

パソコン用統計解析ソフトウェア STATISTICA

立教大学* 山 口 和 範

(1996年6月 受付)

1. はじめに

統計解析においてコンピュータの果たす役割は、従来の統計計算の実行だけでなく、それに伴う報告書の作成やプレゼンテーションの実施まで、幅広くなりつつある。とくに、(可搬性の) パーソナルコンピュータ (PC) の高性能化に伴い、実験計画、データ収集、データ解析、報告書の作成、プレゼンテーションに至る一連の処理が PC 上で可能となった。また、Windows や MacOS の進歩に伴い、複数のアプリケーションでの情報の交換が容易になったことがこのような作業をより簡単なものとした。このような時代においては、統計分析用のソフトウェア自身も単なる統計計算やグラフ作成のための道具ではなく、報告書の作成やプレゼンテーションへつながる道具でなければならないであろう。

コンピュータ上で日本語が自由に扱えるようになり、パソコン上でも現在では当たり前のように日本語が使用できる。しかし、現在でも海外で作成されたソフトウェアの日本語化には様々な問題があり、PC 用の OS の提供においても英語圏に比べ数ヶ月のラグがある。統計解析用のソフトウェアも報告書の作成やプレゼンテーションへの連絡性を考えると、自由に日本語が使用できることは、一般ユーザーにとり必要不可欠なことである。しかしながら、海外で開発されたソフトウェアの日本語版ではこれまで日本語が必ずしも完全に自由に使用できるわけではなかった。統計ソフトウェアがデータ解析の裾野を広げ、さらに PC の普及がこの傾向を顕著にしつつある現況において、統計の専門家だけでなく一般ユーザーも対象としたソフトウェアであれば、完全な日本語化が必須であろう。

STATISTICA は、米国の StatSoft 社が開発した統計解析用のソフトウェアで、1991年に最初の DOS 上のバージョンが発売されている。StatSoft 社は 1984 年設立され、その最初の製品は PsychoStat と呼ばれたデータマネージメント機能を備えた統計計算用のライブラリ集で、メニューから起動できるものであった。このプログラムは PC DOS や Apple II, CP/M などのパソコンで稼働し、必要とするメモリーは 64 k であった。1986 年より開発が始まった CSS (Complete Statistical System) では、グラフ機能の充実が試みられた。またこのパッケージにおいてすでにスクロールシートと呼ばれる出力機能が組み込まれている。

この CSS が STATISTICA の前身となり、91 年に DOS バージョン、翌年 Macintosh バージョン、そして、93 年に Windows バージョンが発売された。日本語化については、Macintosh バージョンと Windows バージョンで各々別会社が行っている。Windows バージョンについては 94 年に正式リリースされ、そこではとくに変数名での日本語の使用が、プログラム内での使用も含め、可能になっている。現在の最新バージョンは英語版、日本語版共に Windows バージョンの 5 である。

* 社会学部：〒171 東京都豊島区西池袋 3-34-1.

本稿での STATISTICA の紹介は Windows 用の日本語バージョン 5 を基に行う。

2. STATISTICA

2.1 STATISTICA の構成

STATISTICA はモジュールと呼ばれる統計解析用プログラム群から構成されている。各モジュールはそれぞれが各統計手法に対応しており、バージョン 5 で 18 個の統計分析モジュールが用意されている (表 1 参照)。またデータ管理用のモジュールとしてデータ管理モジュール、メガファイルマネージャー、ファイルサーバーが用意されている。STATISTICA で一度に使用できる変数の数は通常 4,092 であるが、メガファイルマネージャーを使用すると 32,000 個の変数の管理が行える (関田 (1995))。

STATISTICA を起動すると、モジュールスイッチャ (図 1) が呼び出され、そこで分析した

表 1. STATISTICA の各モジュール名。

一般統計解析モジュール	品質管理関連モジュール	データ管理モジュール
基本統計/集計表	工程解析	データ管理機能
ノンパラメトリック統計	品質管理	ファイルサーバー
線形回帰分析	実験計画法	メガファイルマネージャー
非線形回帰分析		
分散分析/多変量分散分析		
時系列分析		
生存時間分析		
対数線形モデル分析		
判別分析		
因子分析		
共分散構造分析(SEPATH)		
正準相関分析		
クラスター分析		
多次元尺度構成法		
試験の信頼性と項目分析		

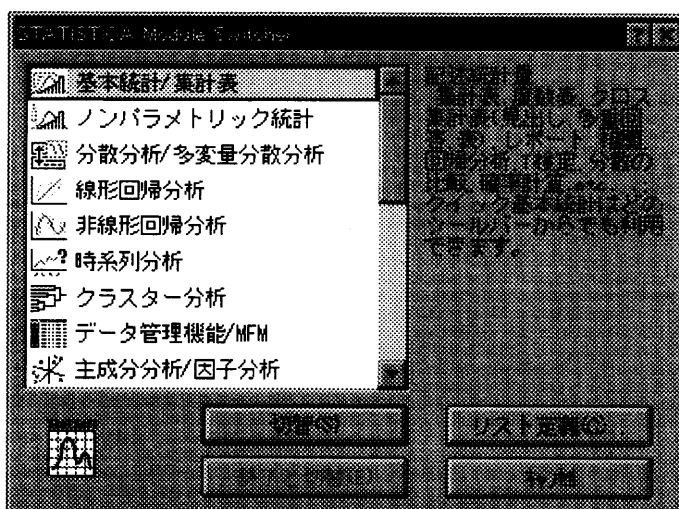


図 1. モジュールスイッチャ。

い手法が含まれるモジュールを選択する。マルチタスク機能により複数のモジュールを同時に起動することも可能で、モジュール間での情報の交換も自由である。

STATISTICAには2種類のプログラム言語が用意されている。SCLとSTATISTICA BASICである。SCLは統計処理をバッチ形式で行いたい場合に使用するもので、SAS（新村（1994a）参照）やSPSS（竹村 他（1991）、Norusis（1990）参照）の言語に近い構成になっている。一方、STATISTICA BASICは、新たな統計手法を組み込んだり、変数変換を行ったりするためのもので、その名の通りBASICによく似た言語である。このSTATISTICA BASICには、行列演算やユーザーインターフェースを提供する関数も用意されており、新たな統計手法

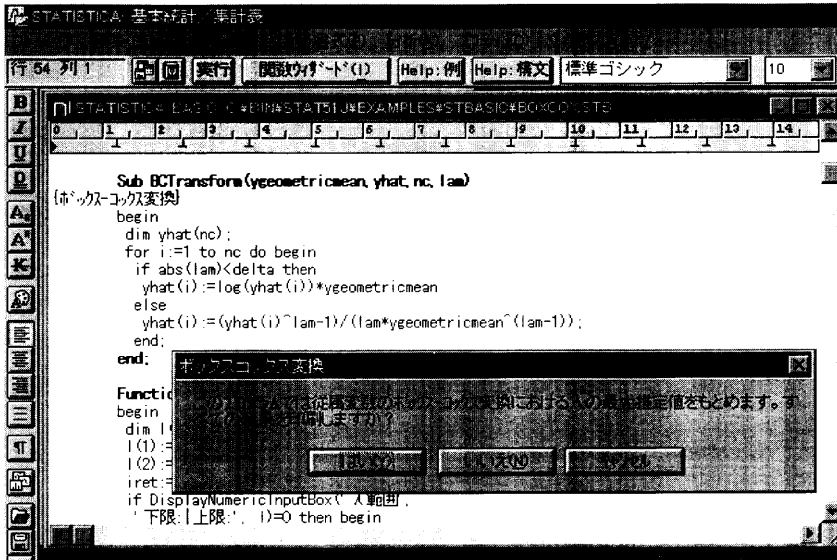


図2. STATISTICA BASIC のプログラムとその実行画面。



図3. STATISTICA のインターフェース。

を、STATISTICA のユーザーインターフェースを生かしたまま、STATISTICA に組み込むことができる (図2)。

STATISTICA のユーザーインターフェースは図3のようにメニューボタンを中心としたもので、基本的にはインタラクティブな統計解析を行うためのものである。統計解析を行う場合は定型的な処理の場合も少なくない。このような目的においては、必要な分析手順をSCLで(必要があればSTATISTICA BASICも使用)記述するか、または、そのような処理手順をメニュー上で一度実行し、その手順を再実行可能なマクロとして記録しておく。こうして作成されたプログラムやマクロをタスクボタンに割り当てておけば、表示されているそのボタンをクリックするだけで必要な処理が実行できる。

2.2 STATISTICA の特徴

STATISTICA の計算結果は必ずスクロールシートに出力される (図4)。このスクロールシートは、表計算ソフトウェアのスプレッドシートによく似たものである。出力先がスクロールシートであることの最大のメリットは、再計算時に利用できる桁数と表示に際しての桁数を変えることができる点である。結果の表示の際には、小数点以下2桁程度でよいものが、出力結果からの再計算が必要とされる場合にはかなりの精度が要求されるであろう。この桁数の違いを吸収させるための工夫が、単なるテキスト出力ではなく、スクロールシート出力である。

ただし、一連の分析において出力が複数のスクロールシートに分かれてしまうことは、結果をまとめるにあたって煩雑になる可能性もある。もちろん、他のソフトウェア同様、出力をテキストにしてしまうこともできるが、バージョン5からワークブック機能が導入され、各分析において必要なデータファイル、出力されたスクロールシートやグラフを必要に応じてまとめて保存できるようになった。

STATISTICA のデータは基本的にはすべて数値で保存されている。文字情報のデータ(名義尺度や間隔尺度とみなされるデータ)も内部的には整数値で保存されている。もちろん文字データをそのまま入力することもできる。ただし、入力された文字データに対し、各変数ごとに連続した整数値が割り当てられる。データはその数値データと各変数ごとのコーディング表から構成される。出力に関しては、ほとんどの場合、文字データで行われるので、ユーザーがこの構造を特に意識する必要はない。しかし、グループ分けの指定を行う際にこのような離散データを含む変数を使用するので、例えば、分析中に1つのグループを選択するような場合に、そのグループに対応する整数値を選択しなければならないことがある。また、名義尺度としてしかみなせないようなデータでも数値で保存されている関係上、指定すれば量的な分析(たとえば、平均や標準偏差の計算)が行われてしまう。ソフトウェアの中には(たとえば、GENSTAT, Lane and Payne (1996)), 変数の属性をあらかじめ固定し量的変数や質的変数を明確に区別し、このような分析を行わせない工夫が施してあるものがあるが、STATISTICA ではこの点は

BASIC STATS		N=60 (従属変数内に欠測値はありません)		
方法	学年	理論値	観測値	標準偏差
普通	1年	24.70	10	6.68
普通	2年	56.40	10	9.95
普通	3年	66.30	10	8.33
普通	4年	41.70	10	8.10
普通	5年	55.60	10	8.73
普通	6年	60.10	10	13.26

図4. STATISTICA のスクロールシート。

ユーザー任せになっている。

Windows 上の他のソフトウェアとの情報交換機能であるが、通常のクリップボード経由の情報交換機能に加え、さらに、DDE、OLE (サーバーとクライアント両モード)機能も装備されている。このため、STATISTICA での計算結果や作成したグラフを他ソフトウェアに貼り付けたり、また逆に STATISTICA で作成中のグラフの中に他のソフトウェアで作成した数式やイラストなどを貼り付けたりすることが可能である。

さらに、Windows DDE/API インタフェースを STATISTICA がサポートしているため、他のソフトウェア (たとえば、Excel のマクロ) から STATISTICA を起動し、データの分析を実行させることもできる (山口 他(1994)参照)。

データベースとの連結については、Microsoft 社提唱の ODBC (Open Data Base Connectivity)の採用により、SQL サーバーを含む各種データベースから STATISTICA へのデータの取り込みなどが行える。

2.3 データ管理

STATISTICA にはデータ管理機能と呼ばれるデータの変換や他のソフトウェアの形式のデータファイルの読み込みのためのモジュールが用意されている。ただし、簡単な変数変換は、スクロールシート上の数式を使えばデータ管理モジュールを呼び出さなくとも各モジュール内で実行可能である。用意されている変数変換用の関数を使用することもできる。複雑な変数変換やコード化については STATISTICA BASIC を使用することもできる。

欠測値はスクロールシート上では空白で表示され、内部的には指定されたコード (デフォルトでは -9999)で保存されている。なお、この欠測値のコードは変数毎に変えることができる。欠測値の処理の方法としては、分析から (欠測値があればすべて、または、対象となった変数が欠測値の場合のみ) 削除、平均値や指定された値での置換などがある。

3. 使用事例—混合実験の計画と解析—

本章では混合実験の事例を用い、STATISTICA の特徴を紹介したい。実験計画モジュールには従来のバージョンに比べ、STATISTICA バージョン5から、かなりの機能が新しく組み込まれている。ここで、紹介する混合実験に関する機能もバージョン5で新しく組み込まれたものである。

3.1 混合実験

混合実験とは、複数の原料を混ぜ合わせて生成物を作る際に、目的変数の最大化または最小化を達成する混合比を探索する実験のことである。このような実験データの解析法として応答曲面法があり、この分析法においてはグラフの活用が特に重要である。詳しくは、岩崎(1994)、Cornell (1990) 等を参照されたい。ここでは Myers and Montgomery (1995) の例題 11.1 でのデータを使い、STATISTICA のグラフ機能により、Myers and Montgomery (1995) で述べられているような分析が簡単に実行できることを示す。

なお、このような混合実験においては、その実験計画の段階において実験点の選択のためにもコンピュータを利用するが、STATISTICA でも実験計画策定のためのルーチンも用意されている。

3.2 STATISTICA での設定

表2がここで分析するデータである。データの入力方法としては STATISTICA のスプレッ

表2. 例題データ (Cornell (1990) からのデータ).

A	B	C	Elongation Value (Y)
1.0	0.0	0.0	11.0
1.0	0.0	0.0	12.4
0.0	1.0	0.0	8.8
0.0	1.0	0.0	10.0
0.0	0.0	1.0	16.8
0.0	0.0	1.0	16.0
0.5	0.5	0.0	15.0
0.5	0.5	0.0	14.8
0.5	0.5	0.0	16.1
0.5	0.0	0.5	17.7
0.5	0.0	0.5	16.4
0.5	0.0	0.5	16.6
0.0	0.5	0.5	10.0
0.0	0.5	0.5	9.7
0.0	0.5	0.5	11.8

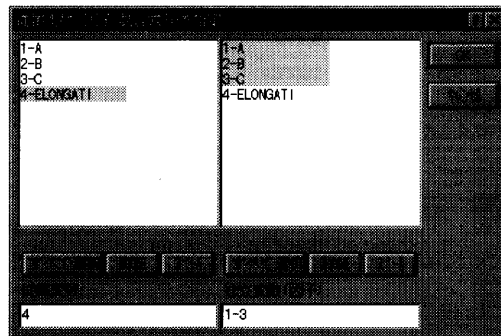


図5. 変数指定のダイアログ.

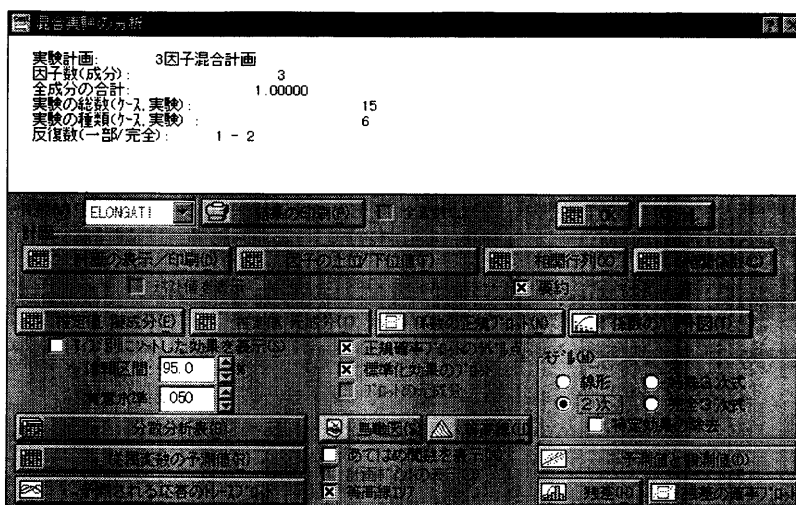


図6. 結果の表示用のダイアログ.

EXPERIM. DESIGN 3 因子混合実験計画；合計 = 1.. 15 実験
より複雑な計画へ急速的にあてはめられている

因子	水準	水準数	水準名	水準値	水準平均	F	p	MS	MS誤差
因子1	1	2	28.31	77.33	12	6.436	4.46	0.95	427
	2	3	23.55	6.56	9	729	32.32	0.000	951
合計		134.9	14	9.63					

EXPERIM. DESIGN 3 因子混合実験計画；合計 = 1.. 15 実験

因子	水準	水準数	水準名	水準値	水準平均	F	p
因子1	1	5	25.65920	35.20317	0.00010		
	2	9	72889				
因子2	1	0	0.00000				
	2	9	72889				
合計		134.8560	14	9.63257			

3 因子混合実験計画；合計 = 1.. 15 実験
DV: ELONGATI, 残差のMS = 726889

因子	水準	水準数	水準名	水準値	水準平均	F	p	MS	MS誤差
因子1	1	11	70000	603692	19.33073	0.00000	10.3344	13.06565	
	2	9	40000	603692	15.57035	0.00000	8.0344	10.76565	
因子2	1	16	40000	603692	27.16516	0.00000	15.0344	17.76565	
	2	19	00000	2.606249	7.29453	0.00046	13.0997	24.90027	
因子3	1	11	40000	2.606249	4.37075	0.01795	5.4997	17.30027	
	2	9	60000	2.606249	-3.69363	0.05071	-15.5003	-3.69973	

図7. 推定結果や分散分析表の出力。

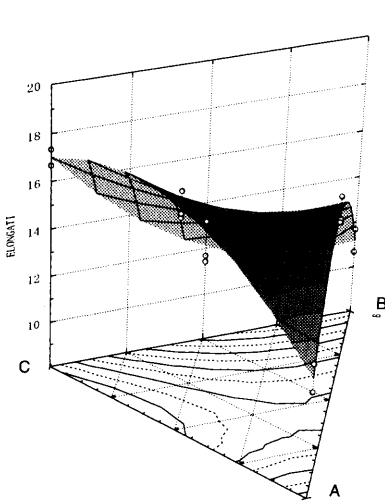


図8. 応答曲面の3次元グラフ。

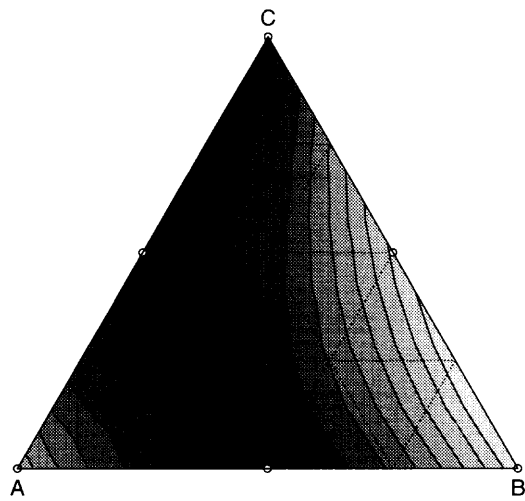


図9. 応答曲面の等高線グラフ。

ドシートに直接キーボードから入力する方法，テキストエディタで作成されたテキスト形式のデータファイルを読み込む方法，他のソフトウェア(スプレッドシートやデータベース，SPSS)でデータの入力を行いそのデータファイルを STATISTICA データファイルに変換する方法(STATISTICA ファイルメニューのインポート)などがある。

分析の設定は表示されるダイアログボックスで行う。ここでは得られた実験結果の分析を行うので，結果の分析を選択する。次に分析する変数の設定であるが，変数ボタンをクリックして，従属変数と独立変数の指定を行う(図5)。変数の指定が終わった後に OK ボタンをクリックすると，結果を表示させるためのダイアログが開く(図6)。ここで，まずあてはめる応答曲面モデルを選択する。Myers and Montgomery (1995) の分析と同様にここでは2次モデルを選択し，後は必要な出力のためのボタンをクリックするだけである。ここでは分散分析表や係

数の推定値, 等高線図, 曲面図を出力しておく (図7~図9).

4. ま と め

STATISTICA はメニューを中心とした対話的なデータ解析のためのツールである。STATISTICA は, 従来からある SAS (市川・大橋 (1987), SAS Institute (1990) 等参照) や SPSS (SPSS (1988), Norusis (1990), 新村 (1994b) 等参照) などの統計解析ソフトウェアに比べ, 必ずしも目新しい分析手法が組み込まれているわけではない (StatSoft (1995)). STATISTICA が他の統計解析ソフトウェアに比べ特徴的である点は, そのインタフェースとグラフ機能である。STATISTICA にも定型的な統計処理のための機能も組み込まれているが, STATISTICA がその特徴を発揮するのは, グラフを活用した探索的なデータ解析や, 分析結果に基づき報告書やプレゼンテーションを作成する場面においてであろう。

パソコン上でも本格的なマルチタスク機能が実現されつつある現状では, 統計解析用ソフトウェアに関しても, 必ずしも1つのソフトウェアがすべての機能を兼ね備えておく必要はない。複数のソフトウェアで情報交換を行いながら分析を進めていくという, 分業の時代である。その意味では統計解析ソフトウェアにとっても, 他のソフトウェアとの情報交換機能の充実は非常に重要な要素である。STATISTICA は Windows の OLE や DDE をサポートしており, 他のソフトウェアとの連携を十分に考慮した設計になっている。

パーソナルコンピュータの低価格化に伴い, ソフトウェアの低価格化競争も始まり, ワープロソフトや表計算ソフトではかなりの低価格が実現されている。これからのマルチタスクの分業の時代においては, 複数の統計解析ソフトウェアを使用することが考えられるが, 一般ユーザーにとって複数の統計解析ソフトウェアを準備することは現状ではかなりのコストになる。その意味でも今後, 統計解析ソフトウェアにおいてもかなりの低価格化が望まれる。

謝 辞

本稿を書くにあたり各種資料を提供いただいたデザインテクノロジー社と StatSoft 社に感謝の意を表す。なお STATISTICA 開発元である StatSoft 社の Web サーバーが 1996 年度より公開されている。アドレスは <http://www.statsoft.com> である。

参 考 文 献

- Cornell, J. A. (1990). *Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data*, 2nd ed., Wiley, New York.
- 市川伸一, 大橋靖雄 (1987). 『SAS によるデータ解析入門』, 東京大学出版会, 東京.
- 岩崎 学 (1994). 『混合実験の計画と解析』, サイエントリスト社, 東京.
- Lane, P. W. and Payne, R. W (1996). *Genstat for Windows*, Lawes Agricultural Trust, Herts, U.K.
- Myers, R. H. and Montgomery, D. C. (1995). *Response Surface Methodology*, Wiley, New York.
- Norusis, M. J. (1990). *SPSS/PC + Statistics 4.0*, SPSS Inc., Chicago, Illinois.
- SAS Institute (1990). *SAS/STAT User's Guide, Version 6*, Cary, SAS Institute, North Carolina.
- 関田素子 (1995). VisualStat (ver 5.0J) による統計解析の視覚的アプローチ, 第9回日本計算機統計学会シンポジウム論文集, 35-44.
- 新村秀一 (1994a). 『SAS 言語入門』, 丸善, 東京.
- 新村秀一 (1994b). 『SPSS for Windows 入門』, 丸善, 東京.
- SPSS (1988). *SPSS-X User's Guide*, SPSS, Chicago, Illinois.
- StatSoft (1995). *STATISTICA*, StatSoft, Tulsa, Oklahoma.

竹村和久, 山本嘉一郎, 吉村 英 (1991). 『パソコン SPSS 基礎編』, 東洋経済新報社, 東京.
山口和範, 谷田則幸, 谷田美穂子, 辻本雅志 (1994). 『VisualStat』, BNN, 東京.

A Statistical Analysis Software for Personal Computers, STATISTICA

Kazunori Yamaguchi

(College of Social Relations, Rikkyo (St. Paul's) University)

STATISTICA is an integrated statistical data analysis software for personal computers, which is developed by StatSoft. STATISTICA has a menu-driven interactive user interface. For the batch processes, two program languages, SCL and STATISTICA BASIC, and 'Auto Task Button' facility are available. STATISTICA offers comprehensive support for the Windows environment, for examples, clipboard, DDE, OLE (Server and Client), DDE/API.

STATISTICA has eighteen modules for statistical analysis and flexible graphics facilities. All numeric outputs in STATISTICA are displayed in Scrollsheet windows. Scrollsheet is highly customizable and users can control display formats of Scrollsheet. Regardless of display format, numbers in Scrollsheet are always saved in a constant precision.

Finally we demonstrate the features of STATISTICA using a small example of the mixture experiment.