PACIFIC SOUTHWEST
FOREST & RANGE EXPERIMENT STATION
Berkeley, California 1965

U. S. FOREST SERVICE
RESEARCH NOTE PSW-92

WALLY1....

A Large, Principal Components Regression Program with Varimax Rotation of the Factor Weight Matrix

JAMES R. WALLIS

ABSTRACT: Written in Fortran IV and MAP, this computer program can handle up to 120 variables, and retain 40 principal components. It can perform simultaneous regression of up to 40 criterion variables upon the varimax rotated factor weight matrix. The columns and rows of all output matrices are labeled by six-character alphanumeric names. Data input can be from punch cards or BCD tape.

Hydrologists solve problems in prediction by using many kinds of multivariate statistical techniques. To fill our particular research needs, we wanted a large, easy-to-use principal component regression program combined with varimax rotation of the factor weight matrix. The result was a computer program, written in Fortran IV and MAP, that we call WALLY1. The program is now operational at the University of California Computer Center, Berkeley, on an IBM 7094 computer using its non-standard monitor.

WALLY1 uses many sophisticated computer programs borrowed from the works of others. To print all matrices and vectors, it uses Russell's[1] CINEMA package; "debugging" was aided by his BULTER and ODS4 (Object Deck Structure Analysis) programs. Eigenvalues and eigenvectors are calculated by a modified version of the Meredith[2]-Matula[3] subroutine HOW, and ranking of factor weights is by Krasnow's[4] ORDER, while tape handling is assisted by Claxton's[5] WINDER program. Dr. William Meredith provided matrix formulation for WALLY1; John Bauer[6] and Robert M. Russell provided programing guidance.

---

[1]Robert M. Russell, chief programer, Pacific Southwest Forest and Range Experiment Station, Berkeley, Calif.

[2]William Meredith, professor of psychology, University of California, Berkeley.

[3]Dave Matula, programer, University of California, Berkeley.

[4]Ellie S. Krasnow, programer, Department of Psychology, University of California, Berkeley.

[5]H. Dean Claxton, timber quality scientist, Pacific Southwest Forest and Range Experiment Station, Berkeley, Calif.

[6]John Bauer, chief programer, Department of Psychology, University of California, Berkeley.

DESCRIPTION OF DECK SETUP

An example of the deck setup for WALLY1 is shown in figure 1. Information on the problem card, title, variable format, name and data cards is given below. It is followed by a list of the printed output. An example of input data for a typical small problem is given in figure 2.

PROBLEM CARD (f)

Columns	Variable		Description
1 - 4	NØBS		Number of observations.
5 - 8	NVAR		Number of predictor (X) plus criteria (Y) variables; must be equal to or less than 120.
9 -12	NVARX		Number of predictor (X) variables; must be equal to or less than 120.
13 -16	NVARY		Number of criterion (Y) variables; must be equal to or less than 40.
17 -20	NTIT		Number of alphanumeric title cards.
21 -24	NFORM		Number of variable format cards used to describe input data array; maximum:5.
25 -28	NTAPE	O	Data is by card input.
		NX	Data should be read from file NX of logical tape unit 5 in BCD.
29 -32	NSTØRE	O	Data are not to be stored on tape. (Note: NSTØRE = O when NTAPE ≠ O.)
		KX	Data are to be stored in file KX of logical tape unit 5 in BCD.
33 -36	NCØN	O	No additional problems to be processed.
		1	Another problem follows.
37 -40	NCØR	O	Correlation matrix, means, and standard deviations are generated from raw data.
		1	Correlation matrix, means, and standard deviations are input using BCTRY Gist format.[7]

[7] User's Manual, BCTRY Computer System of Cluster and Factor Analysis, Project CAP (Cluster Analysis Programs), University of California, Berkeley, July 31, 1965.

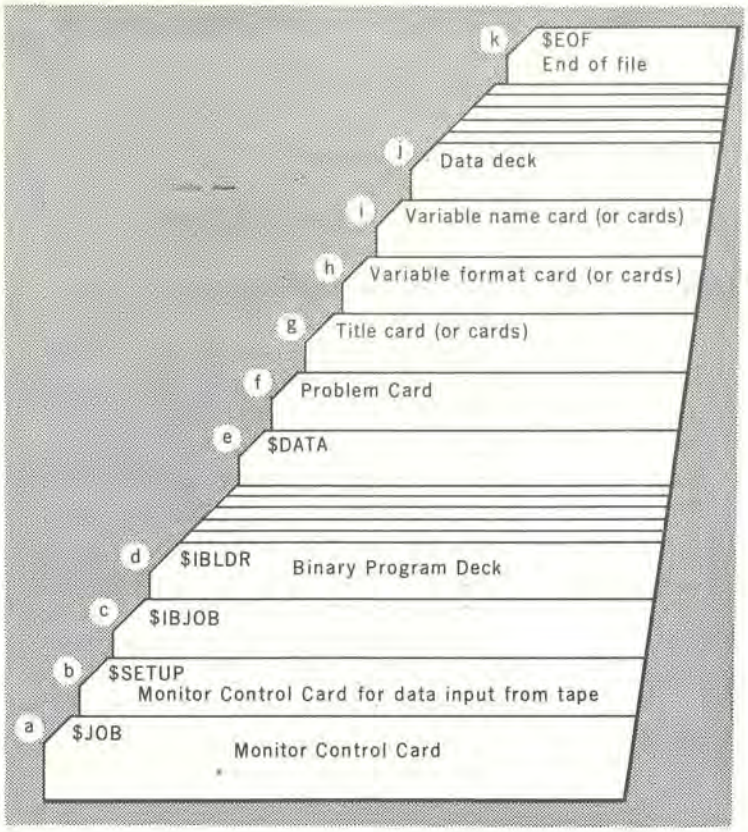

Figure 1.--Deck setup for WALLY1, cards f through j are repeated as needed.

Figure 2.--Example of input data for typical small problem.

```
$DATA
  60  15  13   2   8                          1   1   1   1   1       4       6
 DATA FOR THIS PROBLEM WERE TAKEN FROM --MULTIVARIATE STATISTICAL METHODS IN
 HYDROLOGY-- A COMPARISON USING DATA OF KNOWN FUNCTIONAL RELATIONSHIP-- BY JAMES
 R. WALLIS  WATER RESOURCES RESEARCH VOL. 1,4,P 447-461 (SEE EQUATION 8 AND
 TABLE 4 ). FOR THE UC BERKELEY IBM 7040-7094 DCS THIS TOOK 0.36 MINUTES FOR
 EXECUTION. LOAD TIME FOR THE WALLY1 OBJECT DECK DURING AUGUST AND SEPTEMBER
 1965 WAS HIGHLY VARIABLE (FROM UNDER 2 TO OVER 8 MINUTES) FOR IDENTICAL DECKS.
 INSTALLATIONS WITHOUT DCS SHOULD BE ABLE TO ESTIMATE HOW LONG A JOB MAY TAKE TO
 LOAD AND EXECUTE, BUT FOR BERKELEY THIS IS NOT POSSIBLE.
VNAMS1      1   0
  15
H       HH       2KROH    2KRIH    D       DD       DDIAGO  DDIAGI  RO
KRORO   DIAGO    RI       KRIRI    DIAGI   W
CORRM1      1 1
(9F8.6)
  15  60
  962321
  838235   795783
  681559   676672   739713
   77335    99230   115072   107617
  125556   162450   162535   142411   965436
  123926   168708   321460   245190   912074   892725
  333933   361591   420429   572959   785157   763153   829568
 -154344  -122947   353070   188396    55876    75562   401202   214603
 -156757  -121261   339431   157620    38370    50960   395471   203028   982505
  102410   150271   537685   348355    66614   102964   435380   298254   956526
  949726
  114116   125634   300631   719237   102190    97666   257299   563517   402652
  371088   425452
  110500   116482   296299   694261    75196    75288   240096   579343   401349
  394656   427270   957862
  495742   501478   579758   846005   112960   133260   272904   630327   288961
  272392   422109   902219   889151
  410560   401411   610741   105231   437708   455253   596252   342243   351494
  375769   440350  -126258  -158330    60459
MEANS1      1 1
(9F8.4)
  15  60
  5.2823 36.6881234.3786131.2229  0.5157  0.3373  8.2308  5.2395  7.2450
178.9732 15.8302   3.8084 61.7360  9.9184298.1197
STDEV1      1 1
(9F8.4)
  15
  2.9640 29.9741154.1757118.0292  0.2671  0.2803  4.9374  3.8012  2.1162
 83.4080  3.7375   2.2687 61.2596  4.1129413.1537
   3   4   6   8   9  14
```

PROBLEM CARD (f), Continued

Columns	Variable		Description
41 -44	NAMES	O	Six-character alphanumeric names for each variable are input using the format (9[A6, 2X]).
		1	Variable names are input using BCTRY Gist format.
45 -48	NPO	O	No cards punched.
		1	Correlation matrix punched using BCTRY Gist format.
49 -52	NP1	O	No cards punched.
		1	Variable names, means, and standard deviations are punched using BCTRY Gist format.
53 -56	NP2	O	No cards punched.
		1	Regression coefficients and rotated factor weight matrix punched on cards.
57 -60	NRØØ		Proportion of explained variance for components; program substitutes 0.9950 for a zero or blank.
61 -64	M		Number of eigenvectors to retain if known; otherwise leave blank.
65 -68	NVAL		Number of iterations to be used in varimax rotation of the factor weight matrix. Program sets zero or blank to 10, which is probably satisfactory for all but large matrices.
69 -72	NSLEP		Number of variables to be dropped from the analysis. It is used when the input of means, standard deviations, and correlation matrix contain variables from a previous run not required in the present run. If this option is used, then the indexes of the NSLEP dropped variables arranged in ascending order should follow the standard deviations or raw data (Format 20I4).

TITLE CARD OR CARDS (g)

These are alphanumeric problem identification cards. All 80 columns are usable, and there is no limit on number of cards.

VARIABLE FORMAT CARD OR CARDS (h)

Description of raw data fields as punched on cards or stored upon BCD tape is given on this card. Only X and F type fields may be specified. Statement should start and end with a parenthesis. It may be as long as 80 columns on five cards. (Note: All predictor (X) variables must precede the criteria (Y) variables in the array.)

NAME CARDS (i)

Nine alphanumeric six-character variable names are allowed per card (Format 9[A6, 2X]). Predictor variable names precede the criterion variable names. Sufficient cards for all NVAR names must be used in the deck setup.

DATA CARDS (j)

Either raw data cards as described by the variable format cards or BCTRY Gist input format of correlation matrix (CORRM1), means (MEANS1), standard deviations (STDEV1), and indexes of any variables which are to be suppressed from the analysis can be used as data cards.

PRINTED OUTPUT

The printed output consists of:

(a) Header card.
(b) Title cards (if used).
(c) Format cards (if used).
(d) Vector of variable means.
(e) Vector of variable standard deviations.
(f) Correlation matrix.
(g) Vector of eigenvalues.
(h) Matrix of eigenvectors.
(i) Principal component factor weight matrix.
(j) Varimax rotated factor weight matrix.
(k) Variable names in descending order of their absolute factor loadings on each rotated dimension.
(l) Factor loadings for each variable upon all dimensions arranged in descending absolute order.
(m) Matrix of square roots of factor contribution to the explained variance.
(n) Matrix of factor contributions to the explained variance for each criterion variable (Y) and dimension.
(o) Vector of standard errors.
(p) Matrix of standardized regression coefficients.
(q) Matrix of regression coefficients for original variables.
(r) Vector of regression constants.

SOURCE OF PROGRAMS

Source deck listing of the main program of WALLY1 and most of the subroutines may be obtained as an interlibrary loan from the Librarian, Pacific Southwest Forest and Range Experiment Station, P. O. Box 245, Berkeley, Calif., 94701. The remaining subroutines are available through the University of California, Computer Center Library, Berkeley, Calif., 94704.

The Author

JAMES R. WALLIS is studying problems in watershed hydrology and sedimentation in wildlands. Native of Montreal, P.Q., he earned a bachelor's degree in forestry at the University of New Brunswick (1950), a master's degree at Oregon State University (1956), and a doctorate at the University of California (1965). Before joining the Station staff, he worked in private industry for 5 years and taught forestry at Montana State University for 3 years.