

統計解析プログラム講座 2

回帰分析と主成分分析

芳賀敏郎/橋本茂司

日科技連

統計解析プログラム講座 2

回帰分析と主成分分析

芳賀敏郎/橋本茂司

日科技連

著 者 紹 介

芳 賀 敏 郎 (はが としろう)

1951年 東京大学理学部化学科卒業

現 在 慶応義塾大学講師 医学部

著 書 実験計画法(共著), 培風館, 1969

多変量解析法(共著), 日科技連出版社, 1971

統計数値表(共編著), 日本規格協会, 1972

官能検査ハンドブック(共編著), 日科技連出版社, 1962, 1973

続多変量解析法(共著), 日科技連出版社, 1976

簡約統計数値表(共編著), 日本規格協会, 1977

新版品質管理便覧(共編著), 日本規格協会, 1977

統計解析プログラムの基礎(共著), 日科技連出版社, 1980

橋 本 茂 司 (はしもと しげじ)

1964年 東京大学工学部計数工学科卒業

現 在 東レ株式会社 システム部勤務

著 書 続多変量解析法(共著), 日科技連出版社, 1976

新版品質管理便覧(分担執筆), 日本規格協会, 1977

統計解析プログラムの基礎(共著), 日科技連出版社, 1980

統計解析プログラム講座 2

回 帰 分 析 と 主 成 分 分 析

1980年 5月21日 第1刷発行

定価 2200円

1981年 12月10日 第3刷発行

検 印
省 略

著 者 芳 賀 敏 郎

橋 本 茂 司

発行人 光 明 春 子

発行所 株式会社 日科技連出版社

東京都渋谷区千駄ヶ谷5-4-2

郵便番号 151

電話 03(352) 2231(代)

振替口座 東京 7-7309番

印刷 東京河北印刷株式会社

製本 板倉製本印刷株式会社

Printed in Japan

© Toshiro Haga, Shigeji Hashimoto 1980

は　じ　め　に

筆者らは、情報処理研修センターにおける工業高等専門学校と東京都立高校の教員を対象とするセミナーで、多変量解析を担当してきた。このセミナーは、講義を従として、各自がコンピュータを自由に使うことによって、手法とコンピュータの利用技術を修得するところに特長がある。ここでの5年間の経験によって、多変量解析は、数理的に解説を受けるよりも、身近の問題を取り上げてデータを取り、自分の作ったプログラムに通してみるほうが、ずっとその本質に接近できることがわかった。しかし、短期間のセミナー開講中にプログラムをすべて作成することは困難であるので、必要なサブルーチンを講師側であらかじめ作成しておき、メインルーチンのみを研修生が作るという方法をとった。そのセミナーで使用した、サブルーチンの使い方、メインプログラムの作り方のテキストが本書のもとになっている。

したがって、この本のねらいと特徴は次の点にあるといえる。

- (1) データ解析をしたい人が、身近のコンピュータを使って手軽に行えるように基本的なサブルーチンを提供し、その組合せによる使い方を示している。
- (2) メインルーチンについても標準的なプログラム例を与えているので、ほとんどの場合そのまま利用できる。
- (3) 全体としてアルゴリズムに重点をおき、アルゴリズムを理解することによって手法を理解し、出力の正しい解釈ができるようにくふうしている。
- (4) 重回帰分析については、変数選択の諸手法について統一的なアルゴリズムを示したほか、PSS など最近の進歩についても詳述している。

(5) 層別因子を含む重回帰分析，数量化理論 I 類のデータ解析における変数選択の方法を与えている。

多変量解析の手法の解説は，プログラムの作成と，出力の理解に必要と思われる範囲に限定した。手法の詳しい解説，特に数理的側面については，筆者らの別著^{(1),(2)}を参照されたい。

最近，プログラム作成技術について多くのくふうと試みが行われている。その話題の中心は，“他人が見てわかりやすい”(ということは，“書いた本人にもその内容が把握しやすく，間違いの少ない”ということに繋る)プログラムをどう作るかにある。多くの本が出版されているが，Kernighan 氏⁽³⁾の一読をおすすめする。従来プログラムは，結果が正しく出れば十分であると考えられてきたが，文章と同じように，美しさを備えていることの必要性がわかっていただけたと思う。

この本を書くにあたって，シェイクスピアや山本有三の文章のようなわけにはいかないが，プログラムの手本としての役割も果たしたいと考え，わかりやすく書くよう非力ながら努力したつもりである。

読者には，FORTRAN 文法の基本的知識を前提としている。わかりやすいプログラムを書くためには，FORTRAN 文法に制約をつけたほうがよいと考えられるので，ある規則(書法)に基づいてコーディングしている。その規則は最後の付録に解説してある。

この本にあげた FORTRAN プログラムはいくつかのコンピュータで正しく動くことを確かめてある。またアウトプットは実際の結果で，いずれも見やすくするために活字に組み直したものである。

索引に代えて，本書に出てくるプログラムの出現ページと，メインプログラ

(1) 奥野忠一，芳賀敏郎，久米 均，吉澤 正：『多変量解析法』，日科技連出版社，1971。

(2) 奥野忠一，芳賀敏郎，奥野千恵子，矢島敬二，橋本茂司，古河陽子：『続多変量解析法』，日科技連出版社，1976。

(3) Kernighan, B.W. & Plauger, P.J.: *The Elements of Programming Style*, McGraw-Hill, 1974. — 木村 泉(訳)：『プログラム書法』，共立出版，1976。

ムとサブプログラムの相互関連表を目次のあとに記載した。

これらのサブプログラムおよびそのままで実用に供しうるプログラム(2.1, 3.2, 4.3, 5, 6.1, 7.10)のソースカード(または磁気テープ)の入手を希望する読者は、日科技連出版社編集部まで相談されたい。

最初の草稿に目を通し、有益な助言をいただいた東京大学の奥野忠一教授に感謝の意を表したい。またプログラムのチェックをしていただいた古田桂子嬢にもお礼を申し述べたい。

構想の段階から、原稿を経て、校正の途中においても討議を重ねてお互いに手を入れ、またセミナーの講義テキストとして使用を重ねて改善をはかってきた経緯から、執筆の分担は明らかにしないことにした。この間合宿を含めて十数回に及ぶ会合を重ねながらようやく出版にこぎつけることができたのは日科技連出版社の田原 宏社長、伊藤幸夫氏のご鞭撻のたまものである。記して感謝したい。

1980年4月

著 者

統計解析プログラム講座 芳賀敏郎／橋本茂司

各巻
A 5 判／箱入／
238 ページ

第 1 巻 統計解析プログラムの基礎

四則演算／DO ループ／配列／
サブルーチン／わかりやすいプ
ログラム／基本分布の確率とパ
ーセント点／検定／乱数／マイ
コンによる数値実験／シミュレ
ーション／分布型の検討とロバ
スト推定

第 2 巻 回帰分析と主成分分析

第 3 巻 実験データの解析（続刊）

日科技連出版社

目 次

はじめに	iii
1 基本統計量	1
1.1 基本統計量とそのプログラム	2
1.2 データ行列の印刷(MATPRI)	7
1.3 なまの平方和・積和行列(行列の積MATMUL)	12
1.4 平方和・積和行列(掃出演算SWEEP)	14
1.5 平均と標準偏差(ベクトルの印刷VECPRI)	17
1.6 最大値, 最小値(MINMAX)	20
1.7 相関行列(CORREL)	21
1.8 内部記憶の大きさに制約のある場合	22
1.9 平方和・積和の逐次計算	24
2 基本重回帰分析	31
2.1 重回帰分析とは	32
2.2 基本重回帰分析のプログラム	36
2.3 正 規 方 程 式	38
2.4 分散分析と寄与率(ANOVA)	44
2.5 偏回帰係数の検定	45
2.6 回帰推定値・回帰残差(ESTIM)	50
2.7 残差のグラフ(GRAPH)	55
2.8 桁 落 ち	58
3 変数選択(1)——逐次選択法	61

3.1	変数増加法	62
3.2	変数減少法	67
3.3	変数増減法と変数減増法	72
3.4	強制変数	73
3.5	統合プログラム	73
3.6	逐次変数選択法の変形	85
3.7	対話による変数選択	86
4	変数選択(2)	97
4.1	回帰式の良さの評価	98
4.2	予測平方和(PSS)	102
4.3	総当り法のプログラム(1)	107
4.4	総当り法のプログラム(2)	118
5	層別因子を含む回帰分析	131
5.1	層別因子を含む回帰分析	132
5.2	層別因子を含む重回帰分析のプログラム	136
5.3	変数選択のプログラム	155
6	主成分分析	161
6.1	主成分分析とは	162
6.2	例題	165
6.3	主成分分析のプログラム	172
6.4	実対称行列の固有値・固有ベクトル	175
6.5	固有値・固有ベクトルを求めるプログラム	177
7	重回帰分析, 主成分分析の応用	181
	——時系列データ, 非線形回帰, 異常値の検出など	
7.1	原点を通る回帰式	182
7.2	時系列データ	183
7.3	成長曲線のあてはめ	194

7.4 非線形回帰	204
7.5 異常値の検出	218
付録 FORTRAN 書法	227

1

基本統計量

この章では基本的な統計量を計算し、それらを印刷するサブルーチンを紹介する。

もっぱらプログラムを理解するという立場で読み進んでいただきたい。

第2章以降への導入の準備と FORTRAN の復習が本章の主な目的であるが、行列の印刷(MATPRI)、行列の積(MATMUL)、掃出演算(SWEEP)の各サブルーチンについては、以降の各章で用いられるほか、利用頻度が高いため、汎用性を持たせるくふうをこらしており、かつ簡潔なサブルーチンとしてまとめあげたので、統計計算以外でも活用できるよう十分理解しておくことをすすめる。

1.1 基本統計量とそのプログラム

表 1.1 の形をしたデータにはよくお目にかかる。

表 1.1 データ行列

サンプル \ 変数						
	x_1	x_2	$\cdots$	x_j	$\cdots$	x_p
1	x_{11}	x_{12}	$\cdots$	x_{1j}	$\cdots$	x_{1p}
2	x_{21}	x_{22}	$\cdots$	x_{2j}	$\cdots$	x_{2p}
$\vdots$						
i	x_{i1}	x_{i2}	$\cdots$	x_{ij}	$\cdots$	x_{ip}
$\vdots$						
n	x_{n1}	x_{n2}	$\cdots$	x_{nj}	$\cdots$	x_{np}

一般には次のように表現でき、これをデータ行列と呼ぶ。

$$X = (x_{ij}) \quad (i=1, 2, \dots, n; j=1, 2, \dots, p) \quad (1.1)$$

ここで x_{ij} は第 j 変数(特性値)についての、 i 番目のサンプルの測定値。 n はサンプル数、 p は変数の数

このデータ行列が与えられたとき、基本的な統計量として次のようなものがある。

平均値

$$\bar{x}_j = \sum_{i=1}^n x_{ij} / n$$

最大値

$$x_{\max, j} = \max_i (x_{ij})$$

最小値

$$x_{\min, j} = \min_i (x_{ij})$$

なまの平方和・積和行列

$$X'X = \left(\sum_{i=1}^n x_{ij} x_{ij'} \right)$$

ここで、 X' は X の転置行列を表す。

平方和・積和行列

$$S = (S_{jj'}) = \left(\sum_{i=1}^n x_{ij} x_{ij'} - \frac{\sum_i x_{ij} \sum_i x_{ij'}}{n} \right)$$

分散・共分散行列

$$V = (V_{jj'}) = (S_{jj'} / (n-1))$$

標準偏差

$$s = (s_j) = (\sqrt{V_{jj}})$$

相関行列

$$R = (R_{jj'}) = (S_{jj'} / \sqrt{S_{jj} S_{j'j'}})$$

これらの統計量を1.2節以降で述べるサブルーチンを組み合わせて、計算・印刷するプログラム1.1を示す。

このプログラムでは6つのサブルーチンを利用している。これらサブルーチンの機能については1.2節から1.7節まで順に述べるが、統計計算では基本的なサブルーチンの組合せによって簡潔でわかりやすいプログラムを作成できる一例としてプログラム1.1を掲げた。

2章以降ではこの基本統計量のプログラムで求めた平方和・積和行列または相関行列を初期行列として、重回帰分析、主成分分析等を行うようにしているので、サブルーチンの機能と照らし合わせて十分このプログラムを理解していただきたい。

また2章以降に現れる手法の詳細と応用例については、奥野ほか『多変量解析法』[†]を参照されたい。ただし、本書ではプログラムと記号を合わせるため、記法を次のように変えてある。

サンプルを表す添字 $\alpha \rightarrow i$

変数の番号を表す添字 $i \rightarrow j$ または j'

[†] 奥野忠一，芳賀敏郎，久米 均，吉澤 正：『多変量解析法』，日科技連出版社，1971。

奥野忠一，芳賀敏郎，矢島敬二，奥野千恵子，橋本茂司，古河陽子：『統多変量解析法』，日科技連出版社，1976。

プログラム 1.1 基本統計量

```

001      DIMENSION FMT(20),TITLE(20)
002      DIMENSION X(50,15),S(15,15),R(15,15)
003      DIMENSION AVEX(15),SIGX(15),XMIN(15),XMAX(15)
004      DATA LX,LS/50,15/

C
005      READ (5,500) TITLE
006      WRITE (6,600) TITLE
007      WRITE (6,610)
008      READ (5,510) II,JJ
009      JJ1 = JJ+1
010      READ (5,500) FMT
011      DO 110 I=1,II
012          READ (5,FMT) (X(I,J),J=1,JJ)
013          X(I,JJ1) = 1.0
014      110 CONTINUE
015      CALL MATPRI (8HDATA      ,X,1,LX,II,JJ)
016      CALL MATMUL (X,LX,1,X,1,LX,S,1,LS,JJ1,II,JJ1)
017      CALL SWEEP (S,LS,JJ1,JJ1)
018      DF = II-1
019      DO 130 J=1,JJ
020          AVEX(J) = S(JJ1,J)
021          SIGX(J) = SQRT(S(J,J)/DF)
022          CALL MINMAX (X(1,J),II,XMIN(J),XMAX(J))
023      130 CONTINUE
024      CALL VECPRI (8HMEAN      ,AVEX,1,JJ)
025      CALL VECPRI (8HSIGMA    ,SIGX,1,JJ)
026      CALL VECPRI (8HMINIMUM  ,XMIN,1,JJ)
027      CALL VECPRI (8HMAXIMUM  ,XMAX,1,JJ)
028      CALL CORREL (S,LS,JJ,R,LS)
029      CALL MATPRI (8HR MATRIX,R,1,LS,JJ,JJ)

C
030      500 FORMAT (20A4)
031      510 FORMAT (16I5)
032      600 FORMAT (1H1,10X,20A4)
033      610 FORMAT (1H0,31HDATA AND FUNDAMENTAL STATISTICS)
034      STOP
035      END

```

〈プログラム 1.1 の説明〉

行番号

001~003 配列宣言

FMT(20) データ行列 X の読込のための可変 FORMAT

TITLE(20) データの表題

X(50,15) データ行列

S(15,15) 平方和・積和行列

R(15,15) 相関行列

AVEX(15), SIGX(15) 平均値と標準偏差

XMIN(15), XMAX(15) 最小値と最大値

004 LX, LS は上記配列で扱えるサンプル数, 変数の数+1の最大
数を表す。

大きなデータ行列を扱いたいときは上記配列宣言とこの DATA
文を変えるだけでよい。

005~007 表題の読込と印刷

008 サンプル数 $n(=II)$, 変数の数 $p(=JJ)$ の入力009 $JJ1 \leftarrow p+1$ (1.3 節参照)

010~014 可変 FORMAT の入力とそれに従ったデータの入力(章末補注)

015 データ行列の印刷

016 なまの平方和・積和行列の計算

017 平方和・積和行列の計算

018 自由度(DF) $\leftarrow n-1$

019~023 各変数について平均, 標準偏差, 最小値, 最大値の計算

024~027 サブルーチン VECPRI による平均値, 標準偏差, 最小値, 最
大値の印刷

028 相関行列の計算

029 相関行列の印刷

030~033 入力および出力用の FORMAT

プログラム 1.1 による出力例をアウトプット 1.1 に示しておく。

アウトプット 1.1 プログラム 1.1 でのアウトプット例

EXAMPLE OF MULTIVARIATE DATA

DATA AND FUNDAMENTAL STATISTICS

DATA

	1	2	3	4	5	6
1	8.2500	9.8900	10.8800	8.7300	8.8700	8.3900
2	10.7700	11.4700	11.0200	12.3300	11.8800	10.9300
3	10.9400	9.8200	10.2700	10.9000	10.3600	11.4600
4	9.3400	9.0000	8.8600	8.2800	8.0000	8.4400
5	10.4600	9.9600	8.6200	10.5400	9.8800	10.1400
6	11.5400	10.1700	9.8300	11.9900	11.3700	11.2700
7	9.4200	10.2300	11.4400	10.2900	9.5800	9.6700
8	9.7800	9.9300	10.7400	10.0100	9.8500	10.3900
9	11.3500	10.4700	12.2900	11.8900	10.9200	12.4900
10	9.4100	9.4700	10.3600	9.3600	8.9600	9.9300
11	10.8300	8.1700	9.1400	8.9700	7.5700	9.4900
12	9.6100	10.4600	10.2300	10.1200	10.4100	9.8800
13	9.9500	10.5300	10.9400	10.1300	10.5600	11.5800
14	11.3200	10.3900	11.7500	11.9800	11.5000	12.6800
15	10.3400	9.7100	9.8400	10.3200	9.1800	10.4100
16	8.3100	9.5300	11.4100	9.4900	9.6500	8.2900
17	8.7700	11.5800	10.8600	9.6000	9.9900	10.2200
18	10.2600	8.0500	8.7800	8.8300	8.9600	8.6400
19	7.0500	8.6900	8.9900	7.0400	8.2800	5.9300
20	8.6100	9.3200	8.5600	7.8200	8.5500	7.7800
21	10.2600	8.0300	7.9200	8.4900	8.1400	7.7800
22	8.3400	9.4200	9.6300	8.3500	8.7200	7.3100
23	9.6900	9.6000	10.2600	8.9800	10.1000	8.4800
24	8.6600	8.4400	8.4200	7.8100	7.8300	7.4900
25	8.0200	10.4100	10.3100	8.2400	8.7300	8.0000
26	9.5200	10.0100	9.7000	8.2000	9.4300	10.1500
27	8.8100	8.8500	8.9000	7.5200	8.0100	6.2600
28	9.3300	9.0600	9.0100	8.8800	8.1900	8.1900
29	9.5600	8.2100	8.4100	7.8900	8.0000	8.7000
30	9.3600	9.5500	9.1500	8.3800	8.7000	9.6400
MEAN	9.5953	9.6140	9.8840	9.3786	9.3390	9.3337
SIGMA	1.0879	0.9230	1.1364	1.4340	1.1761	1.6930
MINIMUM	7.0500	8.0300	7.9200	7.0400	7.5700	5.9300
MAXIMUM	11.5400	11.5800	12.2900	12.3300	11.8800	12.6800

R MATRIX

アウトプット 1.1 つづき

R MATRIX

	1	2	3	4	5	6
1	1.0000	0.1711	0.2043	0.7687	0.5599	0.7871
2	0.1711	1.0000	0.7515	0.6491	0.7898	0.6149
3	0.2043	0.7515	1.0000	0.6602	0.7187	0.6190
4	0.7687	0.6491	0.6602	1.0000	0.8856	0.8691
5	0.5599	0.7898	0.7187	0.8856	1.0000	0.7840
6	0.7871	0.6149	0.6190	0.8691	0.7840	1.0000

1.2 データ行列の印刷(MATPRI)

統計計算ではデータ行列，平方和・積和行列，分散・共分散行列，相関行列等を頻繁に扱い，またそれらの行列の要素の値を分析することによって有用な情報が得られることが多い．そのためこれら主要な行列を印刷することは統計解析のプログラムにとって基本となる．

サブルーチンMATPRI(プログラム 1.2)は図 1.1 の様式で行列を印刷するプログラムである．

このプログラムでは行列 A を 1 次元配列で扱っている．プログラムのわかりやすさは 2 次元の整合配列を使ったほうが優るが，同じアルゴリズム(計算手続き)で転置行列も印刷できるようにふうしたため 1 次元配列としている．

この方法は次の 1.3 節のサブルーチンMATMUL(行列の乗算)でも利用しているので基本的な考え方について述べておこう．

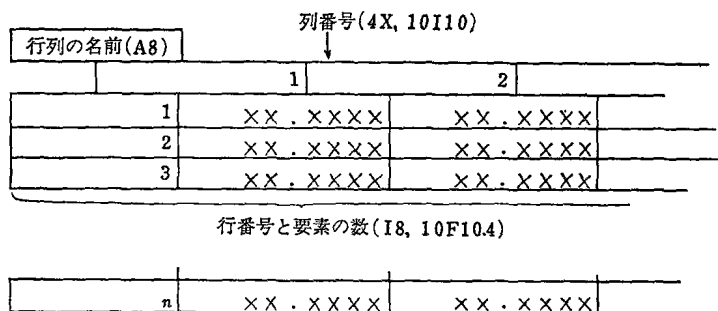


図 1.1 MATPRIの印刷様式