

GCM: An interactive program for the analysis of time-structured data

M. P. F. BERGER

University of Tilburg, 5000 LE Tilburg, The Netherlands

The analysis of longitudinal time-structured data is often referred to as growth curve analysis, since a curve is fitted to the data. Several procedures have been proposed to estimate curve parameters and to test general linear hypotheses (Khatri, 1966; Potthoff & Roy, 1964; Rao, 1965, Tubbs, Lewis, & Duran, 1975). Timm (1980) reviews these procedures.

Potthoff and Roy (1964) developed a generalized MANOVA model:

$$E(Y^*) = XBP, \quad (1)$$

where Y^* is an n by q data matrix, with population covariance matrix Σ , B is a g by p matrix of unknown parameters, and X and P are n by g and p by q across and within individual design matrices, respectively ($p \leq q$, $g \leq n$). The general multivariate theory can be used by reducing Model 1 to the standard MANOVA model:

$$E(Y) = XB. \quad (2)$$

This reduction is obtained by:

$$Y = Y^*G^{-1}P'(PG^{-1}P')^{-1}, \quad (3)$$

where G is any q by q symmetric nonsingular matrix. The choice of G , however, is arbitrary. If the number of parameters equals the number of time points ($p = q$), there is no need to choose G (Bock, 1975; Finn, 1976; Finn & Mattson, 1978). If $p < q$, the choice of G becomes important. A simple choice of G is $G = I$.

To avoid the choice of G , Khatri (1966) and Rao (1965, 1966) proposed an alternative reduction of Model 1 to the multivariate analysis of covariance (MANCOVA) model:

$$E(Y|W) = XB + WQ, \quad (4)$$

where W is an n by $(q - p)$ matrix of covariates and Q is a $(q - p)$ by p matrix of unknown regression coefficients. This model uses the last $q - p$ parameters for covariance adjustments.

Another procedure was given by Tubbs et al. (1975), who used a maximum-likelihood estimator of B (Khatri, 1966) for unconditional tests of the general linear hypothesis.

Although these procedures are asymptotically equivalent (Kleinbaum, 1973), there are differences in variance

of estimates, in relative power of the tests, and in width of confidence intervals, especially for small sample sizes.

The program described here can fit orthogonal or nonorthogonal polynomials of degree $p - 1$. Other transformations, such as a Fourier transform (Zerbi & Jones, 1980), can also be used. The question of whether the average curves can be represented by a $p - 1$ degree polynomial is answered by testing the null hypothesis that the last $q - p$ parameters equal zero.

Assuming an adequate description of the data with p parameters, the program estimates the parameters for each procedure and performs tests for coincidence and parallelism of the curves. Four multivariate test statistics are computed: Wilks' likelihood (with chi-square approximation), Hotelling's trace, Pillai's trace, and Roy's largest root criterion with the largest root distribution parameters.

Finally, the group means are estimated using p parameters, and the corresponding residuals are obtained.

This program can analyze three different designs. Design 1: $S(G)$ by T by V —Subjects (S) are nested within groups (G) and measured on different treatments or variables (V) over time (T). Coincidence and parallelism of group curves are tested. Design 2: S by T by V —Coincidence and parallelism of curves of different treatments or variables are tested. Design 3: S by $T(V)$ —The same hypotheses as in Design 2 are tested. The parameters, however, are estimated by taking into account that time points are nested within the treatments or variables.

Description. The program consists of a main routine and 24 subroutines. The subroutines READIN and CORFMT control the data reading and format specification. ORNOPL (Emerson, 1965), POLYN, and VANDER generate the orthogonal or nonorthogonal polynomial transformation matrix, P . FIT and HYPOGEN generate contrast matrices for hypothesis testing. COVAR computes group and treatment or variable means and the sample estimate of Σ . KHATRI estimates the parameters for each procedure, and SSCP computes the hypothesis and error sums of squares and products (SSCP) matrices, S_h and S_e . Graphical display of group means is given by PLOT and SORT (McNeil, 1977). The eigen structure of $S_h S_e^{-1}$ is solved in NROOT and EIGEN (International Business Machines, 1970), and the four test statistics and largest root distribution parameters are computed in CRITER. WRMATRIX and WRVECTOR are printing routines, and MATINV (Cooley & Lohnes, 1971) is a matrix inversion routine. CHITEST and ZTEST approximate chi-square probabilities. Supplementary operations are performed by MPROD, TRPOSE, DCOPY, and GENER. The subroutines PLOT, NROOT, EIGEN, and MATINV have been altered to conform to the main routine.

Input. The program receives input from a keyboard. This input consists of user replies to Teletype prompts (e.g., "ENTER DESIGN NUMBER:", "ENTER NUMBER OF TIME POINTS:"). The interactive nature of the program allows the user to conduct more than one analysis per execution of the program.

The program requests a specification of (1) the design number, (2) the number of time points, (3) the number of treatments/variables, (4) the number of groups and subjects within groups (only for Design 1), and (5) the name of the data file and the format specification. Data can be read either FREE formatted (default), in which scores in the data file are separated by blanks and/or a comma, or with an F-type variable format enclosed in parentheses.

As soon as these parameters are entered, they are displayed for verification and can be modified. Then the user is asked to specify the polynomial transformation matrix, P: orthogonal (code = 1) or nonorthogonal (code = 0). Other transformation matrices can be entered via file (code = 2).

If the goodness-of-fit test and the coincidence and parallelism tests are wanted, the number of parameters $p \leq q$ is required.

Output. The output is on terminal screen (default), but it can also be directed to a line printer. The output includes (1) a listing of the input parameters, (2) the data for the first five cases, (3) the transformation matrix, P, (4) the sample estimate of Σ and the sample correlation matrix, (5) the correlations between the transformed (polynomial) variables, (6) the means and their plots for each group and variable, (7) the estimated (polynomial) parameters, (8) the hypothesis SSCP matrix, S_h , with the corresponding error SSCP matrix, S_e , and (9) the four multivariate test statistics with the largest root distribution parameters. The estimated group means corresponding with each treatment or variable and residuals are also reported.

Restrictions and Error Messages. The program simply corrects input errors from keyboard by asking the same input statement again. The following error messages may be encountered and will terminate the execution.

PROGRAM DIMENSIONS EXCEEDED: The data matrix may not have more than 500 rows and/or 50 columns. This restriction can easily be changed, however, by modifications of the array sizes. The PARAMETER statement in the main routine specifies these array sizes.

MODEL NOT OF FULL RANK: The sample estimate of Σ is not of full rank. The dependent measures are perhaps linear combinations of each other or $n - g$ is less than q .

INPUT CONVERSION ERROR: A nonnumeric character is found in the data, or a wrong F-type variable format is used.

END OF FILE DURING READ: The input file does not contain enough scores.

ERROR OPENING INPUT DATA FILE: The program cannot find a file with the specified file name.

FORMAT SPECIFICATION ERROR: The F-type variable format must contain a period, the character F, and at least one of the digits 0-9. The specification begins with a left parenthesis and closes with a right parenthesis.

Computer and Language. The main program and some of the subroutines are written in an updated language, known as FORTRAN 77, formally standardized in 1978 by ANSI. The other subroutines consist of standard FORTRAN IV statements.

The program was developed and tested on a VAX-11/780 (DEC) computer. The only machine-dependent statements are the file manipulations. There should be little trouble converting the program to other computer systems with a FORTRAN 77 compiler. No external functions or special library routines are necessary. In order to maintain accuracy, all computations are done in DOUBLE PRECISION. The total allocated space is approximately 300 KB. The source listing of main program and subroutines contains 1,200 statements. The data set reference numbers for input data file, terminal, and line printer are 1, 5, and 6, respectively.

Availability. Source listings, test data, and output listings can be obtained at no cost from the author: Martijn P. F. Berger, Psychological Laboratory, Katholieke Hogeschool, Postbus 90153, 5000 LE Tilburg, The Netherlands.

VAX-11/780 users can send a floppy diskette (RX01 compatible, single density, single sided) to obtain soft-copy versions of the program.

REFERENCES

- BOCK, R. D. *Multivariate statistical methods in behavioral research*. New York: McGraw-Hill, 1975.
- COOLEY, W. W., & LOHNES, P. R. *Multivariate data analysis*. New York: Wiley, 1971.
- EMERSON, P. L. A FORTRAN generator of polynomials orthogonal over unequally spaced and weighted abscissas. *Educational and Psychological Measurement*, 1965, 25, 867-871.
- FINN, J. D. *MULTIVARIANCE: Univariate and multivariate analysis of variance, covariance and regression* (Version 5). Chicago: National Educational Resources, 1976.
- FINN, J. D., & MATTSON, I. *Multivariate analysis in educational research; Applications of the MULTIVARIANCE program*. Chicago: National Educational Resources, 1978.
- INTERNATIONAL BUSINESS MACHINES. *IBM 360 scientific subroutines package (360-CM-03X). Programmer's manual (GH20-0205)*. Author, 1970.
- KHATRI, C. G. A note on a MANOVA model applied to problems in growth curve. *Annals of the Institute of Statistical Mathematics*, 1966, 18, 75-86.
- KLEINBAUM, D. G. A generalization of the growth curve model which allows missing data. *Journal of Multivariate Analysis*, 1973, 3, 117-124.
- MCNEIL, D. R. *Interactive data analysis; A practical primer*. New York: Wiley, 1977.
- POTTHOFF, R. F., & ROY, S. N. A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, 1964, 51, 313-326.

- RAO, C. R. The theory of least squares when parameters are stochastic and its application to the analysis of growth curves. *Biometrika*, 1965, **52**, 447-458.
- RAO, C. R. Covariance adjustment and related problems in multivariate analysis. In P. R. Krishnaiah (Ed.), *Multivariate analysis*. New York: Academic Press, 1966.
- TIMM, N. H. Multivariate analysis of variance of repeated measurements. In P. R. Krishnaiah (Ed.), *Handbook of statistics* (Vol. 1). Amsterdam: North-Holland, 1980.
- TUBBS, J. D., LEWIS, T. O., & DURAN, B. S. A note on the analysis of the MANOVA model and its application to growth curves. *Communications in Statistics*, 1975, **4**, 643-653.
- ZERBI, G. O., & JONES, R. H. On application of growth curve techniques to time series data. *Journal of the American Statistical Association*, 1980, **75**, 507-509.

(Accepted for publication January 20, 1982.)