

# 医薬開発部門におけるSASの利用実態と潜在的ニーズの調査と分析

中島 庸介<sup>1)</sup>, 初澤 潤太郎<sup>1)</sup>, 三沢 秀敏<sup>1)</sup>, 鈴木 昭之<sup>2)</sup>

1) ファイザー株式会社 デベロップメント・オペレーション統括部

2) ファイザー株式会社 クリニカル・リサーチ統括部

Investigation and analysis of utilization and potential needs of SAS in the drug development division

Yosuke Nakajima<sup>1)</sup>, Jyuntaro Hatsuzawa<sup>1)</sup>, Hidetoshi Misawa<sup>1)</sup>, Akiyuki Suzuki<sup>2)</sup>

1) Development Operations, Pfizer Japan Inc.

2) Clinical Pharmacology, Pfizer Japan Inc.

## 要旨

医薬品開発における SAS の利用はさまざまである。近年では EDC や CDISC などの導入に伴い、データベースや CRF の標準化が浸透してきている。これにより臨床統計・解析部門だけでなく臨床薬理部門やデータマネジメント部門などにおいても種々のプログラミングツール(SAS,S-PLUS,SQL など)による標準プログラムを使用することで業務の効率化を図っている。一方で、疾患や試験内容の違いにより標準プログラムがそのままでは利用できないケースも存在する。このような場合には、その疾患や試験のためにプログラムをカスタマイズする必要がある。しかしながら、プログラミングを主な業務としない者が、①標準プログラムを変更し、②動作を保証したりすることは難しい。したがって、臨床統計・解析部門以外の前出の①及び②のような潜在的なニーズがあると仮定し、その打開策を検討することとした。このような背景から、シックスシグマの手法に従い VOC(Voice Of Customer)の収集、分析し、Solution を立案した。VOC の分析には、SAS にて PROC VARCLUS を用いたクラスタリングも用いた。分析の結果、SAS の利用実態と幾つかの趣旨の異なる SAS 利用のニーズが明らかになったので、これらの部門に対し社内での SAS のトレーニングを実施することとした。トレーニングの教材を準備するにあたり、SAS のプログラム・コンテストを計画した。

キーワード：社内 SAS トレーニング, PROC VARCLUS, シックスシグマ, プログラム・コンテスト

## 1 はじめに

医薬品の開発は、薬として可能性のある物質を臨床試験等にて有効性と安全性を評価し、厚生労働省からの承認を得なければならない。この際、臨床試験のデータ解析にSASが多く利用されている。

弊社の医薬開発部門では、これら臨床試験のデータ処理に関わる業務にSAS等を利用したGlobalレベルでの社内標準プログラムが用意され、活用されている。現在では、臨床試験のデータ処理に関わる多くの業務が網羅されており、通常はSASやその他の言語によるプログラミングを意識せずにデータ処理を行うことができる。一方で、疾患や試験の性格によりデータの特殊な確認や取り扱いにおいては、標準プログラムの一部変更が必要になることもある。このように医薬開発部門においてSASに対する潜在的ニーズがあると考えられたことから社内でのSASトレーニング開催の有用性を検討することとした。

この時点では各部門が抱えるこのようなSAS利用の潜在的なニーズを、正確には把握していなかった。そこで医薬開発部門における各部署のSAS利用のニーズを調査・分析し、そのSolutionの立案をおこなった。調査は、選択形式と自由回答形式を組み合わせたアンケートを用いた。自由回答形式の結果の分析では、SASのプロシジャのひとつであるPROC VARCLUSを用いたクラスター分析と検討チームによる分析をおこない、この2つの分析を比較した。

## 2 調査・分析の手順および調査実施の背景と調査対象の検討

SASに係る現状調査とニーズの把握のためのアンケート調査の計画・実施からそのSolution立案までの手順を示す。

### 2.1 調査・分析の手順

本取り組みは、シックスシグマのひとつの手法であるDFACEのステップに倣って調査・分析を進めた。DFACEは、定義(Define)、フォーカス(Focus)、分析(Analyze)、創造(Create)、評価(Evaluate)の5つのフェーズから成る手法の略称で、VOC(Voice Of Customer)により顧客の声を集め、ニーズに合致した新しい価値を創造するために用いられる<sup>1)</sup>。本稿では、VOCによるSASの利用実態と潜在的ニーズの調査を起点に、調査結果の分析およびSolutionの立案までを扱う。これら手順を図1として示す。

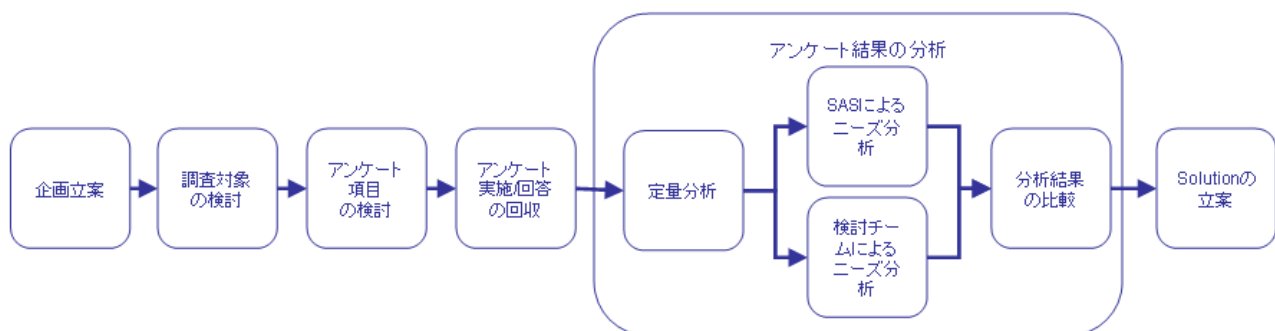


図1 調査・分析の手順

アンケート結果の分析では、定量分析に加え、自由回答形式によって得たニーズをSASのプロシジャであるPROC VARCLUSを用いて分析し、同時に検討チームによる分析も行った。この2つの分析結果を比較し、Solutionの立案をおこなった。

## 2. 2 調査実施の背景と事前情報の整理

調査は、臨床解析部門が企画した。以前に SAS で書かれたプログラムに関する問い合わせや社内での SAS トレーニング開催の要望があった部門から代表者を選出し検討チームとした。まず、これら部門の代表者に簡単なインタビューをおこない、アンケート調査を計画・実施する前にわかっていた SAS トレーニングのニーズを各部門別に整理した(表 1)。

表 1 アンケート調査前の部門別 SAS トレーニングのニーズ

部門	業務上必要か	SAS トレーニングのニーズ
臨床統計部	必須	なし
臨床解析部	必須	なし
CP 部	必須でない	あり
DM 部	必須でない	あり
PMS 統計解析	必須	なし

## 2. 3 調査対象の検討

以上のインタビューをふまえ、今回の SAS トレーニングは①SAS の利用や理解が業務上必須ではないが、②すでに SAS を使える環境があり、③SAS トレーニングを受講したいというニーズがあることが確認できた臨床薬理部門(CP 部)とデータマネジメント部門(DM 部)を対象におこなうこととした。

## 3 アンケートの実施と結果分析と考察

アンケートによる調査の実施とその結果の分析について示す。

### 3. 1 アンケート調査の実施

前章の事前情報を踏まえ、アンケートの調査項目を以下の通りとした。

- A) 社内での SAS トレーニングの希望の有無 (選択形式)
- B) SAS の習熟度 (選択形式)
- C) SAS の使用頻度 (選択形式)
- D) SAS トレーニングへのニーズはどんなものであるか(自由記載形式)

SAS トレーニングへの要望は、自由記載の回答とし、潜在的なニーズを調査した。これら全 4 項目のアンケートを検討チームから CP 部と DM 部のメンバーに配布し、1 週間ほどの期日で回答を得た。

### 3. 2 アンケート調査の結果分析と考察

アンケートによる調査の結果(定量)を示す。はじめに前節のアンケート項目 A)~C)の選択形式 3 項目に対する回答の分析をおこない、次にニーズの分析としてアンケート項目 D)の自由回答形式による回答結果を SAS の PROC VARCLUS を用いて分析を行い、同時に検討チームによるニーズの分析をおこなった。その後、この SAS の PROC VARCLUS を用いた分析と検討チームによるニーズ分析を比較し、考察をおこなった。

#### 3. 2. 1 アンケート結果(定量)の分析と考察(上記アンケート項目の A~C の 3 項目)

アンケート結果(定量)の分析(上記アンケート項目の A~C の 3 項目)の結果とその考察を示す。

### 3. 2. 1. 1 アンケート調査(定量)の結果

CP部,DM部の両部門のTotalで,96%のメンバーが社内でのSASトレーニングの受講を希望した。SASの習熟レベルでは,CP部のみで「SASに触ったことがない」という回答が一部みられたものの,「基礎はばっちり」及び「応用もばっちり」という回答の割合は両部門で同じであった。現状での「SASの使用頻度」は,CP部よりもDM部での利用頻度が高かった。

### 3. 2. 1. 2 アンケート調査(定量)結果の考察

アンケート調査(定量)の結果から,CP部およびDM部の両部門においてSASの社内トレーニングの受講希望者が非常に多いことがわかった。部門間の比較では,SASの習熟レベルにおいてCP部のみで「SASに触ったことがない」という回答が一部にみられたが,これは担当業務で必ずしもSASを必要としないことがそのまま反映されたと考えられた。この影響を除くとDM部とCP部に習熟レベルの差はみられなかった。SASの使用頻度では,DM部がCP部よりも高かった。これらの結果から「SASを使ったことがない」或いは「少しだけ触ったことがある」というグループと,SASの利用経験があり,使用頻度もそれなりというグループがあることがわかった。

これら2つのグループの存在を踏まえて,社内でのトレーニングを実施すること。トレーニングは数種類を用意する必要があること。トレーニングの内容についてはニーズ分析の結果を踏まえて検討することを確認した。

### 3. 2. 2 アンケート結果の分析と考察(2) ニーズの分析(アンケート項目のD)

SAS利用へのニーズの回答は自由記載形式を用いた。回答では,例えばSQL等ではプログラミングしにくいロジックを使ったチェックをしたり,データ・レビューやその他の追加的なデータ確認用資料作成のためにグラフ化したり見やすく可視化したりというニーズが見られた。他にもユーザーの業務上に関するものや,抽象的なものなどさまざまな要望がみられた。

このことから,以降の分析ではニーズの回答に発現している用語の分類による分析と,回答中に省略されたユーザーの業務内容やニーズの優先順位なども考慮した分析をおこなうこととした。

#### 3. 2. 2. 1 PROC VARCLUS を用いたクラスタリングによるニーズの分析手法

SASのプロシジャのひとつであるPROC VARCLUSを用いて,アンケートで回収したSASの利用に対するニーズをクラスタリングした。PROC VARCLUSを用いてニーズをクラスタリングする目的は,可能な限りこれらのニーズの客観的な分析結果を得ること,およびこれらニーズ応えるSolutionを検討するための資料とすることである。

ニーズの分析には,前出のテキスト・データを用いた。通常,文章中に現れる単語は文章を支える基盤レベルとなっており,意味を表現する最小の単位は単語である<sup>2)</sup>。したがって,以降の分析では分析に用いるデータ単位を単語とした。データは,表中の箇条書きの1行がひとつのニーズとなっており,本稿ではこれを各単語の集計の単位とした。

#### 3. 2. 2. 1. 1 クラスタリング対象の決定

PROC VARCLUSを用いる前に,前出のニーズの回答から単語の抽出を行いデータ化した。分析対象とした

ニーズ数は 50 であった。ここからの単語の延べ発現数は 1625 であり、単語数は 404 であった。このうち、分析の対象として意味のある名詞、サ変名詞、固有名詞、動詞の単語数は 225 であった。ニーズに現れた分析に意味のある単語の発現傾向をみるために、Total、CP 部及び DM 部における発現率を求めた。Total の集計は、CP 部と DM 部を区別せずまとめて集計した。CP 部、DM 部の集計はそれぞれに分けて集計をおこなった(表 3)。

表 3 分析で意味のある単語の発現率一覧(Total, CP 部, DM 部)

1)抽出語発現率 - Total			2)抽出語発現率 - CP部			3)抽出語発現率 - DM部		
No.	抽出語	発現率	No.	抽出語	発現率	No.	抽出語	発現率
1	プログラム	46.0%	1	プログラム	51.9%	1	プログラム	39.1%
2	書く	34.0%	2	書く	37.0%	2	書く	30.4%
3	要望	30.0%	3	基礎	29.6%	3	使用/活用	30.4%
4	SAS	26.0%	4	要望	29.6%	4	日付	30.4%
5	基礎	24.0%	5	SAS	25.9%	5	要望	30.4%
6	使用/活用	24.0%	6	Tips役に立つ	18.5%	6	SAS	26.1%
7	チェック	18.0%	7	データ	18.5%	7	チェック	26.1%
8	抽出/検出	16.0%	8	算出	18.5%	8	抽出/検出	26.1%
9	Tips役に立つ	14.0%	9	使用/活用	18.5%	9	整合	21.7%
10	データ	14.0%	10	読める	18.5%	10	リスト	17.4%
11	読める	14.0%	11	社内標準プログラム	14.8%	11	基礎	17.4%
12	日付	14.0%	12	臨床	14.8%	12	方法	17.4%
13	算出	12.0%	13	オーバーラップ	11.1%	13	期間	13.0%
14	読める	12.0%	14	カウンターパート	11.1%	14	今	13.0%
15	社内標準プログラム	12.0%	15	ギャップ	11.1%	15	読める	13.0%
16	服薬	12.0%	16	チェック	11.1%	16	実施	13.0%
17	方法	12.0%	17	教材	11.1%	17	出力	13.0%
18	参照/参考	10.0%	18	業務	11.1%	18	投与/投与量	13.0%

カイ二乗検定の 5%水準にて CP 部と DM 部との間で発現する単語に差があるかを確認した。両部門で有意な差がみられた単語は、『日付』(カイ二乗値 4.042)のみであったこれは DM 部のみに発現しており、逆にその他の単語に大きな差はみられなかった。この結果から以降で実施するクラスター分析では、CP 部と DM 部の所属を区別せずに分析をおこなうこととした。

### 3. 3. 2. 1. 2 Proc Varclus の実行

SAS のプロシジャのひとつである PROC VARCLUS には以下のような SAS データセットを作成し分析をおこなった。PROC VARCLUS とその結果を樹形図で表示するためのプロシジャである PROC TREE のプログラムも同時に示した(図 3)。

①Proc varclusで用いるデータセット

VIEWTABLE: Work.Vc01										
	ID	C1	C2	C3	C4	C5	C6	C7		
1	X01	0	0	0	0	0	0	0		
2	X02	0	0	0	1	0	0	0		
3	X03	0	0	0	0	0	1	0		
4	X04	0	0	0	0	0	1	0		
5	X05	1	1	0	0	0	0	3		

②Proc varclus及びproc treeのSASプログラム

```
*--- Proc varclus ---;
proc varclus data=VC01 outtree=VCTREE01 short;
run;

*--- Proc tree ---;
proc tree data=VCTREE01 horizontal;
height _PROPOR.;
id _LABEL.;
run;
```

図 3 PROC VARCLUS に用いるデータと SAS プログラムのサンプル

PROC VARCLUS は、変数のクラスタリングのためのプロシジャである。相関行列または分散共分散行列を

入力データとして、その変数としての特質を活かしながら直接的なクラスタリングをおこなうもので、明確に分離した分類結果が得られるため結果の解釈が容易である。PROC VARCLUS については、岸本(1996)による解説がわかりやすく詳しい<sup>3)</sup>。分析には、SAS 9. 2(Japanese)を使用した。

変数にはそのユーザーが用いた単語の頻度が入っている。ひとつのニーズの中に何度も同じ単語が登場していた場合は、その発現頻度は1とした。

分析に用いるためにニーズの回答から得られた単語は225であったが、これをそのままPROC VARCLUSに用いると、その結果225の単語がクラスタリングされることになる。これをPROC TREEで樹形図にすることは可能であるが、その結果を解釈するための整理が複雑で、オプションによるアウトプット画面の表示の調整なども必要になる。そこで、今回の分析の主旨は、ニーズの傾向を捉えること、後のSolutionの検討の資料の充足であることから225の単語をすべてクラスタリングする必要がないことを確認し、むしろそれら単語を手際よくまとめて分析を進めることとした。単語のまとめには、コーディングという手法を用いた。コーディングでは、分析の主旨に沿って単語を整理することを心がけた。コーディングを行う際には、その単語の意味するところや使用された文脈を理解する必要がある。このため、それらが用いられている原文に戻るなどしてよく確認する必要がある。最終的には異なる30程度の単語にまとめ、これを分析に用いた。ちなみに、このコーディングによる整理の工程には、コーディングのルールを明確にするなど、できる限り作業者の主観が入らないプロセスの工夫が必要である。

PROC VARCLUS を実行し、今回のニーズのデータの分析を行った。30程度の単語は、PROC VARCLUS のデフォルトの設定で10のクラスターに分類された。これはデータ全体の分散の約72%を説明していた。またこれをPROC TREEにより樹形図として表示した(図4)。

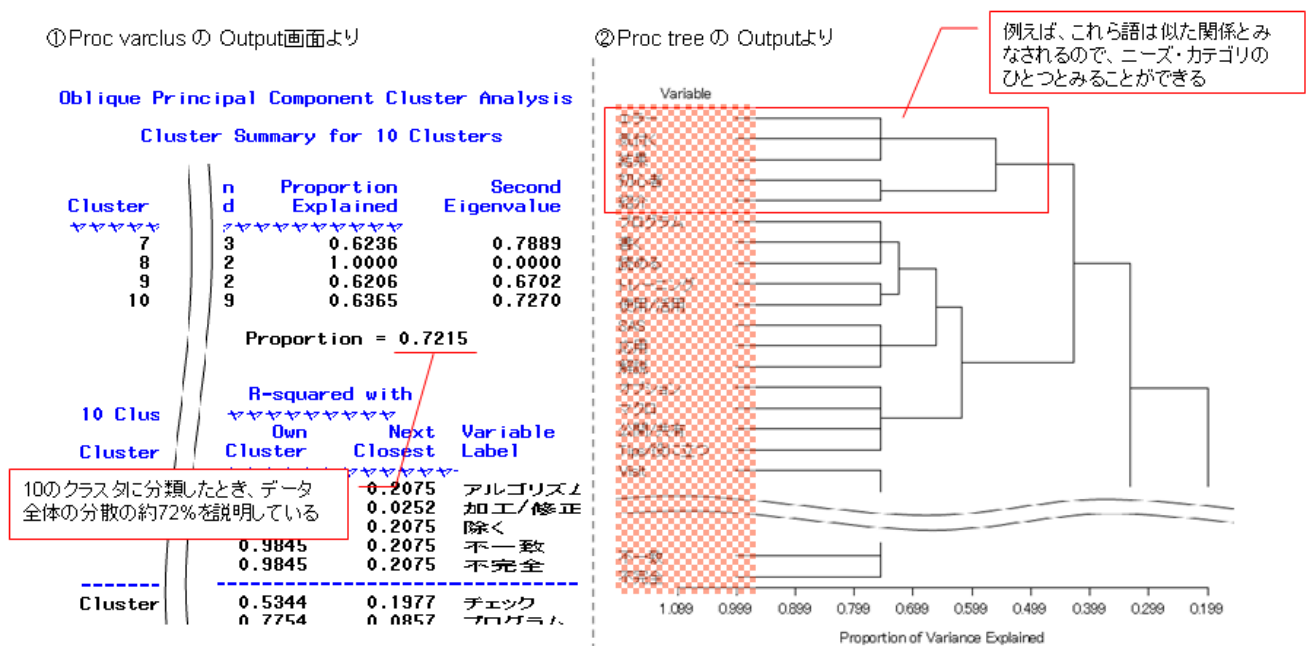


図4 PROC VARCLUS と PROC TREE の実行結果

PROC VARCLUS にて分類され、PROC TREE により樹形図が作成された幾つかのクラスターをまとめることにより、ニーズから抽出された幾つかの単語の集合をひとつのカテゴリとして意味のあるものとして捉えることが可能であった。考察については後で述べる。

### 3. 2. 2. 2 検討チームによるニーズの分析手法

PROC VARCLUS を用いたクラスタリングによるニーズ分析とは別に、検討チームによるニーズ分析をおこなった。まず、SAS トレーニングに対するニーズを調査した結果を、CP 部と DM 部について別々にそれぞれのニーズの傾向を整理した。例えば、<基礎の補完>、<実例からの学習>および<応用>といったカテゴリ等に分類した。さらに、CP 部と DM 部それぞれに整理したニーズを見比べ、検討チームによるニーズの分析結果とした。詳細な分析結果は次節に示した。

### 3. 2. 2. 3 ニーズの分析結果の考察

前節までのニーズの分析結果を示し、2つの分析結果を比較して考察する。

#### 3. 2. 2. 3. 1 PROC VARCLUS を用いたニーズ分析の結果と考察

ニーズの回答から抽出した単語は、PROC VARCLUS により 10 のクラスターに分類され、これを PROC TREE により樹形図して示した。この 10 に分類されたクラスターのそれぞれからは、ごく少ない情報しか得られなかったが、これらクラスターの隣り合った幾つかを組み合わせ、意味が読み取れるカテゴリを作成した。各カテゴリに含まれる単語の集合をクラスター分析によって得られた従属関係を考慮しながら、最終的に 4 つの大きなカテゴリに分類した。この 4 つのカテゴリは、①初心者向けの SAS の紹介に言及したカテゴリ、②SAS トレーニングについて言及したカテゴリ、③役立つ Tips やマクロについてのカテゴリ及び④SAS で実施したい具体的な要望に言及したカテゴリであった。各カテゴリに要望の詳細をみることができる。

#### 3. 2. 2. 3. 2 検討チームによるニーズ分析の結果と考察

一方、検討チームによるニーズの分析結果は以下の通りであった(表 4)。前述の方法で、ニーズを CP 部と DM 部でそれぞれまとめたものを見比べ、今後のトレーニングの実施をイメージして整理した。同時にそれらに関わる検討事項も確認した。当初、検討チームは CP 部と DM 部をまとめて考えようとしたが、回答に記載されていない背景情報(自明のこととして省略されている業務内容など)や解決したい問題などといった情報を踏まえた上で、また、全部の自由回答を見た上で、ニーズの内容が完全に分離していること、また、そのニーズが急ぎか急ぎでないかおよび業務上必須か必須でないかによる優先順位も考慮し、それぞれの部門に対する SAS トレーニングをわけて考えることとした。

表 4 検討チームによるニーズの分析結果

トレーニング内容	対象	根拠	検討事項
①初級 SAS トレーニング	両部門向け	SAS の機能の理解や操作で曖昧なところを整理してもらうため	教材の準備
②特定の SAS プログラムの解説	CP 部向け	特定の SAS コードがすでに存在するため	業務に有用なプログラムがあれば紹介してほしい、テクニカル・チップスなどの情報提供をしてほしいといった要望への対応
③具体的な問題に対しサンプル・プログラムをもとに解説	DM 部向け	ある処理を行うためのプログラムの組み方など、具体的なイメージがあるため	

### 3. 2. 2. 4 PROC VARCLUS を用いたクラスタリングと検討チームによるニーズ分析の比較と結論

ここでは、まず、PROC VARCLUS を用いた分析と検討チームによる分析のそれぞれの特徴について述べ、次に、この 2 つの分析結果を比較して得られた知見を述べる。最後に、この 2 つの分析結果の関係と PROC VARCLUS を用いた分析結果の有用性について述べる。

今回の PROC VARCLUS を用いたクラスタリングの特徴は、ひとつには CP 部と DM 部のニーズを区別しなかったことである。これは、はじめに CP 部と DM 部の部門間の差をカイ二乗検定にて確認し、その結果、一語を除いて差がなかったことを根拠とした。もうひとつは、得られたニーズのそれぞれ異なった意味の単語がクラスタリングされることで、それらが結果として 4 つのカテゴリに集約され、まとまりとしての意味が浮き彫りになったことである。

検討チームによる分析の特徴は、まず全部の自由回答を見た上で、CP 部と DM 部のニーズが完全に分離していると判断し、この 2 部門のニーズを分けて分析したこと。また、Solution の検討という当初の目的を念頭におき、ニーズの回答からトレーニングへの要望を理解し、その上でどのようなトレーニングを実施すべきかという観点から分析したことである。

PROC VARCLUS を用いたクラスタリングと検討チームによる分析の結果を比較し、以下のようなことがわかった。まず、分析の開始時のアプローチにおいて違いが見られた。クラスタリングでは CP 部と DM 部のニーズに現れる単語の頻度の差をカイ二乗検定を用いて確認し、「日付」を除いて差が見られなかったため CP 部と DM 部をまとめて分析することとした。反対に、検討チームによる分析では、当初両部門をまとめて分析しようとしたが両部門の業務背景、解決したい問題、優先順位などを考慮した上で部門によるニーズに差があると判断し、CP 部と DM 部への SAS トレーニングの内容を別々に検討した。この CP 部と DM 部をまとめる検討するか別々に検討するかという違いが生じた理由は、ひとつには検討チームによる分析では、回答に記載されていない業務内容、解決したい問題などといった背景情報に加え、ニーズから優先順位を判断しこれを分析に用いることでクラスタリングには含まれない情報を利用していたこと、また、クラスタリングに用いたニーズに係る単語はコーディングにより情報の縮約をおこなっていたことが、このような違いに繋がっていると考えられた。次に、分析結果として得られた内容の性格に違いがみられたが、その関係は、クラスタリングによる分析結果が検討チームによる分析結果をサポートするものであった。検討チームによる分析結果は、トレーニング実施に向け得られた情報を集約して分類した Solution に直結するものであり、分析結果に Solution とその対象部門、根拠および検討事項が含まれていた。一方で、PROC VARCLUS を用いた分析結果は、得られた情報をコンパクトにまとめたもので、分類により回答に含まれる内容を際立たせたものであった。この検討チームによる分析結果と、PROC VARCLUS を用いた分析結果との関係は、PROC VARCLUS を用いた分析結果が検討チームによる Solution をサポートする関係であるように思えた。そこで、これが正しい考察かどうかを確認した。この確認では、検討チームが分析結果として導いた SAS トレーニング 3 種と PROC VARCLUS を用いた分析結果と見比べて、大きな乖離がなく、さらに PROC VARCLUS を用いた分析の結果が検討チームによる分析結果をサポートしているかどうかを判断基準とした。確認の結果は、検討チームの分析結果の「①初級 SAS トレーニング」に対しては、PROC VARCLUS を用いた分析結果の「初心者へ向けについて」と「トレーニングについて」のニーズ・カテゴリが参考になると考えられた。同じく、「②特定の SAS プログラムの解説」と「③具体的な問題に対しサンプル・プログラムをもとに解説」に対しては、「SAS で実施したいこと」のニーズ・カテゴリにその内容を詳しくみることができた。また、検討チームによる検討項目の「業務に有用なプログラムやテクニカル・チップス」に関する部分に対し、「SAS Tips」として存在した。以上から検討チームの分析結果の詳細を PROC VARCLUS を用いた分析結果がサポートしていると判断した。



以上の考察から、分析の初期の段階で CP 部と DM 部をまとめて検討するか、別々に検討するかという違いがみられたが、これは分析に利用した背景情報の使用の有無、コーディングによる情報の縮約による違いがその理由として考えられた。また、分析結果は、PROC VARCLUS を用いた分析結果が、検討チームがニーズ分析から導いた Solution をサポートする関係にあることが確認できた。

#### 4 Solution としての SAS トレーニングとプログラム・コンテストの立案

Solution は以下の通り。

- ・ 3 種の SAS トレーニングの実施
  - ① 初級 SAS トレーニング
  - ② 特定の SAS プログラムの解説
  - ③ 具体的な問題に対しサンプル・プログラムをもとに解説

##### 4. 1 SAS トレーニング

Solution の①については、両部門の希望者を対象におこなうこととした。また、②と③のトレーニングについては、それぞれ CP 部と DM 部のメンバーをそれぞれ対象とすることとした。

##### 4. 2 プログラム・コンテスト

また、③に絡めて CP 部、DM 部向けトレーニングのサンプル・プログラムの募集を目的に、SAS 経験者によるプログラム・コンテストの実施を立案した。これは、特定のデータエラーを検出するための汎用性の高いプログラムを募集するもので、主に SAS の利用頻度の高い臨床解析部門のメンバーを対象とした。通常、臨床解析部門のメンバーが特定の題材に対してプログラム技術を競うことはないため、また、メンバーの他部門向け教材作成への参加と臨床解析部門のレクリエーションとしても良い機会なのでコンテストとして企画した。

#### 5 まとめと課題

近年の医薬品開発における EDC や CDISC などの普及で、データベースや CRF の標準化が浸透するにつれ SAS の利用方法や現在のニーズをあらためて把握する必要性が生じている。本稿では、製薬企業における医薬品開発部門において、SAS のプログラミングを主な業務としない者を対象に SAS の利用における潜在的なニーズがあると仮定し、そのニーズの掘り起こしと Solution を検討した。アンケートにより、SAS トレーニング希望の有無、SAS の習熟レベル、利用頻度及びニーズを調査した。調査により把握したニーズの分析には、SAS の PROC VARCLUS を用いたクラスター分析と検討チームによるニーズ分析を行い、この 2 つのニーズ分析の結果を比較した。当初、クラスタリング手法を用いたニーズの分析結果が果たして意味のある検討資料になるかという懸念もあった、しかしながらクラスター分析による結果は、直接的には Solution となるものではなかったものの、検討チームが導いた Solution を支持する内容のものであった。今回 PROC VARCLUS を用いた一連のクラスター分析の結果は、検討チームの分析による Solution を支持する有用な資料であった。Solution として、異なるニーズに対応した 3 種の SAS トレーニングの実施と、トレーニングで用いるサンプル・プログラムの募集を目的としたプログラム・コンテストを開催することとした。

今回は、SAS トレーニング実施のための現状の把握から Solution の立案までを記述した。SAS トレーニングおよびプログラム・コンテストは、現時点ではまだ実施しておらず、したがってその結果や評価を述べるできない。また、PROC VARCLUS を用いた分析に係るコーディングの部分で、抽出語の整理/まとめの工程にお

いて、かならずしも完全に人の主観を除いたとは確証できない部分があった。分析の初期の段階で CP 部と DM 部をまとめる検討するか、別々に検討するかという違いが生じた理由については、ニーズの回答に含まれない背景情報とコーディングによる情報の縮約の違いがその理由のひとつとして考えられた。したがって、クラスター分析でおこなうコーディングでの情報の縮約は、その結果の利用方法をふまえて十分な考慮と工夫が必要である。例えば、「知っておくと役立つ Tips の紹介」というニーズが CP 部の回答にみられた場合、コーディングで「SAS Tips」とするのか「業務に役立つ SAS Tips」とするのか或いは背景情報ではあるが CP 部という単語を付与して「CP 部の業務に役立つ SAS Tip」とするのかはその影響を十分に考慮する必要がある。この場合、追加する背景情報を別途データとして持っておくなどの工夫が考えられる。また、クラスター分析の結果の考察では、回答に含まれるに単語をクラスター分析にて階層的に分類した後に、一定の水準で再度分類を見直すことで、異なる性格の要望が浮き彫りになった。今回は一定の水準として、性格の似通っていると判断できるクラスターの集合で見直しをおこなったが、この水準の判断にも客観性を保つ配慮が必要である。これらについては引き続き評価および検討していければと思う。

全体として、今回の調査開始からトレーニング実施までのプロセスとその手法は、Solution を導くために適切であったと考えられる。この過程で、シックスシグマの DFACE に倣った VOC の分析に SAS のプロシジャのひとつである PROC VARCLUS を用いた。この分析結果が、検討チームが検討した Solution をサポートしていたことはひとつの発見であった。この一連のプロセスと手法は、医薬品開発での SAS トレーニング実施の検討に有用であるに留まらず、例えば、その他業界/業務における業務改善のためのタスク・チーム等による活動の参考になると考えられた。

## 参考文献

- 1) DFACE: Consulting – method, Toshiba-sigma.com ([http://www.toshiba-sigma.com/consul/3\\_4\\_2\\_dfacer.htm](http://www.toshiba-sigma.com/consul/3_4_2_dfacer.htm))
- 2) Ronen Feldman, James Sanger (2007), The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data, Cambridge University Press
- 3) 岸本 敦司 (1996), 変数のクラスタリング –PROC VARCLUS 再発見–, 日本 SAS ユーザー会論文集