Factor analysis in BASIC for Minicomputers

JOHN C. SCHMITT

*State University of New York at Binghamton
Binghamton, New York 13901*

A common procedure in the analysis of both large and small data sets involves calculating intercorrelation matrices and subsequently factor analyzing these matrices. These procedures are useful in a variety of situations, including both hypothesis testing and exploratory applications. (Refer to Guertin & Bailey, 1970; Horst, 1965; and Nunnally, 1967, for information on the nature and applications of factor analysis.)

A common feature of factor analysis programs is their development in languages such as FORTRAN for use on large computer systems. In many instances, less formal and less expensive access would be desirable; laboratory minicomputers afford that accessibility.

The program described here is designed to perform preliminary factoring of matrices, later to be rotated and further explored using larger program packages (e.g., Nie, Hull, Jenkins, Steinbrenner, & Bent, 1975). Its main feature is the small amount of core storage used to calculate the correlation matrix. Factoring is through the centroid method; although the method of principle axes is more common, the iterative

mechanics and matrix inversion prohibit its use on small computers. The centroid method, in addition to providing almost identical extraction of at least the first factor, requires only three versions of the correlation matrix (for the calculation of residual matrices) and a variables by factors loading matrix. The program also allows input of the correlation matrix directly. This feature not only saves time when the matrix is available, but also allows reanalysis of the same matrix with communalities placed on the diagonal (obtained from initial factoring). With data input only, unities are placed on the diagonal.

Development and limitations. This program was developed on an Interdata Model 70 computer. Size limitations are independent of the number of subjects and arise from the correlation and loading matrices which must be dimensioned. It is unlikely that more than one-third as many factors as variables would be desired. With approximately 16.8K bytes available after loading the BASIC interpreter, the size limitation for the machine is about 31 variables and 10 factors. Minor modifications (available from the author) allow calculation of the correlation matrix alone for a much larger number of variables.

Availability. Reprints of this article and listings of the program may be obtained from the author at no cost. Requests should be addressed to John C. Schmitt, Department of Psychology, State University of New York at Binghamton, Binghamton, New York 13901.

PROGRAM LISTING

```
10 ;"INPUT A 1 IF YOU WISH TO ENTER THE CORRELATION MATRIX DIRECTLY"
20 INPUT Z
30 IF Z=1 THEN GO TO 1170
40 ;"INPUT NUMBER OF SUBJECTS, VARIABLES, AND DESIRED FACTORS."
50 INPUT S,V,F
60 DIM D(1,V), C(V), L(V,F), L1(1,V), L2(V,1), R(V,V), R1(V,V), R2(V,V), W(V)
70 MAT L1=(0)
80 MAT R=(0)
90 FOR K=1 TO S
100 ;"INPUT SUBJECT";K
110 MAT INPUT ON (6)D
120 FOR I=1 TO V
130 LET L1(1,I)=L1(1,I)+D(1,I)
140 FOR J=1 TO I
150 LET R(I,J)=R(I,J)+(D(1,I)*D(1,J))
160 NEXT J
170 NEXT I
180 NEXT K
190 ;"VARIABLE", "TOTAL", "MEAN", "VARIANCE", "STAND. DEV"
200 FOR I=1 TO V
210 FOR J=1 TO I
220 LET R(I,J)=(R(I,J)-((L1(1,I)*L1(1,J))/S))
230 NEXT J
240 LET W(I)=R(I,I)
250 ;I, L1(1,I), L1(1,I)/S, R(I,I)/(S-1), SQR(R(I,I)/(S-1))
260 NEXT I
270 FOR I=1 TO V
280 FOR J=1 TO I
290 LET R(I,J)=R(I,J)/(SQR(W(I))*SQR(W(J)))
```

```

300 LET R(J,I)=R(I,J)
310 NEXT J
320 NEXT I
330 ;"THE CORRELATION MATRIX IS--"
340 MAT PRINT R
350 MAT R2=R
360 REM *****CONTROLS ANALYSIS; CALLS SUBROUTINES*****
370 FOR K=1 TO F
380 GOSUB 610
390 GOSUB 430
400 GOSUB 540
410 NEXT K
420 GOTO 840
430 REM *****EXTRACTS FACTOR; ASSIGNS IT TO THE LOADING MATRIX*****
440 LET C1=0
450 FOR J=1 TO V
460 LET C(J)=C(J)+R2(J,J)
470 LET C1=C1+C(J)
480 NEXT J
490 FOR I=1 TO V
500 LET L(I,K)=(C(I)*W(I))/SQR(C1)
510 LET L1(I,J)=L(I,K)
520 NEXT I
530 RETURN
540 REM *****SUBTRACTS KTH FACTOR CROSS-PRODUCTS FROM THE K-1TH
550 REM *****RESIDUAL MATRIX TO GET THE KTH RESIDUAL MATRIX*****
560 MAT L2=TRN(L1)
570 MAT R1=L2*L1
580 MAT R2=R-R1
590 MAT R=R2
600 RETURN
610 REM *****CHECKS COLUMN SUMS FOR NEGATIVE VALUES; CALLS REFLECTION
620 REM *****SUBROUTINE IF REFLECTION IS NECESSARY*****
630 FOR J=1 TO V
640 LET W(J)=1
650 NEXT J
660 FOR M=1 TO 1000
670 LET Z=0
680 LET Z1=0
690 FOR J=1 TO V
700 LET C(J)=0
710 FOR I=1 TO V
720 LET C(J)=C(J)+R2(I,J)
730 NEXT I
740 LET C(J)=C(J)-R2(J,J)
750 IF C(J)<Z THEN LET Z1=J
760 IF C(J)<Z THEN LET Z=C(J)
770 NEXT J
780 IF Z1>0 THEN GOSUB 1090
790 IF Z1>0 THEN NEXT M
800 RETURN
810 REM *****PRINTS FACTOR MATRIX; CALCULATES AND PRINTS VARIANCE
820 REM *****IN EACH VARIABLE ACCOUNTED FOR BY ALL FACTORS (H-SQUARE)
830 REM *****AND VARIANCE IN DATA ACCOUNTED FOR BY EACH FACTOR*****
840 ;"THE FACTOR LOADING MATRIX, WITH ROWS VARIABLES"
850 ;"AND COLUMNS FACTORS IS--"
860 ;
870 MAT PRINT L
880 ;
890 ;"VARIABLE", "H-SQUARE"
900 FOR I=1 TO V
910 LET H2=0
920 FOR J=1 TO F
930 LET H2=H2+((L(I,J))*L(I,J))
940 NEXT J
950 ;I, H2
960 NEXT I
970 ;
980 ;"FACTOR", "VAR.EXPL.", "PROP.VAR.EXPL.", "CUM.PROP.VAR.EXPL."
990 LET V2=0
1000 FOR J=1 TO F

```

```

1010 LET V1=0
1020 FOR I=1 TO V
1030 LET V1=V1+((L(I,J))*L(I,J))
1040 NEXT I
1050 LET V2=V2+V1/V
1060 J,V1,V1/V,V2
1070 NEXT J
1080 STOP
1090 REM *****REFLECTS VARIABLES AND KEEPS TRACK OF SIGNS*****
1100 FOR I=1 TO V
1110 LET R2(I,Z1)=0-R2(I,Z1)
1120 LET R2(Z1,I)=0-R2(Z1,I)
1130 NEXT I
1140 LET W(Z1)=0-W(Z1)
1150 LET Z=Z+1
1160 RETURN
1170 ;"INPUT NUMBER OF VARIABLES AND NUMBER OF FACTORS DESIRED"
1180 INPUT V,F
1190 DIM C(V), L(V,F), L1(1,V), L2(V,1), W(V), R(V,V), R1(V,V), R2(V,V)
1200 ;"INPUT DESIRED CORRELATION MATRIX"
1210 MAT INPUT R
1220 GOTO 350
1230 END

```

REFERENCES

GUERTIN, W. H., & BAILEY, J. P. <i>Introduction to modern factor analysis</i>. Ann Arbor, Mich: Edwards, 1970. HORST, P. <i>Factor analysis of data matrices</i>. New York:	Holt, Rinehart, & Winston, 1965. NIE, N. H., HULL, C. H., JENKINS, J. G., STEINBRENNER, K., & BENT, D. H. <i>Statistical package for the social sciences</i>. New York: McGraw-Hill, 1975. NUNNALLY, J. C. <i>Psychometric theory</i>. New York: McGraw-Hill, 1967.