

製造販売後調査のデータ解析で 受領するデータ形式について

山内 浩嗣
イーピーエス株式会社 統計解析2部

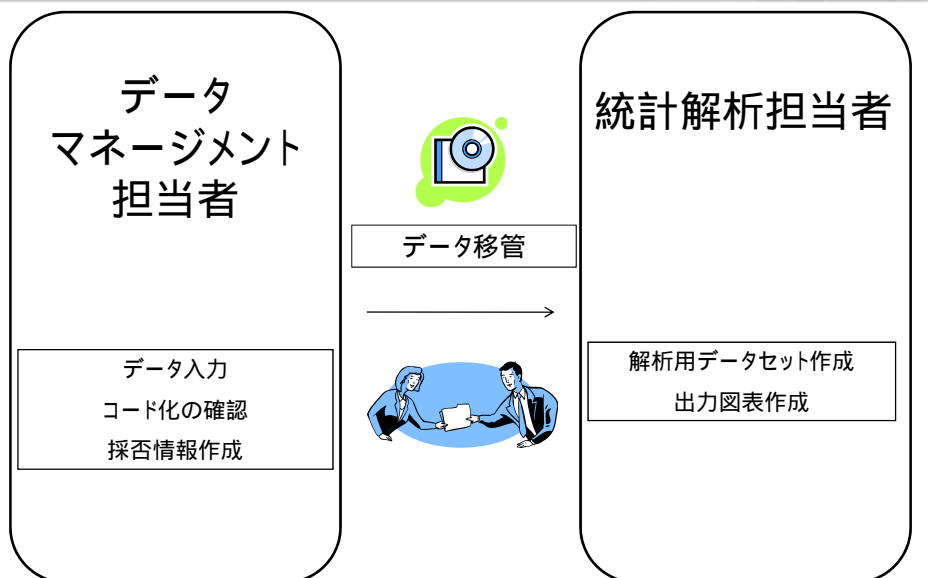
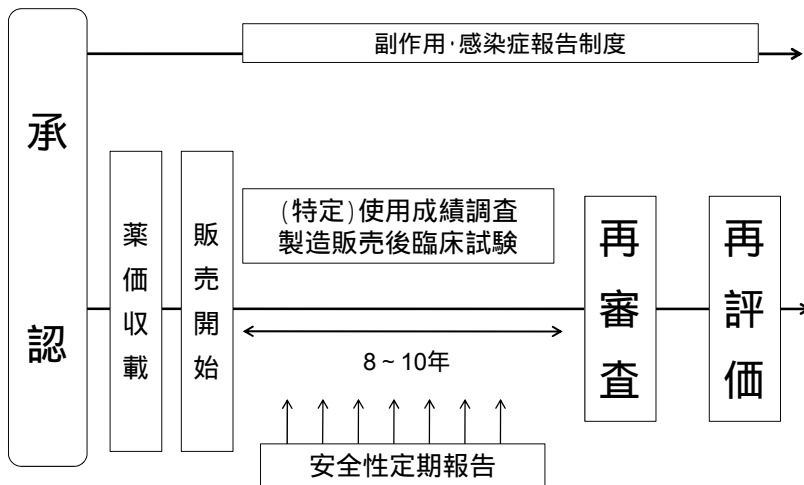
About the data format received in the data analysis of the PMS investigation

Kouji Yamauchi
Statistics Analysis Department 2, EPS Corporation

要旨:

製造販売後調査においては、データマネージメント担当者から解析担当者へのデータ移管の際に様々な移管形式が利用されている。特に外部データ(症例及びデータの取り扱い方法をまとめたデータ)については、SASデータセット形式だけでなくmdb形式、xls形式、csv形式など、様々な形式のデータが利用されている。これらのデータのSASへの取り込み方法について紹介するとともに、外部データ構造の標準化を目指す上で留意すべき内容についてまとめた。

キーワード: 移管データの標準化、データのSASへの取り込み



データの種類

調査票データ: データベースより取得

→ データベース変更が無い限りデータ構造変化無

外部データ(症例及びデータの取扱い方法をまとめたデータ):

解析毎にデータの変更、手作業による作成

→ データ構造の変化有

データ移管→解析業務

効率よく作業を開始するためのデータ構造の標準化

移管データの形式

SAS



アクセスファイル



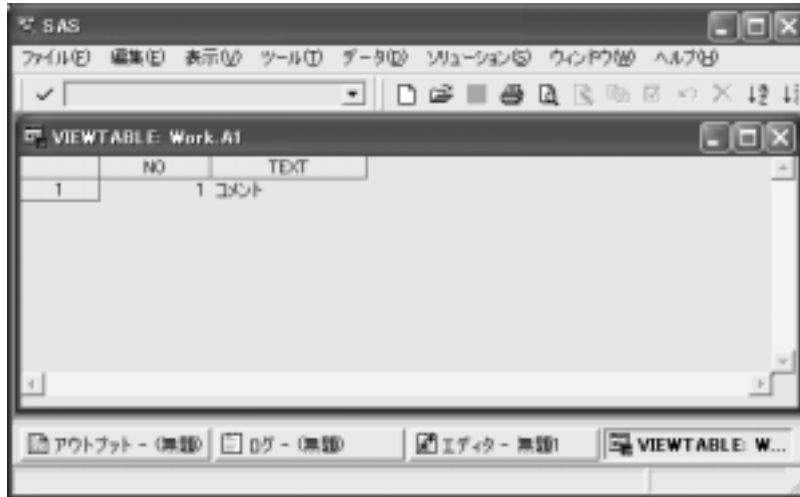
エクセルファイル



テキストファイル



SASデータ



7

SASデータ

データの特徴:

変換作業が不要

デメリット:

ライセンス数の問題

プログラムスキル

8

ACCESSファイル



9

ACCESSファイル

データの特徴:

- リレーショナルデータベース
- SASへ変数名や属性が原則継承可
- テキスト型にフィールドサイズの設定が可

デメリット:

- ライセンス数の問題
- 変数名に全角文字の使用可
- フィールドサイズのカウント方法

10

EXCELファイル



11

EXCELファイル

データの特徴:

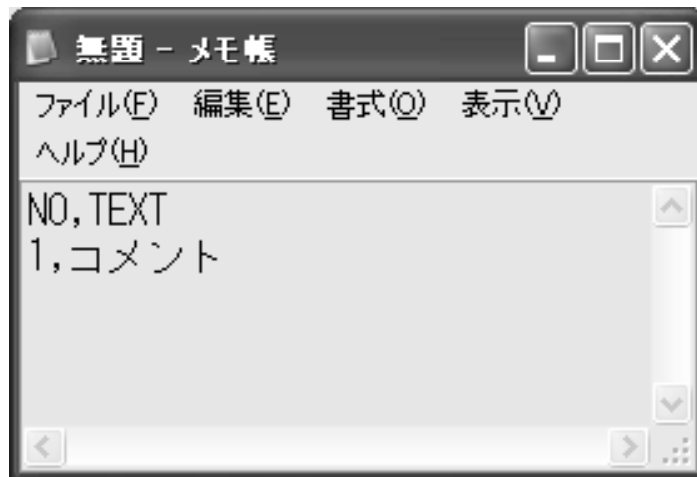
- 表計算ソフト
- 作成の容易性

デメリット:

- セルの書式に依存
- 変数名に自由度が多い

12

テキストファイル



13

テキストファイル

データの特徴:

ファイルサイズが小さい

デメリット:

SASへの取り込みが煩雑

データの作成ルールが必要

14

各ファイル形式のまとめ

取り込み方法	特徴	問題点	SASへの取り込み方法
SASデータ	変換が不要	ライセンス数の問題 プログラムスキル	LIBRARY
ACCESSファイル	リレーショナルデータベース SASへ変数名や属性が原則継承可	ライセンス数の問題 変数名に全角文字の使用可	LIBNAME,SQL,DDE, IMPORT
EXCELファイル	表形式 作成の容易性	セルの書式に依存	LIBNAME,SQL,DDE, IMPORT
テキストファイル	ファイルサイズが小さい	SASへの取り込みが煩雑	INFILE,IMPORT,DDE

15

各ファイル形式のまとめ

解析担当者の立場から・・・

移管データ形式：SASデータもしくはACCESSファイル形式を推奨

メリット：変数名、属性が継承できる

デメリット：ライセンス数の問題

16

各ファイル形式のまとめ

データマネージメントの立場から・・・

移管データ形式:

調査票データ: SASデータもしくはACCESSファイル形式

外部データ: EXCELファイル形式

EXCELデータの読み込み



データの取り込み



データの取り込み方法

LIBNAMEステートメント

SQLプロシジャ

DDE機能

IMPORTプロシジャ

LIBNAMEステートメント

プログラム例:

```
LIBNAME [ライブラリ名] "[リンク先のアドレス及びファイル名]";  
LIBNAME [ライブラリ名] ODBC REQUIRED=  
    "DRIVER={Microsoft Excel Driver (*.xls, *.xlsx, *.xlsm, *.xlsb)};  
    DBQ=[リンク先のアドレス及びファイル名]";
```

特徴:

リンク先のアドレス及びファイル名の指定

デメリット:

セルの書式に依存

19

SQLプロシジャ

プログラム例:

```
PROC SQL;  
CONNECT TO ODBC (NOPROMPT=  
    "DRIVER=MICROSOFT EXCEL DRIVER (*.xls, *.xlsx, *.xlsm, *.xlsb);  
    DBQ=[リンク先のアドレス及びファイル名]" DBMAX_TEXT=32767);  
CREATE TABLE [出力データセット] AS SELECT * FROM CONNECTION  
    TO ODBC (SELECT * FROM "[シート名]");  
DISCONNECT FROM ODBC;  
QUIT;
```

特徴:

多彩なSASプロシジャ

デメリット:

文法の理解が必要

20

DDE機能

プログラム例:

```
FILENAME [出力データセット1] DDE "EXCEL |  
  [[ファイル名]][シート名]![データ範囲]";  
DATA [出力データセット2];  
  INFILE [出力データセット1] NOTAB DLM='09'x DSD MISSOEVER;  
  FORMAT _ALL_ ;  
  INFORMAT _ALL_ ;  
  LENGTH [変数名];  
  INPUT [変数名];  
  
  KEEP [変数名];  
RUN;
```

21

DDE機能

特徴:

セルの書式に関係なく取得可

デメリット:

取得先のファイルを開く必要性

取得先のデータの範囲の指定

非表示セルの読み込み不可

22

IMPORT プロシジャ

プログラム例:

```
PROC IMPORT OUT= [出力データセット]
  DATAFILE="[リンク先のアドレス及びファイル名]"
  DBMS=EXCEL REPLACE;
  SHEET="[シート名]";
RUN;
```

特徴:

一般的なSASプロシジャ

デメリット:

セルの書式に依存

EXCELデータの読み込み

データ形式	特徴	問題点
LIBNAME ステートメント	リンク先のアドレス及びファイル名の指定	セルの書式が統一されていない場合読み込みが困難
SQL プロシジャ	多彩なSAS procedure 取り込みからデータ加工、集計まで一括 実施が可能	文法の理解が必要
DDE機能	セルの書式に関係なくデータの取得可能	取得先のファイルを開く必要有 取得する範囲の指定が必要
IMPORT プロシジャ	一般的なSAS Procedure	セルの書式が統一されていない場合読み込みが困難

EXCELデータの読み込み

IMPORTプロシジャを推奨

メリット:セルの書式に関係なく読み込める(プログラムによるが)

デメリット:文法の理解が必要

変数名の設定が必要

念のため・・・

二人で異なる方法で読み込む → 読み込みエラーの抑制

25

まとめ

移管データ形式: SASデータもしくはACCESSファイル形式を推奨

外部データについて:

EXCELファイルでも移管も可能(一部制約を含む)

セル書式の統一、入力方法の統一

データ構造の標準化 → 作業の効率化、品質の向上に貢献

26