり、QDFが一番良くなりました。わずか2変数の正規乱数であるため、LDFよりQDFの判別成績が良くなります。乱数を用いることで、頭の中で考えた分散共分散が異なる場合は、LDFに代わってQDFを用いるというガイドラインの正しいことが分かります。しかし、多重共線性にあるCPDデータの実証研究で、QDFには問題であることが分かったので、このガイドラインは利用しない方がよいことになります。

（3）学位取得

1999年秋に、田中豊先生を主査、垂水先生を副査とした論文審査が行われました。英語での1時間の発表を何とか乗り切りました。そして2000年3月に学位授与式に臨みました。1971年の大学卒業式以来でした。実は研究を始めた当初、100頁以上の英語の学位論文と、ドイツ語の筆記試験があると聞かされ、辞退しようかと思いました。数学科にはいったのは、Gaussにあこがれてゲッチンゲン大学に留学したいという漠然とした夢がありました。入学後、フランスのブルバキと呼ばれる数学者集団の書籍の影響か、フランス語を取りました。大学のドイツ語の先輩教員に速習法を聞きましたが、できの悪い学生のような質問に対して要領を得ませんでした。学位をとるために、ドイツ語の勉強に時間を割くつもりもありませんでした。しばらくして、今年から学位論文が英語であれば、第2外国語の筆記試験がなくなっていたと連絡があり、首一つで助かりました。実は学位取得後の研究で一番苦労したのは、現実に合った評価データを得ることでした。実は、2004年にウイーンでJMPの解説書のレビューをしていて乱数生成法が載っていて、もう少し早くJMPユーザーになっておればと思いました[24]。しかし苦しまぎれに、群1を回転させ、群2は平行移動させて135組の異なった2群の判別問題を作るアイデアは、自分としては気に入っています。読者も分析したいデータを用いて条件を変えて評価データを自由に作成すればよいと考えます。

また、2000年にヘルシンキで開催されたISIの後で、エストニアの古都のタルトの会議で発表しました。幸いに論文の投稿を勧められ10頁の論文が掲載されました[58]。幸先は良かったのですが、その後に外国語論文の投稿をしたところ、「信じられない」とか「分からない」というおよそ研究者のコメントとは思えない理由でリジェクトされることが多かったです。何も理論だけでなく、検証可能な実証研究も行っているのに中々理解されないことは、自分にとって残念に思います。すなわち、信じられない結果であれば自分で簡単に追試してみればすぐに分かることです。

文献：

[24] 新村秀一（2004）.（レビュー）JMPを用いた統計およびデータ分析入門. SAS Japan
[41] A.Miyake & S.Shinmura (1976). Error rate of linear discriminant function, F.T. de Dombal & F.Gremy , editors 435-445, North- Holland Publishing Company
[50] 新村秀一（2007 ）. ExcelとLINGOで学ぶ数理計画法. 丸善.
[51] 新村秀一(2010). 最適線形判別関数. 日科技連.
[52] 新村秀一（2011）. 数理計画法による問題解決法. 日科技連.
[53] 三宅章彦, 新村秀一（1980a）. 最適線形判別関数のアルゴリズムとその応用, 医用電子と生体工学, 18-1, 15-20.
[54] 新村秀一（1998）. 数理計画法を用いた最適線形判別関数. 計算機統計学. 11-2,89-101.
[55] 新村秀一, 垂水共之（1999）. 2変量正規乱数データによるIP-OLDFの評価. 計算機統計学 12-2, 107-123.
[56] 新村秀一（2000）. 数理計画法を用いた最適線形判別関数（1）. 成蹊大学経済学部論集, 30-2, 193-215.
[57] 新村秀一（2000）. 最適線形判別関数（2）ーアイリスデータへの適用ー. 成蹊大学経済学部論集, 31-1, 297-313.
[58] Shuichi Shinmura. (2000). A new algorithm of the linear discriminant function using integer programming. New Trends in Probability and Statistics 5,133-142.
[59] 新村秀一（2001）. 数理計画法を用いた最適線形判別関数（3）ー多重共線性のある医学データの分析ー. 成蹊大学経済学部論集, 31-2, 207-236.
[60] 新村秀一（2001）. 数理計画法を用いた最適線形判別関数（4）ー2変量正規乱数データによるIP-OLDFの評価—. 成蹊大学経済学部論集, 32-1, 249-273.

Q&A

- 複数の SAS ウィンドウを起動すると WARNING が表示される
- バブルチャートの作成
- SAS® Enterprise Guide® 上で使用可能な機能の確認
- 前0なしの日付データを出力したい
- 現在適用可能なHot Fixの確認と入手方法について

Q 複数の SAS を同時に起動すると次の WARNING が表示されます。また、2 つ目以降のウィンドウでは以前設定した SAS ウィンドウに関する設定がロードされていません。

WARNING : SASUSER レジストリを WORK レジストリにコピーできません。このため、このセッション中にレジストリをカスタマイズすることはできません。

A 複数の SAS を同時に起動した場合、二つ目以降の SAS は画面表示などの設定情報をロードすることができない仕様となってます。
これは、設定情報を格納するファイルに対して複数の SAS が情報の更新を行うと不整合が発生する恐れがあるためです。
次のような設定を行うことで、複数の SAS に設定情報を引き継ぐことができます。

1. SAS のデスクトップショートカットを用意します。

例

```
SAS 9.3
```

（必要に応じて、ショートカットのコピーを作成します。ショートカット名の例：RSAS 9.3）

2. ショートカットを右クリックし、プロパティを選択します。ショートカットタブの「リンク先」に格納されているパスの末尾に -RSASUSER という文字列を追加します。
3. OK ボタンを押してショートカットへの変更を保存します。
4. ショートカットから SAS を起動することで WARNING を回避できます。また、2 つ目以降のウィンドウでも画面の設定情報がロードされます。

Q ODS統計グラフ機能にて、バブルチャートを作成できますか。

A SAS® 9.2 TS2M3 以前では、ODS 統計グラフ機能にてバブルチャート作成に対応しているプロシジャはありません。このため、SAS/GRAPH® の GPLOT プロシジャ、BUBBLE ステートメントを用います。
SAS 9.3® にて、新たに SGPLOT プロシジャ、SGPANEL プロシジャに BUBBLE ステートメントが追加され、以下のようにバブルの位置を示す変数を X=、Y= オプション、バブルのサイズを示す変数を SIZE= オプションにて指定し、バブルチャートを作成できます。

例

```
PROC SGPLOT DATA=one;
  BUBBLE X=xvar Y=yvar SIZE=sizevar;
RUN;
```

SAS 9.3 では、BUBBLE ステートメントの他、以下のステートメントが SGPLOT プロシジャ、SGPANEL プロシジャに追加されています。

HBARPARM/VBARPARM
集計されているデータセットに基づく棒グラフの作成

HIGHLOW
上限と下限をつなぐライングラフ

LINEPARM
点と傾きの情報から直線をグラフ上に描画

WATERFALL（SGPLOT のみ、評価版）
ウォーターフォールチャートの作成

Q SAS® Enterprise Guide® 5.1 を使用しています。システム管理者により、ユーザーアカウントごとに Enterprise Guide の機能に制限がかけられているのですが、どの機能が使用可能かわかりません。システム管理者に確認するしか方法はないのでしょうか。

A SAS® Enterprise Guide® 4.2 以降のリリースでは、システム管理者へ確認、もしくは SAS 管理コンソール等のプロダクトを使用しなくても、現在 Enterprise Guide を使用しているユーザーアカウントがどの機能を使用できるか確認できるようになりました。
Enterprise Guide のメニューにある下記の項目で確認可能です。

ヘルプ -> SAS Enterprise Guide のバージョン情報 -> 機能設定

Q 1998/07/01 ではなく、1998/7/1 のように、月日の前に 0 を付加しない形で出力したいのですが、可能でしょうか。

A FORMAT プロシジャにて、%Y、%m、%d ディレクティブを用いることで可能です。

例

```
DATA _date;
  INPUT date YYMMDD10.;
CARDS;
1976/04/05
1976/09/02
2002/09/03
;
RUN;

PROC FORMAT;
  PICTURE datefmt LOW-HIGH='%Y/%m/%d'
  (DATATYPE=date);
RUN;
PROC PRINT DATA=_date;
  FORMAT date datefmt10.;
RUN;
```

Q 現在 SAS® 9.2 を利用しています。SAS の Hot Fix ページでは 9.2 でも種類がありどの Hot Fix を適用するのかよくわかりません。一括して最新の Hot Fix を入手することはできますか。

A SAS 9.2、SAS® 9.3 の双方で利用可能な Hot Fix の分析、ダウンロード（と適用）用のツールとなる SASHFADD ツールが利用可能です。
SASHFADD ツールを利用しますと外部のネットワークに接続できないサーバーやクライアント向けの Fix についても、別の FTP が利用可能な環境から現在使用している SAS に対応する Fix ファイルを一括して入手することが可能です。

次の手順にてこのツールを利用することが可能です。

1.SASHFADD ツールの入手

-1 FTP が利用可能な環境を用意します。
※後でファイル入手を行うには外部サイトへのFTPアクセスが必要です。
-2 SAS の Hot Fix の先頭ページにアクセスします。
http://ftp.sas.com/techsup/download/hotfix/hotfix.html
-3「SAS Hot Fix Analysis, Download and Deployment Tool」の項目にて「SASHFADD download page」のリンクをクリックしてツールのダウンロードページへ移動します。
-4 ダウンロードページにある「Download: SASHFADDwn.exe」をクリックして入手し、任意のフォルダに保存します。
-5 ダウンロードした「SASHFADDwn.exe」を実行すると WinZip の Self-Extractor が起動します。
-6 任意のフォルダを指定して「UnZip」を押し、ツールを展開します。
※この手順だけはHot Fixを適用する対象のマシン上で行う必要があります。

2.インストールされている SAS と関連プロダクト一覧ファイルの入手

-1 Hot Fix の対象とするマシンに管理者権限にてログインします。
-2 Windows のエクスプローラから SASHome ディレクトリ下の deploymentreg フォルダへ移動します。
※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。

SAS 9.2の場合

```
C:¥Program Files¥SAS
```

SAS 9.3の場合

```
C:¥Program Files¥SASHome
```

-3 deploymentreg フォルダにある "sas.tools.viewregistry.jar" をダブルクリックするなどにて実行します。
※通常ファイルの関連付けにて自動で.jarファイルの実行が可能ですが、セキュリティ関連の設定などで実行不能な場合には管理者権限で起動されたコマンドプロンプトよりJavaのコマンドから実行するなどの対応が必要となる場合があります。

例

```
java -jar sas.tools.viewregistry.jar
```

-4 ファイルの実行が完了したら "DeploymentRegistry.txt" ファイルが作成されますので任意の場所にコピーするか、同一マシン内であれば<1.-5>で作成したフォルダ内に保存します。

3.SASHFADDツールの実行

-1 SASHFADD を展開したマシンに管理者権限のある ID でログインします。
-2 Windows のエクスプローラから SASHFADD を展開したフォルダへ移動します。
-3 <2.-4>で保存した "DeploymentRegistry.txt" を同じフォルダに保存します。
-4 メモ帳やワードパッドなどのエディタから "SASHFADD.cfg" を開きます。
-5 対象の SAS リリース項目の先頭にある '#' をスペースで置き換えます。

SAS 9.2の場合

```
/* 変更前 */
# -SAS 9.2
# -SAS 9.3

/* 変更後 */
  -SAS 9.2
# -SAS 9.3
```

-6 入手対象とする Hot Fix のレベルを指定します。

すべての対象を入手する場合

```
/* 変更前 */
# -GENERATE_ALERT_ONLY_SCRIPTING
# -GENERATE_SCRIPTING_FOR_ALL_ELIGIBLE_HOT_FIXES

/* 変更後 */
# -GENERATE_ALERT_ONLY_SCRIPTING
  -GENERATE_SCRIPTING_FOR_ALL_ELIGIBLE_HOT_FIXES
```

-7 エディタを閉じて保存します。
-8 "SASHFADD.exe" を実行します。

※Windows Vistaや7の場合には必要に応じて「管理者として実行」で実行します。

-9 プロンプト画面で「任意の名称」をつけるかそのま "Enter" するかの確認に応答します。

※入力したタイトルもしくは標準の"SASHFADD"で始まる_ddMonyyyy_hh.mm.ssのフォルダが作成されます。
※自動でSASのFTPサイトに接続"DeploymentRegistry.txt"の内容に適合するこの時点での入手可能な最新のHotFixの一覧が取得され、その後の処理に必要なフォルダやスクリプトが自動生成されます。

-10 Press "Enter" のメッセージに応答して終了させます。

4.Hot Fixのダウンロード

-1 <3.-8>で作成されたフォルダに 4 つのサブフォルダが作成されます。
-2 "DOWNLOAD" で始まるフォルダに移動します。
-3 "ftp_script.bat" を実行します。
-4 SAS の FTP サイトへ自動的に接続され、"DEPLOY" で始まるディレクトリ内に Hot Fix がダウンロードされます。

上記までの手順でそのサイトの SAS および関連プロダクトのリリースに適した Hot Fix を一括で入手できます。
Hot Fix の適用については SAS のリリースや適用するプロダクトの内容により手順が変動します。
また、上記以外にも SASHFADD ツールにはさまざまな機能がありますので、詳細については次のアドレスから SASHFADD「Usage Guide」を参照願います。
また SAS 9.3 ではメディアから実行される Deployment Manager からも同様の機能を実現可能です。
詳細は次のマニュアルにて記載しております。

・SAS Deployment Wizard、SAS Deployment Manager ユーザーガイド (SDW and SDM User's Guide)
・SAS Deployment Wizard 9.3、SAS Deployment Manager 9.3 ユーザーガイド
http://www.sas.com/offices/asiapacific/japan/service/documentation/installcenter/deploywiz/9.3/user.pdf
第 4 章　ホットフィックスを探し適用する
付録 C　以前のバージョンの SAS Deployment Wizard でホットフィックスを適用する

Latest Releases

最新リリース情報

PCプラットフォーム

Windows版(32-bit/64-bit)	SAS 9.1.3 / 9.2 / 9.3
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 / 9.3

UNIXプラットフォーム

SunOS/Solaris版	SAS 9.1.3 / 9.2 / 9.3
HP-UX版	SAS 9.1.3 / 9.2 / 9.3
HP-UX(Itanium)版	SAS 9.1.3 / 9.2 / 9.3
AIX版	SAS 9.1.3 / 9.2 / 9.3
Linux(Intel) 版	SAS 9.1.3 / 9.2 / 9.3

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

STN SAS Technical News WINTER 2013

■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 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©2013, SAS Institute Inc.All rights reserved.