

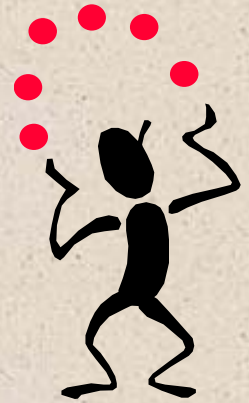
SAS未経験者を SAS内部構造を理解した DATAステッププログラマに 短期間で育成するカリキュラムの紹介

アストラゼネカ株式会社
臨床統計・プログラミング部

山田大志 小澤康彦 宮浦千香子

Aug, 1st, 2003

1. はじめに



- ◆ SAS内部構造を教えた動機
- ◆ セミナーで使ったSAS内部構造の説明資料
- ◆ 卒業試験問題例
- ◆ SAS内部構造の教育が実際の業務にどのような影響を与えたか

2. SASセミナーの開始

- ◆ 「臨床データ論理チェック用プログラムが作成できるSASプログラマを短期間で育成して欲しい」
- ◆ オリジナル教育カリキュラムを作成

3. 呪文の限界

◆ 受講者からの様々な質問に対して、その答えのみ教えた

✗ 応用問題に対応できない

◆ SAS言語を「呪文」のように教える限界

4. セミナー方針の見直し

◆ Program Data Vector (PDV)

SASシステム固有の様々な規則に従って、データを処理した結果や出力SAS データセットの変数についての情報を保持するための一時的なメモリ領域

- ◆ SAS内部構造を理解した上でDATAステップの仕組みを習得することが、応用力のあるSASプログラマになる近道

4. セミナー方針の見直し (追加カリキュラム)

- ◆ それまでのセミナー内容を反映したプログラム演習の実施
- ◆ 本来外側からは見ることができない、SASの内部構造を交えたDATA ステップの説明



*** 表2 演習問題 ***

(LIBRARY) ライブラリにあるデータセット [DEM] を読み込み、以下の処理を実行して下さい。

・以下の 2 変数を新規変数として設定してください。

変数名	型 / 長さ	ラベル
DAY	数値 / 8 バイト	Day number
COUNT	数値 / 8 バイト	OBS No.

・現在の各症例 1 オブザベーションの [DEM] データセットから、新規変数 [DAY] の値が「1」「8」「15」「22」となるような各症例 4 オブザベーションの [DEM_NEW] データセットを WORK ライブラリに作成してください。

出力変数 (順不同) : [SUBJECT] [PATIENT] [SEX] [BIRTHDAT] [DAY] [COUNT]

・新規変数 [COUNT] に全オブザベーションを通して連番をつけてください。

[DEM] データセット

	FNAME	LNAME	SUBJECT	PATIENT	BIRTHDAT	SEX
1	YASUHIKO	OZAWA	E0001004	101	1980-01-31	1
2	TAIJI	YAMADA	E0001010	102	1979-05-05	1



[DEM_NEW] データセット

	SUBJECT	PATIENT	BIRTHDAT	SEX	DAY	COUNT
1	E0001004	101	1980-01-31	1	1	1
2	E0001004	101	1980-01-31	1	8	2
3	E0001004	101	1980-01-31	1	15	3
4	E0001004	101	1980-01-31	1	22	4
5	E0001010	102	1979-05-05	1	1	5
6	E0001010	102	1979-05-05	1	8	6
7	E0001010	102	1979-05-05	1	15	7
8	E0001010	102	1979-05-05	1	22	8

*** 表2 演習問題 ***

- ❖ SASの経験が乏しい場合、自分の作成したプログラムがどのような結果を出力するのか明確ではないため、正しいプログラムを作成することは難しい
- ❖ SAS内部の動きを順に追ったPowerPointファイルを作成



*** 表2 演習問題（受講者の反応）***

- ◆ セミナー中はほとんど反応がなかった受講者も質問
 - ➡ これまでとは違った雰囲気
- ◆ スライドショーを見ながら説明を聞くことで、DATAステップの流れ、及びSAS内部構造に対する理解が深まった

*** データセット ***

【DEM】データセット

PATIENT	STUDY	CENTRE	SUBJECT
101	AAA	1	E0001004
102	AAA	1	E0001010

【DEM_NEW】データセット

	PATIENT	DAY	CNO
-1	101	1	1
-2	101	8	2
-3	101	15	3
-4	101	22	4
-1	102	1	5
-2	102	8	6
-3	102	15	7
-4	102	22	8



*** *Compile phase* ***

構文チェック

コンパイル (マシンコードへの変換)

ディスクリプタ情報の作成

“ *Input Buffer* ” の作成

入力ファイルが **SAS** データセット のみの場合は作成されない

“ *Program Data Vector* ” (*PDV*) の作成

宣言 (情報) ステートメントの情報を
Program Data Vector (*PDV*) に提供

*** *Program Data Vector* ***

```
data DEM_NEW ;  
  set DEM ;  
  do I = 1 to 4 ;  
    if I = 1 then DY = 1 ;  
    else DY = DY + 7 ;  
    COUNT + 1 ;  
  output ;  
end ;  
rename DY = DAY COUNT = CNO ;  
keep PATIENT DY COUNT ;  
run ;
```

PATIENT	STUDY	CENTRE	SUBJECT	I	DY	COUNT	_N_	_ERROR_

*** *Compile phase* ***

構文チェック

コンパイル (マシンコードへの変換)

ディスクリプタ情報の作成

“ *Input Buffer* ” の作成

入力ファイルが **SAS データセット** のみの場合は作成されない

“ *Program Data Vector* ” (*PDV*) の作成

宣言 (情報) ステートメントの情報を
Program Data Vector (*PDV*) に提供

*** Program Data Vector ***

```

data DEM_NEW ;
  set DEM ;
  do I = 1 to 4 ;
    if I = 1 then DY = 1 ;
    else DY = DY + 7 ;
    COUNT + 1 ;
  output ;
end ;
rename DY = DAY COUNT = CNO ;
keep PATIENT DY COUNT ;
run ;

```

PATIENT	STUDY	CENTRE	SUBJECT	I	DY	COUNT	_N_	_ERROR_
	Drop	Drop	Drop	Drop	↓ DAY	↓ CNO	Drop	Drop

*** Execution Phase ***

```

data DEM_NEW ;
  set DEM ;
  do I = 1 to 4 ;
    if I = 1 then DY = 1 ;
    else DY = DY + 7 ;
    COUNT + 1 ;
    output ;
  end ;
  rename DY = DAY COUNT = CNO ;
  keep PATIENT DY COUNT ;
run ;

```

宣言
ステートメント

PATIENT	STUDY	CENTRE	SUBJECT	I	DY	COUNT	_N_	_ERROR_
101	AAA	1	E0001004	1	1		1	0

Drop

Drop

Drop

Drop

DAY

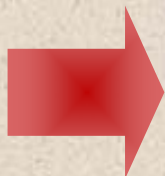
CNO

Drop

Drop

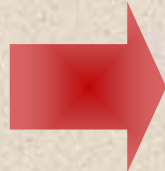

*** *DEM_NEW* データセット ***

PATIENT	DAY	CNO
101	1	1



*** *Execution Phase* ***

```
data DEM_NEW ;  
  set DEM ;  
  do I = 1 to 4 ;  
    if I = 1 then DY = 1 ;  
    else DY = DY + 7 ;  
    COUNT + 1 ;  
    output ;  
  end ;  
  rename DY = DAY COUNT = CNO ;  
  keep PATIENT DY COUNT ;  
run ;
```



PATIENT	STUDY	CENTRE	SUBJECT	I	DY	COUNT	_N_	_ERROR_
101	AAA	1	E0001004	2	1	1	1	0
	Drop	Drop	Drop	Drop	↓ DAY	↓ CNO	Drop	Drop

*** *Execution Phase* ***



```
data DEM_NEW ;
```

```
set DEM ;
```

```
do I = 1 to 4 ;
```

```
    if I = 1 then DY = 1 ;
```

```
        else DY = DY + 7 ;
```

```
    COUNT + 1 ;
```

```
    output ;
```

```
end ;
```

```
rename DY = DAY COUNT = CNO ;
```

```
keep PATIENT DY COUNT ;
```

```
run ;
```

PATIENT	STUDY	CENTRE	SUBJECT	I	DY	COUNT	_N_	_ERROR_
102	AAA	1	E0001010	.	.	4	2	0

Drop

Drop

Drop

Drop



DAY



CNO

Drop

Drop

*** *DEM_NEW* データセット ***

PATIENT	DAY	CNO
101	1	1
101	8	2
101	15	3
101	22	4
102	1	5
102	8	6
102	15	7
102	22	8

表2 演習問題（特に理解して欲しいこと）

- ◆ 宣言ステートメントの情報はCompile PhaseでProgram Data Vector (PDV) の変数に保持
- ◆ PDVのデータはOUTPUTステートメントでSAS データセットに出力
- ◆ 反復DOループはENDステートメントでカウントアップしてから条件判定
- ◆ DATAステップ内で新規に作成した変数はDATAステップループ開始ごとにPDV内の値を初期化する
- ◆ SETで読み込んだ変数、合計変数はDATAステップループ開始ごとにPDV内の値を初期化しない

5. 卒業試験

- ◆ 複数のSAS経験者に模擬的に試験を行い、試験の難易度を調整
- ◆ ほとんどの受講者が合格ラインを超えた

6. 卒業試験問題例

*** 表5 卒業試験問題 1 ***

次のプログラムを実行し、作成されるSASデータセット【TEST5】の変数[X]の値を記述して下さい。

```
data TEST5 ;  
    S = 1 ;  
    do X = 1 to 7 by 2 ;  
        S + 1 ;  
        X + S ;  
    end ;  
run ;
```

6. 卒業試験問題例

*** 表6 卒業試験問題 2 ***

次のプログラムを実行した結果はどうなるでしょう？ 選択肢の中から選んでください。但し、データセット【VIT】は変数【CTEMPORA】【WEIGHT】のみが存在することとします。

```
data TEST6 ;  
  set VIT ;  
  rename CTEMPORA = CTMP ;  
  format WEIGHT 8.1 ;  
  label CTMP = 'TEMP (C)' ;  
run ;
```

- A: フォーマットは設定されない
- B: ラベルは設定されない
- C: 全て問題なく設定される

7. まとめ (1)

◆ 受講者からの声 (その1)

予想とはちがう結果が出たときにSAS内部構造を思い出すとエラーの原因が容易に分かった

◆ 講師陣の見解

- ◆ SASの内部構造だけを学んでも、それを実践に生かせなければ意味はない
- ◆ SASの内部構造を理解することは非常に重要であり、また、結果的に実際の業務において非常に有用

7. まとめ (2)

◆ 受講者からの声 (その2)

最初から業務に必要なことのみ教えてくれれば
すぐに作業ができるようになったのでは

◆ 講師陣の見解

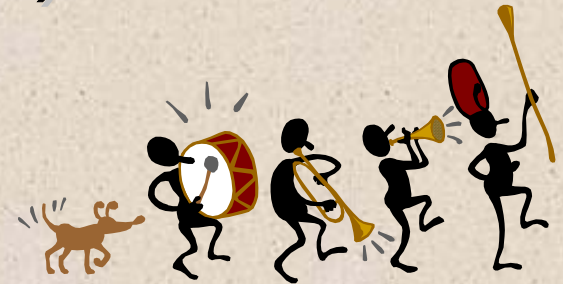
◆ SASの内部構造がブラックボックスでは・・・

- ✗ 実力を上げるために多くの「呪文」を覚える必要
- ✗ 経験という名の恐ろしく時間がかかる方法でしか成長できない

◆ SAS内部構造を学べば、SAS言語を幾つかのパターンに分類でき、その後の自主学習もスムーズ

7. まとめ (3)

◆ 受講者からの声 (その3)
SASに対して興味を持った



◆ 講師陣の見解

◆ SAS内部構造を学ぶ喜びを知ることが高いレベル
のSASプログラマに成長する近道

◆ 最後に・・・

SAS経験者にもSAS内部構造の理解は有用

参考文献

1. SAS言語: 解説編 バージョン8 (2003), SAS出版局
2. SASランゲージ リファレンス
Version 6 First Edition (1995), SAS出版局
3. Step-by-Step Programming with Base SAS®
Software (2001), SAS Institute Inc.
4. Base SASソフトウェア使用法ガイド
Version 6 First Edition (1993), SAS出版局
5. Robert Virgile, An Array of Challenges
- Test your SAS® Skills (1996), SAS Institute Inc.