

千葉大学総合情報処理センターニュース

平成8年7・8月発行
千葉大学総合情報処理センター

データ解析用のソフトウェアについて

— SPSS を中心として —

総合情報処理センター講師 橋本明浩

センターで利用できるアプリケーションサーバ(以下 cuaps)で稼働するデータ解析用のアプリケーションソフトウェア SPSS を中心にして簡単に説明します。データ解析はデータを見ながら、構造をみつけたり、自分の仮説を検証したりしていくことが重要です。あらかじめ目的を定めた、データ解析用の複雑なプログラムを作成することは現実的ではなく、実績のあるアプリケーションソフトウェアを利用することを推奨します。他方、実績のある統合型アプリケーションソフトウェアは、非常に高価であり、多くの計算機資源を必要とし、各研究室で利用する人の数だけ購入することは非効率的ですので、cupas の利用をおすすめします。そこにはデータ解析用アプリケーションソフトウェアとして、用途別に Wingz, SPSS, SAS の3つを準備しています。詳細は <http://cuaps.ipc.chiba-u.ac.jp> を参照して下さい。

各アプリケーションソフトウェアの概要

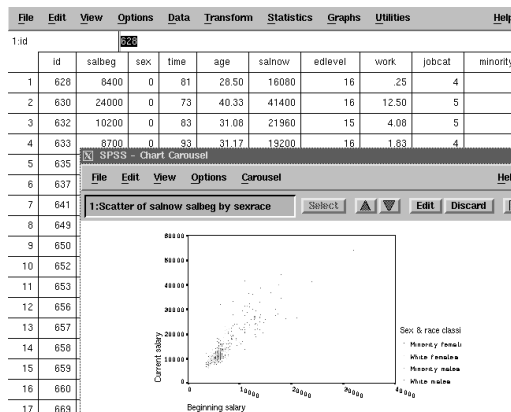
Wingz 表計算及び簡単なグラフ作成機能を有するソフトウェア。操作は極めて簡単である。起動は cuaps で wingz と入力し、終了はファイルメニューで終了を選択すればよい。

SPSS 表計算及びグラフ作成機能に加えて、カイ2乗検定、t,F検定、重回帰分析、因子分析、ANOVA,MANOVAなどの統計的な手法が利用できるソフトウェア。日本語対応はしていないが、操作はウィンドウですべて行うことができる。起動は spss とコマンドを投入し、終了はファイルメニューで終了を選択すればよい。

SAS 多彩な統計手法を含む統合型情報解析ソフトウェア。グラフィック機能も充実しており、簡単に利用できる SAS/ASSIST も含まれる。日本語に一部対応している。

SPSS の基本操作

ウィンドウの種類 SPSS の起動後、出力ウィンドウ(Output Window)、データウィンドウ(Data Window)の2つのウィンドウが現れます。その他に重要なウィンドウとその機能および保存時の拡張子の表を下に示しました。



ウィンドウ	説明・機能	拡張子
出力	はじめに現れるウィンドウで操作結果が表示されます	*.lst
データ	2番目に現れるウィンドウ。変数と各値が表示され、このウィンドウの操作バーで操作を選択します。操作の履歴は spss.jnl に保存されます。(標準)	*.sav
シンタックス	プログラム入力用。データウィンドウのメニューで File → Open → SPSS syntax で現れます。	*.sps
グラフィック	データウィンドウのメニューで Graph → で現れます。	*.cht

アスキー(生)データの読み込み エディタで作成した数値、文字だけのデータはアスキーファイルと呼びます。(このような生のデータを含んだファイルに対して拡張子 `*.dat` を使う取り決めになっています。)アスキーファイルはそのままで利用できません。数値、文字についての情報を含んだファイル形式に直す必要があります。この形式をSPSSでは、データファイルと呼びます。

具体的な例で変換方法を示します。番号、氏名、性別、身長、体重の情報が記述してあるアスキーファイルとして `my.dat` があると仮定します。(----+---- はカラムを示す為に便宜的に記述してあるだけで、実際は入っていません。)

```
cuaps% cat my.dat      (my.dat の内容)
----+----|----+----|----+----|
1200 Suzuki Takeshi M 168  56
1201 Miyano Teruo   M 170  56
.....
```

各個人に対応する並びをオブザベーション(観測値)と呼びます。各観測値はこの例では、1行のデータ(これをレコード)を有しています。レコードの中身は、1-4カラム目が番号、6-19カラムが名前、21カラム目が性別、...身長、体重と入っています。これをフィールドまたは変数と呼び、名前をつけて区別します。ここでは、番号にID、名前にはNAME、性別にはSEX、身長にはHEIGHT、体重にはWEIGHTという名前を与えることにします。名前の規則は以下の通りです。

- 利用できる文字は英数字、記号a, #, - で、先頭の1文字は英字でなくてはならない。長さは8文字以内。
- 変数名で大文字、小文字の区別はなく、(Love と love は同じとみなされる) 予約名の ALL, AND, BY, EQ, GE, GT, LE, LT, NE, OR, THRU, TO, WITH は利用不可。
- 変数の型は計算の可能な数値型(Numeric)が標準、文字列を入れる場合は文字型(Strings)、日付は日付型(Date)を利用。前出の例では、NAME, SEXが文字型で、ID, HEIGHT, WEIGHTは数値型となります。
- 指定がない場合はVAR0001, VAR0002, ... と仮定。

アスキーファイルを読み込みデータファイルを作成する手順を以下に示します。

1. データウィンドウで File → Read ASCII Data .. と選択します。

2. ファイル選択画面が現れ、ファイル名を指定します。(この例では、ファイルの形式を fixed とします。)

3. Define ボタンを押すと、変数定義画面が現れます。ここに、アスキーファイルの数値文字並びを変数として定義します。(一つの観測値に対するレコード数が1になっていることを確かめます)

- (a) 変数名(Name) を記入
- (b) 対応する変数データの始まるカラム数(Start Column)、終了カラム数(End Column)を記入
- (c) 変数の型と形式(Data Type)を選択、NAME と SEX は文字列ですので、strings を選択します。他は数値(numeric)で小数点形式はおまかせ(as is)を選びます。
- (d) Add ボタンを押して、Define Variables: 枠に入れます。間違えた場合は、変数を選択して修正し、Modify ボタンを押します。変数除去の場合も同様で、除去(remove)を選択します。

4. ファイルがうまく読み込めていたら、File → Save を利用して `my.sav` と保存します。(拡張子は必ず `*.sav` とします。)

次の起動からはデータウィンドウで、作成したデータファイルを指定し (File → Open → Data)、様々な作業を行うこととなります。

グラフの作成 折れ線・棒・円グラフの他に度数分布(histogram)、2変数プロット(scatter)などができます。方法は、データウィンドウで Graph→scatter.. などと選択し、対象の変数を決めればグラフィックウィンドウが現れます。さらに Edit ボタン(メニューではない)を選択するとグラフィックウィンドウの編集も可能です。

検定・推定など 主成分分析、因子分析、重回帰分析、分散分析などが可能です。データウィンドウで Statistics→ を選択します。

困ったら SPSS のデータウィンドウで HELP ボタンを押すと、概念・用語の説明、キーワード検索、などが可能です。SPSS など様々な質問集などは Mosaic コマンドで表示されます。

入出力サーバ cuios の使い方

入出力サーバ **cuios** では以下のデバイスが使用可能です。

装置名	デバイス名	
テープ装置	QIC-150	/dev/rmt/1
	DAT (DDS2)	/dev/rmt/2
	EXABYTE	/dev/rmt/0
CDROM ドライブ	/dev/dsk/c0t6d0s2	
CD 作成装置	/dev/dsk/c1t2d0	
カラーイメージスキャナ	/dev/jxis0	
画像入力装置 (カメラ)	—	

テープ装置

1. ファイルをテープに記録する。

```
cuios% tar cvf デバイス名 ファイル名
```

デバイス名はテープ装置のデバイス名をファイル名には、記録したいファイル名かディレクトリ名を入力。(ファイル名は、複数指定することができます。)

例) ホームディレクトリ全体を DAT に保存する。

```
cuios% tar cvf /dev/rmt/2 ~
```

2. テープに記録されている内容の一覧を見る。
(tar コマンドで記録した場合)

```
cuios% tar tvf デバイス名
```

3. テープの内容をディスクに戻す。(tar コマンドで記録した場合)

```
cuios% tar xvf デバイス名 [ファイル名]
```

ファイル名には、戻したいファイルを入力。ファイル名を省略した場合、記録されている内容全てをディスクに戻す。

CDROM ドライブ

1. 備え付けのキャディーに CD をセットして CDROM ドライブに入れると、自動で /cdrom/cdrom0 に CD がマウントされる。

```
cuios% ls /cdrom/cdrom0
```

2. CD を取り出す。

```
cuios% eject
```

CD 作成装置 (CD-R)

1. /cuios の下にログイン名のディレクトリを作成し、そこを作業ディレクトリとする。

```
cuios% mkdir /cuios/ログイン名  
cuios% cd /cuios/ログイン名
```

以降の作業は、作業ディレクトリ上で行ないます。

2. 作業ディレクトリに、CD 作成ソフトの設定ファイルをコピーする。

```
cuios% cp /usr/gear/.gearrc .
```

3. CD 作成ソフトを立ちあげる。立ち上がると、CD> というプロンプトになる。

```
cuios% msgen
```

4. ボリューム名 (CD の名前) を付ける。

```
CD> newvol ボリューム名  
Please enter volume ... ← CDの種類を入力  
Volume ボリューム名 created.  
Please enter type ... ← 記録形式を入力  
  
Track 1 created.  
  
CD>
```

5. データのコピー

```
CD> cp -r dir_name CD_dir
```

dir_name : コピー元のディレクトリ名

CD_dir : CD 上のディレクトリ名

6. パフォーマンスの測定

```
CD> estimate  
MESSAGE: This is only an estimate ...  
.  
.  
(0)k/(C)ancel (0k) ← リターン  
Using the virtual image for...  
.  
.  
CD>
```

7. CD への書き込み

```
CD> write
```

8. 終了

```
CD> exit
cuio% 
```

カラーイメージスキャナ

1. 原稿をスキャナにセットする。
2. スキャナユーティリティを立ちあげる。立ち上がると、メニューが表示される。

```
cuio% nss
```

3. メニューから **Scanner** を選ぶ。新たに、解像度などの設定メニューが表示される。
4. 解像度や色数等を設定する。
5. メニューから **Execute** を選び、画像を取り込む。

6. メニューから **Quit** を選び、終了。

画像入力装置 (カメラ)

1. カメラの裏のスイッチを入れる。
2. ビデオツールを立ちあげる。

```
cuio% videotool
```

3. **view menu** で **input1** を選択。画像が表示される。
4. **edit menu** から **config** を選択し、設定をする。
必要に応じて **config** 以外の項目も設定。
5. **file menu** から任意の項目を選択。
6. 終了するには、**videotool** ウィンドウの逆三角形から **quit** を選択。

総合情報処理センター講習会

— CUAPS の FORTRAN —

講師：伊賀泰之 (キャノンスーパーコンピューティング)

場所：総合情報処理センター電算実習室 1

日時：9月26日(木) 14時 - 16時

対象：センター利用者番号をお持ちの方

内容：SUNSOFT f77, CRAY CF90 の使用方法

LibSci を使ったプログラミング

並列化について

編集後記

秋の虫の音も聞こえるようになりました。皆さんいかがお過ごしでしょうか。夏休みに5日間にわたってセンターで高校生のためのサマースクールが開かれました。FORTRAN を用いたベクトル計算についての講義で、実習では **cuhas** が使われました。たいへん熱心にプログラミングを学びとり、満足して帰っていった様子です。彼らが研究者として活躍する頃にはどういう計算機が主流になっていることでしょう。(山下)

[予定]

以下の日時はセンター内利用ができません。

9/24 定期点検日 (9時~13時)

9/30 月末処理日 (全日)

[広報編集部門]

千葉大学総合情報処理センター
〒263 千葉市稲毛区弥生町1-33
TEL 043-290-3536
FAX 043-290-3544
E-mail editor@yuri.ipc.chiba-u.ac.jp