

SASによるデータベース利用法：医療用データベースの経験から

竹下, 節子
九州歯科大学内科

廣田, 安夫
九州歯科大学内科

<https://doi.org/10.15017/1474923>

出版情報：九州大学大型計算機センター広報. 15 (1), pp.3-10, 1982-03-10. 九州大学大型計算機センター
バージョン：
権利関係：

SASによるデータベース利用法 －医療用データベースの経験から－

竹下 節子*, 廣田 安夫*

1. はじめに

疫学調査データ, 臨床的研究データ等を, 総合的に解析するために, 汎用統計パッケージSASを使ってデータベースを作った[1].

SASはDBMSとして作られたソフトウェアではないが, データベースを維持管理するのに必要な機能を備えていると考えられるので, 私共の経験を紹介してユーザーの便宜を計るのが本稿の目的である.

2.1 データベース作成上の機能

SASのプログラムは, 基本的にはDATAステップとPROCステップの2つのステップの組の列から成る. 2つのステップに含まれる機能は, 図1のように非常に多いが, データベース作成上からは, DATAステップの機能が重要である.

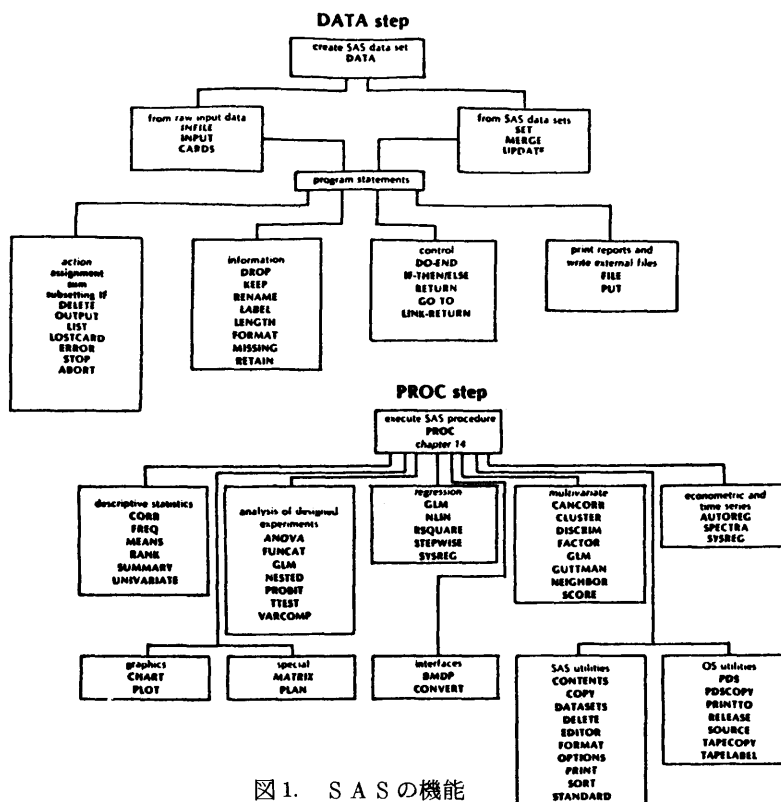


図1. SASの機能

* 九州歯科大学内科

2.1.1 生データの入力方法

カード、ターミナルからの入力およびすでに磁気テープやディスクにあるOSファイルからの入力が可能であるが、SASで有利なのは、入力の形式がリスト形式、フォーマット形式等、多様な入力形式があり、SASのために改めてデータを作る必要がない点である。

2.1.2 SASデータセットの作成

DATAステップには、SASデータセット*に加工をして新たなデータセットを作ったり、複数のデータセットを連結したり、MERGEしたり、UPDATEする機能等が豊富で(図2)、データベースを作成するときには、これらの機能を繰返しながら、次第に完全なものに作り上げることができる。

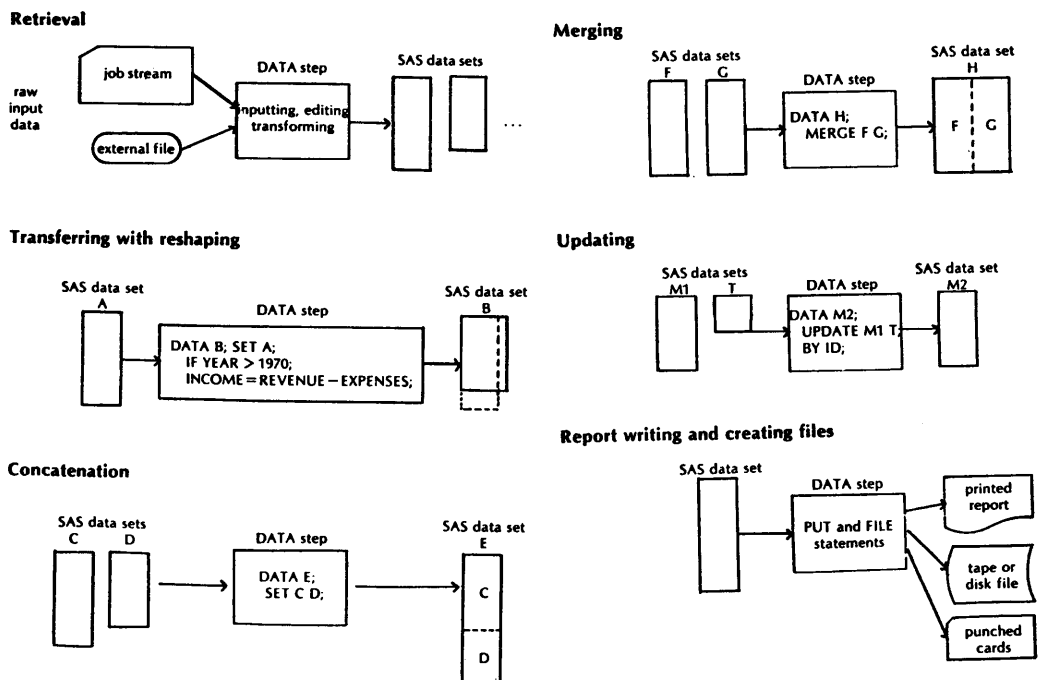


図2. DATAステップの諸機能

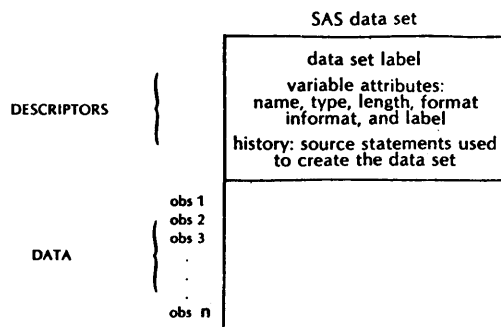


図3. SASデータセットの構造

* SASで作られたデータセットで、図3のような構造を持つ。以下データセットと言う。

2.1.3 変数および観測値^{*}の選択

解析に必要な変数の選択,あるいは観測値の選択は,無駄な処理を省き効率の向上に繋る.図4のFIRSTOBS,OBSは,データセットから取出す観測値の範囲を,KEEPは取出す変数を示す.作成するデータセットに含む変数の指定にも使える.KEEPと反対の役割をもつものにDROPもある.DELETEは観測値を消去する.

```

19? DATA;
20? SET A.HIHSAS5<READ=① FIRSTOBS=1 OBS=200 KEEP=ID NAME1 NAME2 SEX
21? BIRTH DEATH S37U5 S37U2>;
22? IF S37U2=. THEN DELETE;
23? PROC PRINT;④
NOTE: DATA SET WORK.DATA4 HAS 42 OBSERVATIONS AND 8 VARIABLES. 296 OBS<TRK>
24? RUN;⑧

```

OBS	ID	S37U2	S37U5	NAME1	NAME2	SEX	BIRTH	DEATH
1	4	3370517	60			1	1340827	420522

- | | |
|---|-----------------|
| ① データセットA・HIHSAS5より取りだすオブザーベーションの先頭を指定する。 | ④ オブザーベーションの消去 |
| ② 取りだすオブザーベーションの最後を指定する。 | ⑤ SASのもつ出力プロシジャ |
| ③ 取りだす変数名の指定 | ⑥ ⑤の出力結果の一部 |
| | ⑦ SASのメッセージ |
| | ⑧ ⑤の実行 |

図4. 変数,観測値の選択

2.1.4 変数の長さの変更

SASの変数は,数値変数と文字変数の2種類がある.文字変数は,1~200バイトの指定が可能で,住所,名前などに利用できる.数値変数は,DATAステップで処理をされる場合は,通常8バイトである.データの性質により8バイトを必要としないときには,バイト数の少ない方が,記憶域も少なくてすみ,効率的でもある.図5のLENGTHは,変数の長さの変更を行うSAS文である.整数値の場合には,図6のように2~8バイトの変更が可能である.

```

39? DATA C;
40? SET A.SASBP;
41? LENGTH KGAN1-KGAN16 AGE KAGE1-KAGE16 2;
42? LENGTH KWEI1-KWEI16 KHEI1-KWEI16 3;①
43? LENGTH KDATE1-KDATE16 4;
44? RUN;
NOTE: DATA SET WORK.C HAS 1003 OBSERVATIONS AND

```

- ① KGAN1などの変数の長さを2バイトに変更する。

図5. 変数の長さの変更

Length	Significant Digits Retained	Largest Integer Represented Exactly
2	2	255
3	4	65,535
4	7	16,777,215
5	9	4,294,967,295
6	12	1,099,511,627,775
7	14	281,474,946,710,655
8	16	72,057,594,037,927,935

図6. 整数値のバイト長,有効桁と最大値

*オブザーベーションのこと

2.1.5 自動的な欠測値

医療データには欠測値が多く、そのコーディングは、労力のいるものである。SASには自動的な欠測値が用意されているので、入力データ上はブランクのまゝでよい。数値変数は、(ピリオド)、文字変数は、_ (ブランク)が入る。新たにデータを作成する場合には、特に便利かと思われる。

以上、データベース作成上からSASの機能を見た。2.1.1, 2.1.2は当然のことであるが、2.1.3の選択機能、2.1.4の変数の長さの変更機能は、できるだけ大量の情報を少ないスペースに格納することを可能にする。2.1.5の欠測値以外に、特定の欠測値を作ることとはもちろん可能である。このようにSASにはデータ作成側の労力を省くことが考慮されている点がありがたい。

3. 編集機能EDITOR

できあがったデータベースの内容の修正、変更、追加の機能にはEDITORがある。これは、データセットに直接にアクセスし、変数名によって内容を取扱う。データセットの記述部(図3のDESCRIPITORS)の変更ができないので、変数名の追加はできないが、観測値の追加は、データセットの末尾にその観測値をおくことによりできる。直接アクセスのため、一旦変更されると復元はできない。観測値の消去は、そのすべての変数の内容を欠測値にすることであるために、多数の観測値を消去しても観測値の総数、したがってデータセットの大きさが小さくなる訳ではない点に注意すべきである。図7はEDITORの例である。データベースを直接に編集できる機能を備えている点は、プログラムによる編集あるいはUPDATEによる編集などと異り、効率的であり、SASがDBMSとして使える1つの利点である。

```

2? PROC EDITOR DATA=A.TEST(READ=TK PROTECT=HISA);
NOTE: SAS STOPPED PROCESSING THIS STEP BECAUSE OF ERRORS.
3? RUN; ②
ENTER COMMANDS OR 'HELP';
4? HELP; ③
SAS EDITOR: FOR DIRECT INTERACTIVE EDITING OF SAS DATA SETS.
ALL COMMANDS MUST BE FOLLOWED WITH A SEMICOLON (;).
THERE ARE 10 COMMANDS (ALL SIMILAR TO THE TSO TEXT EDITOR).
LIST      REPLACE FIND      DELETE
END        HELP      ADD      VERIFY
DOWN      UP

ENTER 'HELP' AND ONE OF THE COMMANDS FOR A DESCRIPTION
OF THAT COMMAND. ENTER 'HELP OTHER;' FOR OTHER HELP.
5? L 20; ④
20 ID=100 _TYPE_=1 _FREQ_=10 HEIGHT=148.93 WEIGHT=52.6 CHOL=148.75
SBP=193.6667 DBP=91.2 HEISE=0.6363175 WEISE=1.10755 CHOLSE=24.43267
SBPSE=6.329573 DBPSE=4.664338 NAME1= NAME2= SEX=1 AGE=76
FIRST=3360428 LAST=3531003
6? REP 20 HEIGHT=148;
7? VERIFY ON; ⑤
8? L 20; ⑥
20 ID=100 _TYPE_=1 _FREQ_=10 HEIGHT=149 WEIGHT=52.6 CHOL=148.75
SBP=193.6667 DBP=91.2 HEISE=0.6363175 WEISE=1.10755 CHOLSE=24.43267
SBPSE=6.329573 DBPSE=4.664338 NAME1= NAME2= SEX=1 AGE=76
FIRST=3360428 LAST=3531003
9? ADD ID=990 HEIGHT=170;
21 ID=990 HEIGHT=170 ⑦
10? L 21; ⑧
21 ID=990 _TYPE_= . _FREQ_= . HEIGHT=170 WEIGHT= . CHOL= . SBP= . DBP= . HEISE= .
WEISE= . CHOLSE= . SBPSE= . DBPSE= . NAME1= NAME2= SEX= . AGE= . FIRST= . LAST= .

```

```

11? DELETE 20; ⑨
12? 20; ⑩
20 ID=, _TYPE=, _FREQ=, HEIGHT=, WEIGHT=, CHOL=, SBP=, DBP=, HEISE=,
WEISE=, CHOLSE=, SBPSE=, DBPSE=, NAME1= NAME2= SEX=, AGE=, FIRST=, LAST=,

107? FIND ALL 1, LAST ID=50;
FOUND AT OBS=154 ⑪
FOUND AT OBS=155
FOUND AT OBS=156
FOUND AT OBS=157
FOUND AT OBS=158
FOUND AT OBS=159
FOUND AT OBS=160
FOUND AT OBS=161
FOUND AT OBS=162
NOT FOUND. OBS=451

6? VERIFY ON; ⑬
7? FIND ALL 1, LAST ID=50;
NOT FOUND. OBS=22
8?L 22;
22 ID=93 NAME1= NAME2= SEX=2 BIRTH=1390403 DEATH=0 DATE=3530906
AGE=72 HEIGHT=149.7 WEIGHT=38.5 ECG=1 CHOL=251 URINEP=0 SBP=120 DBP=77.33333
YEAR=53 GANKEN=9
9? REP 22 NAME1=NODA NAME2=MOTO; ⑬
22 NAME1= NAME2=
22 NAME1=NODA NAME2=MOTO

```

- ① EDITOR コマンドの入力
- ② 起動
- ③ SAS EDITORに含まれるコマンドの一覧表の出力
- ④ LIST 20;の意味。20番目のオブザーベーションの出力
- ⑤ 内容が変化したときに変化した変数の値を出力する。
- ⑥ オブザーベーションの追加。最後のオブザーベーションの後に追加される。
- ⑦ ⑥で追加された内容
- ⑧ 追加されたオブザーベーションの出力

- ⑥で指定のない変数には、missing value が入る。
- ⑨ オブザーベーションの消去
- ⑩ ⑨で消去したオブザーベーションは、すべての変数の値がmissing value にかわるが、オブザーベーションとしては、のっている。
- ⑪ データセット全体でID=50のものをさがして、すべてを出力する。
- ⑫ } オブザーベーションの内容を変更する。
- ⑬ } ⑬では、⑬のために変更の前後の変数の値を出力する。

4. データの保護

データベースの保護は、OS上のものと、SASのものと2種類がある。他の課題番号からのファイル消去に対する保護、OLD指定による無資格者の使用からの保護などは、OS上の保護である。SASの持つ保護機能としては、読出し保護（READパスワード指定）、消去保護（PROTECTパスワード指定）の2つがある。これらは、DATAステップでデータセットを作成するときに指定する。図8がその例である。保護をつけて作成したデータセットは、パスワードが一致しなければ使用できない。

```

READY
ALLOC DA(TAKE.DATA) F(0) OLD ①
3? DATA D.TEST(READ=TK PROTECT=HISA); ②
NOTE: DATA SET WORK.A HAS 20 OBSERVATIONS AND 19 VARIABLES. 108 OBS/TRK.
4? SET A;
5? PROC EDITOR DATA=D.TEST; ③
NOTE: DATA SET D.TEST HAS 20 OBSERVATIONS AND 19 VARIABLES. 108 OBS/TRK.
5 PROC EDITOR DATA=D.TEST;
-----
202 } ④
202: DATA SET IS PROTECTED BUT PROTECT= KEYWORD WAS NOT SUPPLIED.

9? PROC EDITOR DATA=D.TEST(PROTECT=HISA); ⑤

9 PROC EDITOR DATA=D.TEST(PROTECT=XXXX);
-----
163 } ⑥
163: DATA SET ACCESS IS LIMITED BY READ= PASSWORD.
CHECK WITH CREATOR OF DATA SET FOR PASSWORD TO USE.
10? PROC EDITOR DATA=D.TEST(READ=TK PROTECT=HISA); ⑦
NOTE: SAS STOPPED PROCESSING THIS STEP BECAUSE OF ERRORS.

11? RUN; ⑧
ENTER COMMANDS OR 'HELP:'

```

- ① SAS データセットを含むファイルの割り付け
- ② D・TEST に読み取り保護のために READ のパスワードと消去保護のために PROTECT のパスワードをつける。
- ③ D・TEST のパスワードを指定せずに EDITOR で使おうとする。
- ④ ③ に対するエラーメッセージ、パスワードがないため。
- ⑤ D・TEST の PROTECT のパスワードを指定して EDITOR で使おうとする。
- ⑥ ⑤ に対するエラーメッセージ
- ⑦ } EDITOR の実行
- ⑧

図8. READパスワード、PROTECTパスワードの例

パスワードの出力方法はないので、作成者は絶対これを忘れてはならない。これらの機能により、データベースを使う資格のない者から読まれることはなく、管理者以外は、消去できない。前述した EDITOR を使用するときには、消去のパスワードも必要であるから、勝手に内容を書替えられることはない。これらの機能は、データベースを維持管理してゆく場合には、なくてはならぬものである。

5. SAS ユーティリティ

EDITOR 以外に、SAS ユーティリティとして 9 個のプログラムがある。ここでは、特に便利であると思われる 2, 3 をあげる。

5.1 SORT

PROC ステップのプログラムを使う場合、あるいは、データセットの MERGE をする場合等に、キーとなる変数の値によって分類する必要がある。SORT はこの分類をする。必要なワークファ

イルの指定は不要なので、簡単であり使いやすい。しかし、SASを起動する前に、SASSORTコマンドを入力しておく必要がある。

5.2 DATASETS

1つのOSファイルの中にSASデータセットを多数格納しているとき、すべてのデータセット名、それぞれに含まれる観測値の数、占有するトラック数、ファイルの中の使用トラック数等の情報を出力する。セッション中に作成したWORKの中のデータセットについても同様である。

5.3 CONTENTS

1つのデータセットに含まれる情報—データセットラベル、変数の属性、履歴（データセットが作られるときに使用されたプログラム文）等が出力される。READ, PROTECTのパスワードの指定のある場合には、パスワードそのものの出力はない。

この他に、DELETE, FORMAT, COPY, PRINT等がある。

6. SASの問題点

6.1 マクロ文

一まとまりのプログラム文に名前をつけて、その名前の引用で代行する方法としてMACRO文がある。図9は簡単なMACRO文の例である。MACRO文は、定義後は、SASプログラムのどこでも引用できる。あらかじめ別の順編成ファイルにMACRO文を作り、使用時にSYSINファイルの先頭に連結するという方法をとれば、後日も使用できる。これは現在のMACRO文による定義の方法では、パラメータ渡しができないのでそれを緩和する1つの方法であろうが、必要に応じて、MACRO文の一部をTSS EDITORによって変更し使用するのは不便であり、同じパターンのMACRO文を多数用意しておかねばならない結果になる。

```

OSIU/F4 E40 U12L02 <<< JCL STATEMENTS LIST >>>
1 //F M JOB CLASS=A,MSGCLASS=R JOB 2528
2 // EXEC SAS 00000020
18 //D DD DSN=F .TAKE.DATA,DISP=OLD 00000025
19 //SYSIN DD DSN=F .TKMACRO.CNTL,DISP=SHR ① 00000030
20 // DD * 00000040

1 STATISTICAL ANALYSIS SYSI
NOTE: THE JOB F M HAS BEEN RUN UNDER RELEASE 79.3B OF SAS AT KYUSHU UNIVERSITY.

1 MACRO HYPER65 (SBP)=160 OR DBP>=95 & AGE>=65 % 00000010
2 DATA H; 00000050
3 SET D.RESIII T; ② 00000060
4 IF HYPER65; ④ 00000070
5 KEEP ID NAME1 NAME2 SEX AGE CHOL SBP DBP; 00000080
6 00000090

NOTE: DATA SET WORK.A HAS 20 OBSERVATIONS AND 8 VARIABLES. 264 OBS/TRK.
NOTE: THE DATA STATEMENT USED 0.13 SECONDS AND 140K.

```

- ① MACRO文のあるSequential fileを
SYSIN fileの直前におく。
- ② MACRO文の定義
- ③ MACRO名
- ④ MACRO文の引用

図9. マクロ文の使用例

6.2 TSS使用上の問題点

SASをTSSで使用する時、特に問題になるのが、ステップの途中でエラーのあるときの処理である。エラーのあるステップを強制的に終了するのは、すっきりしている反面、利用者にとっては、再度ステップの先頭から入力しなければならず、不便である。この点は、センターで用意されたSAS-RUN等のコマンドを利用することにより幾分か軽減されている。SASの今後のバージョンアップでは、これらの点の改良を望みたい。

7. ま と め

医学的データ解析のためにSASを使っているが、DBMSとしての観点からSASを取りあげた。データベース作成のための柔軟な機能、編集機能、保護機能等を備えている点を考慮すると、SASはDBMSとしても十分使えることがわかる。SASのデータ構造は、表形式であり、簡単な関係モデル様のものとも言われており、DBMSとしての使い方がそれほど難しいはずではないと思われる。また、我々の身近なところで使用できるDBMSがあれば、紹介していただけると幸いである。

参考文献

1. 廣田，竹下他，長期追跡研究用データベースの試作（第3報）SASの利用，第9回日本行動計量学会予稿集，1981，100-101.
2. 武富，大賀，平野，石田 センターにおける統計解析プログラムパッケージSASへの案内を兼ねて一，九大大型計算機センター広報，14，1，1981，6-30.
3. 武富，大賀，平野，石田，景川 統計解析システムSAS概説（1），九大大型計算機センター広報，14，2，1981，149-171.
4. 武富，大賀，石田，永井，景川 統計解析システムSAS概説（2），九大大型計算機センター広報，14，3，1981，366-387.
5. 武富，大賀，石田，永井，遠矢，景川 統計解析システムSAS概説（3），九大大型計算機センター広報，14，4，1981，491-514.
6. SAS User's Guide 1979 Edition, SAS Institute Inc.
7. SAS Views 1980 Edition, SAS Institute Inc.
8. SAS Applications Guide 1980 Edition, SAS Institute Inc.