SAS-汎用統計プログラム-の概要

SAS－汎用統計プログラム－の概要

森川　利信*

SAS とは何か

SAS システムは Statistical Analysis System の名前が示すように、主に統計解析用のコンピュータソフトウェアの一つである．SAS は 1966 年に米国ノースカロライナ州立大学で IBM メインフレーム（大型汎用コンピュータ）用に開発されたもので，その後何度も機能拡張が行われ，すべての種類のデータ処理に適用できるエンドユーザー用ソフトウェアとして，世界中に多くの熱烈なユーザーをもっている．生物科学や社会科学における統計解析は，ほとんど SAS を用いて行われているといっても過言ではない．それは，既成の統計パッケージの中で，最も大きく最も良く整備されているからである．伝統的なパッケージとしては，SPSS，BMDP，GENSTAT-V などがあり，かなり高度な統計計算が可能であるが，FORTRAN 書式の入力やバッチジョブのみといった不自由さがあった．本学の大型計算機として長年稼働してきた NEC S-3700/10 ACOS6 上では STATPAC と SPSS が利用できたが，SAS は NEC の OS と互換性がないので移植されていなかった．この近辺では，京都大学大型計算センターの Fujitsu M-1800E の MSP 上だけで SAS が稼動していた．また，メインフレームでは，プログラム作成やデータ入力はラインエディタで行うのが普通であり，日本語スクリーンエディタは使えなかった．PFD(Fujitsu)などのスクリーンエディタは，研究室のリモート端末からエミュレーションモードでようやく使えるもののすぐにフリーズするし，グラフィック画面が使えないなどの欠点があって，ほとんど実践的でなかった．計算結果を京都大学のプリンターに直接出力することができるが，配送サービスを受けなければならないなどの不自由さがあった．

私は，1985 年から一年間在外研究員として，英国の University College of Wales Aberystwyth, Plant Breeding Station（現在は Institute of Glass land and Environmental Research）に滞在していた．そのとき，VAX 上で稼働する Minitab や SAS を利用して，その使いよさとパワフルさに驚いた．その後，統計パッケージのとりこになり，大学間ネットワークを利用してメインフレーム版の統計パッケージをいろいろと使ってみた．その中で，SAS の使い勝手が最も良かった．SAS スーパーバイザーとよばれる管理機能，GLM（一般線形モデル）の多機能性，SAS データーセットに代表されるファイル管理とプログラミング機能，それにプロシジャとよばれる統計コマンドの豊富さである．最近，SAS はメインフレーム以外のワークステーション(WS)やパソコン(PC)システムへの移植作業が進められ，個人レベルで使えるようになった．ユーザーの多さからこのパソコン版は，むしろメインフレーム版より新しいバージョンが供給されており，同等以上の機能をもっている．

三年前の教育研究用情報処理システムの更新時に，待望の SAS システム ver 6.12 版が本学にも導入され，現在，実習室１やオープンスペースのパソコンと汎用 Unix(namihaya)上で稼働している．１０万～２０万円ほどのパソコンでも一頃の汎用機と同じ計算能力を持つようになったので，最新の SAS システムがインストールされていれば，有り余るほどの情報処理能力が発揮でき，適切なデータ分析と情報に基づく，教育研究開発の意思決定を支援することができる．

---

*大阪府立大学農学部応用植物科学科助教授

このような高度な情報処理システムを研究用に使わない手はない．情報処理教育ではよく使われているようだが，まだ利用者は少ないように思う．ワープロと表計算ソフトを卒業して，統計システムを使って複雑なデータの山から有用な情報をどのように引き出すか．このバリアーを超えたいと考えている人に，SAS システムはうってつけのソフトウエアである．しかし，SAS システムにも欠点があるので，熱烈な信望者がいる反面，なかなか新しいユーザーが増えないのも事実である．その第一理由は，分厚いマニアルを統計解析項目ごとに読まなければならないことにある．最近，日本でも解説書が出版されるようになったが，その内容は豊富過ぎてまだ難解である．第二の理由は，マニアルやソフトウエアは，一部日本語に翻訳されているが，ほとんど英語のままである．第三の理由は，ソフトウエアがライセンス契約（レンタル）でしか供給されていないことである．また，ユーザーが増えれば価格も下がるのだろうが，その価格が個人で買えるほど安くないことである．

SAS の内容

どのような解析作業ができるのか解説してみよう．SAS システムは，広範な統計ツールの集合体で，幅広い分析に対応していて新薬の臨床試験，マーケティング，健康調査，顧客意識調査や株式市場のトレンドなど，あらゆる種類のデータが扱える．しかし，特別専門的なユーザーでないかぎり Base SAS, SAS/STAT, SAS/INSIGHT, SAS/ASSIST および SAS/GRAPH の５つのソフトウエアを活用することで十分な作業はできる．これらは独立したソフトではなく Base SAS を中心にモジュールを形成し，特別意識せずにすべてのソフトをツールとして利用することができる．以下にそれらを，個別に解説する．

1. Base SAS は中心になるソフトウエアであり，データアクセス，ファイル管理，基本分析およびプレゼンテーションを掌っている．データアクセスは，あらゆるフォーマットやファイルからも可能である．また，記述統計量，相関や連関性，クロス集計や推計統計量を計算できる．

2. SAS/STAT は，データ分析用総合ツールで，SAS システムの統合コンポーネントで拡張統計機能を専門的なデータの解析に使えるようになっている．分散分析，回帰分析，カテゴリーデータ分析，多変量解析，生存分析，精神測定分析，クラスター分析およびノンパラメトリック分析などの広範囲な統計解析に対応している．

3. SAS/INSIGHT は，データの視覚化と対話型データ分析のための高度な対話機能を持つツールで，ビジュアルなデータ解析ができる．自分で実験データをとったり調査データを整理した経験のある人なら，生データには必ずと言っていいほど"はずれ値"が含まれていることを知っている．すぐに高度な解析を行うことはまれで，まず，"はずれ値"を見つけ，データの傾向を知るのだが，この作業には最適のツールである．また，最初から，データの類推ができない場合が多いが，強力なモデリング機能を使っていろいろテストすることができる．具体的には，1 変数の統計量と分布，多変量データの視覚化，回帰モデル，共分散分析および一般化線形モデルへのあてはめが可能である．

4. SAS/ASSIST は，経験度合いに関係なくすべてのユーザーが適切な解析作業ができるように，対話型であらゆる統計解析の必須フィールド，選択リスト，変数の選び方を解説してくれる．メニュースクリーンではキーワードに従って適切なアイコンを選択することで，試行錯誤の末に最終結果を得ることができる．SAS プログラミングの構文を知らなくても一応使えることを前提としている．

5. SAS/GRAPH は，情報およびプレゼンテーションカラーグラフィック機能をもっていて，多

彩な色とパターンによるさまざまなチャート図，プロット図および地図グラフを作成することができる．SAS システムのデータ管理および分析ツールの能力を拡大することにより，データから人目を引くフルカラーの三次元グラフィックおよび等高線図に変換することができる．

SAS プログラムの作成と実行

SAS の初期画面

WINDOWS98 上で SAS を起動すると，PROGRAM EDITOR(PGM)，LOG，OUTPUT のウインドウが現れる．メニューバーのウィンドウ(W)から分割画面や単独画面を自由に選択できる．PGM はプログラムの編集を行うところで，行番号を表示させたり，コピー，切り取り，貼り付けが自由にできる．WORD などワープロソフトで作ったプログラムテキストファイルを読み込んでもかまわない．LOG は実行時に SAS 処理系から出されるメッセージを表示する．ここに出される赤文字のエラーメッセージをたよりに，プログラムを修正する．OUTPUT は，統計処理等の結果を表示する．プログラムの実行には，SUBMIT コマンドを使うが，ランニングマークのアイコンをクリックする方が簡単である．エラーがあれば，PGM に戻り，RECALL コマンドを押せば，プログラムが再表示されるので，修正してから再度実行する．

SAS のプログラム構成

SAS のプログラムは，基本的な４つの部分からなる．それは，SAS ステートメント，SAS データセット，DATA ステップおよび PROC ステップである．

1. SAS ステートメントは，SAS に対してある処理をさせるための命令文である．自由書式で書き，一つのステートメントを複数行に，複数のステートメントを一つの行に書いてもよい．セミコロン(;)で終わる．
2. SAS データセットは，SAS の作業用ファイルである．SAS は起動するとデータセットを次々に作成していく．一時的な作業用のデータセットの名前は"WORK.SAS データセット"がついている．このなかには，各個体に対するいくつかの変数(variable)のデータ値が行列ではいっている．個体のことをオブザベーション(observation)とよんでいる．個体×変数の形でデータ行列を作る．WORK.SAS データセットは，SAS セッション終了後には消去される．永久 SAS データセットを作るには，"ライブラリ参照名.SAS データセット"を指定する．
3. DATA ステップは，DATA ステートメントで始まり，SAS データセットを作成・編集する．生データを入力する，新しい変数を作る，データ値を変換する，および外部ファイルにデータ値を出力する．基本的には，DATA ステップは，オブザベーションの数だけ回るループになっている．
4. PROC ステップは，DATA ステップや他の PROC ステップですでに作られた SAS データセットを入力して，統計処理を行う．統計機能を表す名前がつけられたサブプログラム（プロシジャ procedure）を呼び出し，データを解析する．
5. RUN ステートメントは，SAS ステートメントの一つで DATA ステップや PROC ステップの終了を示し，統計用サブプログラムを実行に移す．

SAS によるプログラミングの実例

応用植物科学科の学部カリキュラムの中で，SAS を使った実験実習を行っているので，その一部のデータを利用して，SAS によるプログラミングの実例を紹介しよう．２回生対象の応用植物科学実験第１と応用植物情報処理演習では，イネにおける矮性遺伝子の形質発現を，散布図，平均値の差の検定（t 検定）および主成分分析を用いて解析している．以下に，その内容を簡単に解説する．

目的：イネの矮性品種の一つである短銀坊主と品種日本晴の成熟植物体の形態形質を比較し，矮性遺伝子 *d35* の形質発現の様式を知る．

概説：イネ矮性遺伝子系統の多くは，内在するジベレリン様物質の含有量が極めて少なく，ジベレリン酸(GA3)を経時的に投与することによって，その草丈を正常に回復させることができる．また，矮性遺伝子の多くは，草丈の矮化だけでなく，他の多くの形態形質を縮小させる多面的な作用があることが知られている．ここでは，イネの矮性品種と高性品種について，多くの形態形質を比較する．

平均値の差の検定：イネの矮性品種と高性品種について，成熟植物体の各形態形質の平均値を比較し，統計的に有意な差があるかどうか検定する．これを行うために，二集団の差の標準誤差を推定し，平均値の差を比較して，t 値を求める．この独立する平均値の差の検定は，二つの集団の分散が同じであることを前提にしている．また，この二組の測定値は，それぞれ，お互いに独立していることを想定している．

材料：イネ(*Oryza sativa* L.)矮性品種 短銀坊主（*d35d35*）．品種日本晴（*d35⁺d35⁺*）．

方法１：両品種の成熟植物体 10 株の分けつ数を数える．次に，主幹を選び出し，草丈，穂長，小花数，止葉葉身長，第Ⅱ葉葉身長，第一節間長および第二節間長を測定する．

方法２：各形質について，二つの品種の平均値，分散，標準偏差，標準誤差を求める．

方法３：検定統計量 t を求める．それは，以下の式で与えられる．

$$t = (\text{平均値の差}) / (\text{差の標準誤差})$$

$$= \frac{(\overline{Xa} - \overline{Xb})}{SEd} \quad SEd = \sqrt{SEa^2 + SEb^2}$$

方法４：有意確率 p を求める．帰無仮説が真である時の t よりも大きいか等しい検定統計量の確率 p を計算する．自由度は Na+Nb-2=18 である．Na と Nb は，それぞれ，集団の標本の大きさを示しここでは共に 10 である．なお，t[0.05, 18]=2.101，t[0.01, 18]=2.878 および t[0.001, 18]=3.922 である．

方法５：もし p<0.05 ならば，帰無仮説を捨てて対立仮説を採用する．すなわち，二つの平均値の差は，有意であると判断する．もし p＞0.05 ならば，帰無仮説を保留する．すなわち，二つの平均値の差は有意でないと判断する．

考察１：各形質について，二品種間の平均値の差は，有意であるといえるか．

考察２：考察１の結論から考えて，矮性遺伝子 *d35* は多面発現しているといえるか

図表１．は SAS データセット dwarf のプリント出力結果を示している．これが，イネの２品種８形態形質についての生データである．SAS データセットの作成と t 検定を行う TTEST プロシジャの SAS プログラムを作る．ただし，n は日本晴，t は短銀坊主を示している．

T-test and PCP analysis on two rice plants

OBS	CV	TILLER	HEIGHT	PANICLE	FLORET	FLAG	LEAF2	NODE1	NODE2
1	n	10	90.0	19.0	101	27.0	35.5	36.5	16.0
2	n	6	103.0	23.5	159	37.0	44.5	42.0	19.0
3	n	6	96.0	24.5	151	35.5	47.5	39.0	19.5
4	n	6	100.0	24.5	164	37.5	47.5	40.0	13.0
5	n	5	104.0	23.5	158	35.0	42.0	60.5	20.0
6	n	7	87.0	20.0	142	20.5	40.5	48.0	12.0
7	n	4	111.5	22.0	136	31.0	34.5	60.0	11.5
8	n	6	103.5	21.5	132	25.5	36.5	60.0	12.5
9	n	11	101.5	20.5	98	30.5	40.0	60.0	11.5
10	n	5	92.0	20.0	129	22.0	32.0	33.0	16.0
11	t	9	64.0	18.0	52	16.5	26.0	23.0	15.0
12	t	7	66.0	17.0	92	18.0	23.5	22.5	14.5
13	t	6	66.0	16.5	68	18.0	28.0	22.5	13.0
14	t	8	72.0	15.0	74	17.0	22.0	22.0	13.5
15	t	5	71.0	18.0	86	14.0	22.0	26.0	15.0
16	t	8	67.0	20.0	88	20.0	24.0	29.0	15.5
17	t	6	69.0	15.0	81	15.0	22.0	21.0	12.0
18	t	10	72.0	17.0	101	12.5	24.0	23.5	15.0
19	t	11	75.0	18.5	105	16.0	25.0	25.5	15.0
20	t	8	69.0	17.0	93	15.0	23.0	23.0	14.0

図表１．SASデータセットdwarfのPRINTプロシジャによる出力

```
title 'T-test and PCP analysis on two rice plants';
options ps=60 ls=80;
data dwarf;
input cv $ tiller height panicle floret flag leaf2
      node1 node2;
cards;
 n 10  90   19   101 27   35.5 36.5 16
 n  6 103   23.5 159 37   44.5 42   19
 n  6  96   24.5 151 35.5 47.5 39   19.5
 n  6 100   24.5 164 37.5 47.5 40   13
 n  5 104   23.5 158 35   42   60.5 20
 n  7  87   20   142 20.5 40.5 48   12
 n  4 111.5 22   136 31   34.5 60   11.5
 n  6 103.5 21.5 132 25.5 36.5 60   12.5
 n 11 101.5 20.5  98 30.5 40   60   11.5
 n  5  92   20   129 22   32   33   16
 t  9  64   18    52 16.5 26   23   15
 t  7  66   17    92 18   23.5 22.5 14.5
 t  6  66   16.5  68 18   28   22.5 13
 t  8  72   15    74 17   22   22   13.5
 t  5  71   18    86 14   22   26   15
 t  8  67   20    88 20   24   29   15.5
 t  6  69   15    81 15   22   21   12
 t 10  72   17   101 12.5 24   23.5 15
 t 11  75   18.5 105 16   25   25.5 15
 t  8  69   17    93 15   23   23   14
;
proc print; run;
proc ttest;
 class cv;
var tiller height panicle floret flag leaf2
    node1 node2;
run;
```

プログラムの表題.
ページ設定は６０行８０字.
データセット名はdwarf.

９変数を設定する.
データの始まりを示す.

データの終りを示す.
データセットdwarfの出力.
平均値の差の検定を行う.
品種間でｔ検定を行う.

ｔ検定の対象変数を指示する.

図表２．SASデーターセットdwarfの作成とTTESTプロシジャのSASプログラム

演習課題

1. 応用植物科学実験第１（イネにおける矮性遺伝子の形質発現）で得られたデータを使って，平均値の差の検定を行うSASプログラムを作成し実行する．
2. 手計算で行った検定とTTESTプロシジャのSASプログラムの結果を比較する．
3. 穂長と小花数の散布図を作るSASプログラムを作成し，両品種における両形質の関係を視覚化する．

図表２．にSASステートメント，DATAステップによりSASデータセットdwarfの作成・出力，平均値の差の検定を行うSASプログラムを示した．

TTEST PROCEDURE

Variable: TILLER

CV	N	Mean	Std Dev	Std Error	Minimum	Maximum
n	10	6.60000000	2.22111083	0.70237692	4.00000000	11.00000000
t	10	7.80000000	1.87379591	0.59254629	5.00000000	11.00000000

Variances	T	DF	Prob>|T|
Unequal	-1.3059	17.5	0.2085
Equal	-1.3059	18.0	0.2080

For H0: Variances are equal, F' = 1.41 DF = (9,9) Prob>F' = 0.6206

****

Variable: HEIGHT

CV	N	Mean	Std Dev	Std Error	Minimum	Maximum
n	10	98.85000000	7.50943851	2.37469296	87.00000000	111.50000000
t	10	69.10000000	3.41402337	1.07960898	64.00000000	75.00000000

Variances	T	DF	Prob>|T|
Unequal	11.4046	12.6	0.0001
Equal	11.4046	18.0	0.0000

For H0: Variances are equal, F' = 4.84 DF = (9,9) Prob>F' = 0.0279

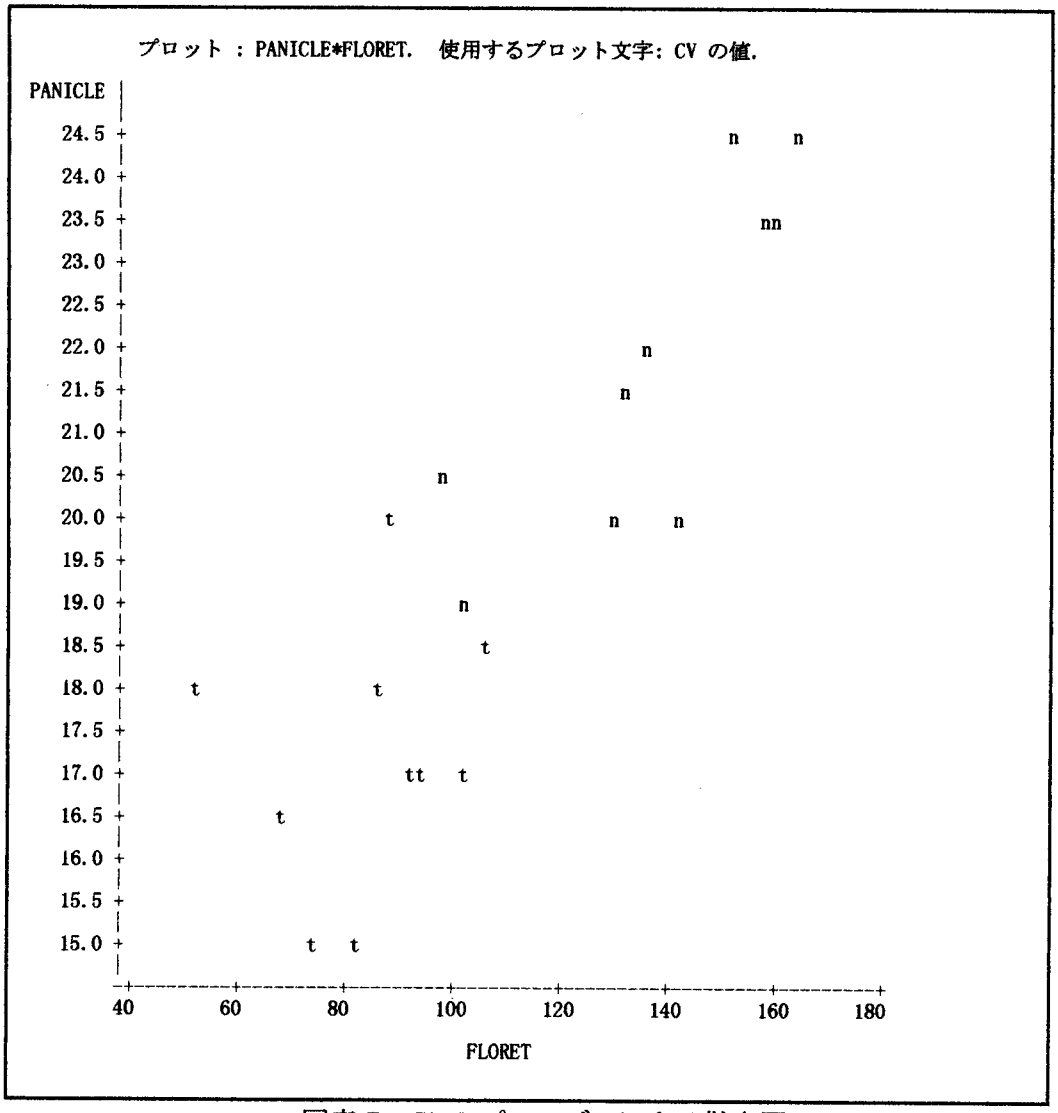
図表３．TTESTプロシジャによる分けつ数と草丈に関する平均値の差の検定

```
proc plot;                          散布図を書く．
     plot panicle*floret=cv;        穂長と小花数の間で．
run:
```

図表４．PLOTプロシジャによる散布図作成SASプログラム

図表３．には，８形態形質中分けつ数と草丈について，TTESTプロシジャによる平均値の差の検定を示した．TTESTプロシジャは，基礎統計量の出力とともにt値，等分散性を検定してくれる．分けつ数は，等分散（Pro>F'=0.6206 >0.05）なので，EQUALの行を見る．Tの絶対値は，1.3059でProb>T=0.2080 >0.05なので，有意差はないと判断する．草丈は，非等分散

(Prob>F′=0.0279 <0.05)なのでUNEQUALの行を見る．Tの絶対値は，11.4046でProb>T=0.0001 <0.05なので，高い有意水準で差があると判断する．残りの6形質中，第二節間長だけ有意差がなかったので，草丈，穂長，小花数，止葉葉身長，第Ⅱ葉葉身長および第一節間長では，品種間差が有り，分けつ数と第二節間長では差がないことがわかった．したがって，矮性遺伝子 *d35* は草丈の矮化作用以外にも多面発現しているといえる．



図表５．PLOTプロシジャによる散布図

両品種の形態特性を把握するために，穂長と小花数による散布図を描き二変量のデータを視覚化する（図表５．）．両形質の間には高い正の相関が認められるが，両品種のデータは連続していて明確に区別することできない．そこで，8形質を使って主成分分析を行い，両品種を明確に区別する総合指標を抽出してみる．

主成分分析

主成分分析とは，ある問題についていくつかの要因が考えられるとき，それらの要因を一つ一つ独立に扱うのではなく，総合的に取り扱おうとする分析法である．つまり，いくつかの説明変量 x1, x2, ・・・・, xp の総合特性を

$$a1x1+a2x2+\cdots+apxp$$

の様な少数個の１次式で表現することである．この式によって表されるものを主成分(principal component)という．別の言い方をすれば，主成分分析とは多くの変量 x1, x2, ・・・, xp の値を出来るだけ情報の損失を少なくし，１個または互いに独立な総合指標 z1, z2, ・・・, zm で代表する手法である．

$$z1= a11x1+a12x2+\cdots+a1pxp$$
$$z2= a21x1+a22x2+\cdots+a2pxp$$
$$\vdots$$
$$zm= am1x1+am2x2+\cdots+ampxp$$

z1, z2, ・・・, zm をそれぞれ第１主成分，第２主成分, ・・・, 第m主成分と呼ぶ．

具体例として，二変量の場合を考えてみる．説明変量 x1 を穂長，説明変量 x2 を小花数とおく(図表５．)．目標は，この二つの説明変量の総合的特性を求めることにある．すなわち，a1x1+a2x2 という１次式によって表される主成分を探してゆく．この式の係数 a1, a2 は主成分直線Ｚの傾きを表している．また，各点からＺにおろした垂線の長さを，情報量の損失と呼ぶ．主成分は，情報量の損失を最小にする係数 a1, a2 を求めることによって得られる．

主成分分析を理解するためのキーワード

1. 固有値(eigenvalue)：各主成分の分散を表す．情報の損失量の平方和と等しい．
2. 固有ベクトル(eigenvector)：a1, a2, ・・・・, am の係数を示す．主成分の意味する総合特性を表す．
3. 主成分得点(principal component score)：各点からＺ軸に下す垂線との交点のＺ軸での値．
4. 寄与率(propotion)：{(元の情報の平方和)-(情報の損失量の平方和)}／(元の情報の平方和)
5. 累積寄与率(cumulative proportion)：第１から第 i 主成分までの寄与率を累積したもの．主成分の数 i はなるべく少なくデータの情報を反映できることが望ましい．第１から第 i 主成分までの累積寄与率が 0.8 以上であることを一つの基準としている．

演習課題

1. イネの８形態形質を用いて，主成分分析を行う SAS プログラムを作成し実行する．
2. 第一主成分と第二主成分の固有値(eigenvalue)，累積寄与率(cumulative proportion)を求める．
3. 第一主成分と第二主成分の固有ベクトル(eigenvector)は，それぞれ，どのような総合指標を表しているか．

```
proc princomp out=out_prin;                    主成分分析を行う.
var tiller height panicle floret flag leaf2
    node1 node2;                               主成分分析の対象変数を指示する.
proc print; run;                               out_prin を出力する.
proc means;                                    各変数の平均値を求める.
proc plot;                                     散布図を書く.
     plot prin2*prin1=cv/vref=0 hreh=0;        第一と第二主成分の間で.
run;
```

図表６．PRINCOMP プロシジャによる主成分分析のプログラム

Simple Statistics

	TILLER	HEIGHT	PANICLE	FLORET
Mean	7.200000000	83.97500000	19.55000000	110.5000000
StD	2.092593455	16.28324534	2.97312524	33.3332456

	FLAG	LEAF2	NODE1	NODE2
Mean	23.17500000	32.00000000	35.85000000	14.67500000
StD	8.46397163	9.15940701	14.65578743	2.53021218

Correlation Matrix

	TILLER	HEIGHT	PANICLE	FLORET	FLAG	LEAF2	NODE1	NODE2
TILLER	1.0000	-.3304	-.3443	-.4331	-.3171	-.2471	-.2487	-.1213
HEIGHT	-.3304	1.0000	0.8282	0.8316	0.8690	0.8432	0.9030	0.1419
PANICLE	-.3443	0.8282	1.0000	0.8683	0.9058	0.8958	0.7288	0.4221
FLORET	-.4331	0.8316	0.8683	1.0000	0.7959	0.8393	0.6766	0.3531
FLAG	-.3171	0.8690	0.9058	0.7959	1.0000	0.9219	0.7316	0.3598
LEAF2	.2471	0.8432	0.8958	0.8393	0.9219	1.0000	0.7353	0.2975
NODE1	-.2487	0.9030	0.7288	0.6766	0.7316	0.7353	1.0000	-.0088
NODE2	-.1213	0.1419	0.4221	0.3531	0.3598	0.2975	-.0088	1.0000

図表７．PRINCOMP プロシジャの要約統計量と相関行列の出力

T-test and PCP analysis on two rice plants

Variable	N	Mean	Std Dev	Minimum	Maximum
TILLER	20	7.2000000	2.0925935	4.0000000	11.0000000
HEIGHT	20	83.9750000	16.2832453	64.0000000	111.5000000
PANICLE	20	19.5500000	2.9731252	15.0000000	24.5000000
FLORET	20	110.5000000	33.3332456	52.0000000	164.0000000
FLAG	20	23.1750000	8.4639716	12.5000000	37.5000000
LEAF2	20	32.0000000	9.1594070	22.0000000	47.5000000
NODE1	20	35.8500000	14.6557874	21.0000000	60.5000000
NODE2	20	14.6750000	2.5302122	11.5000000	20.0000000
PRIN1	20	2.220446E-16	2.3184298	-2.6030352	3.7218451
PRIN2	20	-2.8727E-16	1.0405410	-2.1933953	1.7762903
PRIN3	20	-1.06512E-16	0.9397754	-1.3728076	1.9627944
PRIN4	20	-2.77556E-17	0.5056806	-1.5223115	1.0625194
PRIN5	20	9.15934E-17	0.4468593	-0.7671237	0.8312846
PRIN6	20	2.925177E-16	0.3175038	-0.7157273	0.5892521
PRIN7	20	4.510281E-17	0.2731254	-0.7490107	0.4956231
PRIN8	20	-2.18575E-16	0.1678624	-0.3372290	0.3174429

図表８．MEANS プロシジャによる OUT_PRIN の出力

Eigenvalues of the Correlation Matrix

	Eigenvalue	Difference	Proportion	Cumulative
PRIN1	5.37512	4.29239	0.671890	0.67189
PRIN2	1.08273	0.19955	0.135341	0.80723
PRIN3	0.88318	0.62746	0.110397	0.91763
PRIN4	0.25571	0.05603	0.031964	0.94959
PRIN5	0.19968	0.09887	0.024960	0.97455
PRIN6	0.10081	0.02621	0.012601	0.98715
PRIN7	0.07460	0.04642	0.009325	0.99648
PRIN8	0.02818	.	0.003522	1.00000

Principal Component Analysis

Eigenvectors

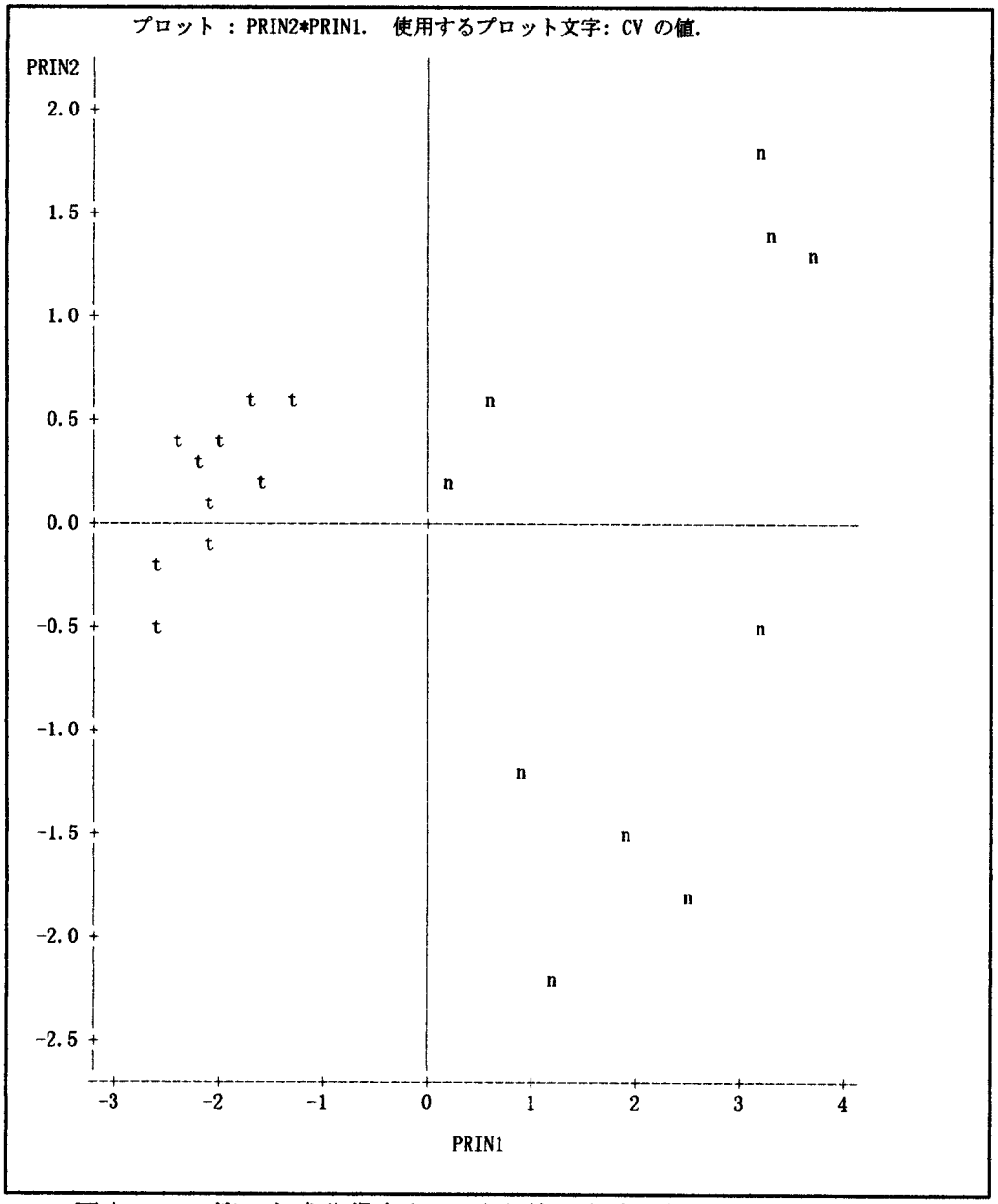
	PRIN1	PRIN2	PRIN3	PRIN4
TILLER	-.178799	-.120356	0.955096	-.035486
HEIGHT	0.403970	-.234211	0.029322	0.271905
PANICLE	0.410253	0.114491	0.075137	-.147514
FLORET	0.395102	0.080771	-.086344	-.263522
FLAG	0.407456	0.039623	0.110450	.188016
LEAF2	0.404087	-.015415	0.174344	-.431631
NODE1	0.358729	-.404330	0.061111	0.658521
NODE2	0.149776	0.863561	0.165837	0.422255

	PRIN5	PRIN6	PRIN7	PRIN8
TILLER	0.169666	0.013416	0.096868	0.041577
HEIGHT	0.085606	0.502765	0.301682	-.596697
PANICLE	-.039351	-.749970	0.393103	-.269528
FLORET	0.789514	0.124027	0.035415	0.347135
FLAG	-.551318	0.297064	0.342779	0.524560
LEAF2	-.187223	0.047385	-.718094	-.258492
NODE1	0.000520	-.266198	-.305486	0.325980
NODE2	0.000434	0.088636	-.131687	-.028828

図表９．PRINCOMP プロシジャによる固有値と固有ベクトルの出力

第一主成分と第二主成分の固有値は，それぞれ，5.37512 と 1.08273 であり，第二主成分までの累積寄与率は 0.80723>0.8 である(図表 9.)．したがって，情報量の損失は少なく第一主成分と第二主成分がうまく抽出できたといえる．また，第一主成分の固有ベクトル，すなわち重み係数は，分けつ数を除いたすべての変数に対してほぼ同じような正の値である．したがって，第一主成分は，植物体のバイオマスのような総合指標を意味していると思われる．各主成分に対する重みのベクトルは直交するので，第二主成分以後の重み係数は，正・負入り混じったものとなっている．この例では，第二主成分の重み係数は主として節間長を表し，第一節間長は負の重みを表し，第二節間長は正の重みを表していることがわかる．

第一主成分得点と第二主成分得点を用いて，散布図を描くと(図表 10.)，穂長と小花数の二変量の散布図より，明確に両品種を区別することができた．その上，短銀坊主は１グループにまとめられるが，日本晴には，第一節間長や第二節間長が異なる２グループが混在する可能性が推測できる．



図表１０．第一主成分得点(PRIN1)と第二主成分得点(PRIN2)の散布図