UNIVERSITY OF ILLINOIS

DIGITAL COMPUTER

LIBRARY ROUTINE K 16 - 263

TITLE: Modified Multiple Regression Analysis (DOI Only)

TYPE: Entire program

DESCRIPTION: This routine, which is a revision of K 14, will calculate and print any or all of the following:

(1) Zero-order correlations,

(2) Covariances,

(3) Means and standard deviations,

(4) Multiple correlations with associated regression weights, both standardized and unstandardized,

(5) The standard errors of the regression weights and the standard error of estimate.

This analysis can be done on any subset of the variables on the data tape, and in addition, the user has a choice of five transformations that can be made on the variables.

This routine differs from K 14 in that the raw scores are stored on the drum. For repeated use, this routine is preferred as it is necessary to read the data once only.

CAPACITY: s: sample size

p: total number of variables on tape

n: subset of variables used in the analysis

$s(p + 1) \leq 8137$; $p \leq 50$; $n \leq 22$; $s > n + 1$

METHOD OF USE:

	Stops
1. Master tape	24012
2. Parameter tape, Type N	24113
3. Data tape	24012

There are two types of parameter tapes:

(1) Type N directs the computer to read data from tape, store information on the drum, and compute.

(2) Type J directs the computer to read from the drum and compute.

Any number of Type J tapes can follow step 3 above

and each will stop on 24012. The data on the drum can be replaced at any 24012 stop by reading a Type N tape followed by a data tape as in steps 2 and 3 above.

PARAMETER TAPES, TYPE N AND TYPE J:

Type N and Type J parameter tapes are alike in every respect except for the final terminating symbol which for Type N is an N and for Type J is a J.

The first section of each parameter tape must contain six parameters, s, n, f, ℓ, k, and j. These are punched on tape with terminating symbols as follows:

s S n N f F ℓ L k K j J

(The terminating symbols, S and K, can also be punched as - and +).

(1) "s" is the sample size or the number of observations.

(2) "n" is the number of non-eliminated variables; i.e., the subset selected by the parameter tape for analysis.

(3) "f" is the number of decimal places for the zero-order correlations. To suppress this print-out, set f = 0.

(4) "ℓ" is the number of decimal places for the covariances. To suppress this print-out, set ℓ = 0.

(5) "k" is the number of places for the multiple correlations, the standardized regression weights and their standard errors. To suppress this print-out, set k = 0.

(6) "j" is the number of places for the standard errors of estimate, the unstandardized regression weights, and their standard errors. To suppress these, set j = 0.

The second section of each parameter tape consists of a sequence of 1's and 0's punched on tape in the following way:

00101011 space

These digits correspond to the n non-eliminated variables. The 1's specify that these variables are to be considered successively as dependent, with the other (n - 1) variables in the subset as independent,

variables. In the example above, the third, the fifth, the seventh, and the eighth would be treated in turn as dependent variables. The sequence is terminated by any fifth-hole character such as a space, a carriage return, or a delay. It is not necessary to punch 0's after the final "1" in the sequence. As an example, if only the first variable is to be considered as dependent, then punch only a "1" followed by a space.

The third section of each parameter tape is the identification. This can be a name, a date, a problem description, or any combination of teletype characters whatever except single fifth-hole delays. This portion of the tape is reproduced exactly as a heading for the results. It is terminated by a single hole delay. An example of such a heading follows:

4/1/59

PROBLEM NO. 17 Single fifth-hole delay

If no heading is desired, punch two fifth-hole delays.

The fourth section of each parameter tape specifies the transformations to be made on the variables and whether a variable is to be eliminated or not. This section consists of a set of p signed integers where these integers correspond to the numbers on the data tape. If x is the variable, the meaning of the integers is shown in the chart below:

Integer	Meaning
+0	Eliminate x
+1	x is unchanged
+2	x is transformed to x^2
+3	x is transformed to x^3
+4	x is transformed to the square root of x
+5	x is transformed to $\frac{\log x}{10}$

(Note: x must be positive for the +4 and +5

transformations, and x must be greater than 10^{-10} for the +5 transformation.)

The set of signed integers must be terminated by an NN for a Type N parameter tape and by an NJ for a Type J tape. An example may be as follows:

+1+0+0+1+1+1+2+2+0+1NN

DATA TAPE: The p variables are punched on tape as signed fractions with an N terminating symbol after each of the s observations. If an F is punched instead of an N, the computer will stop on 3012K. An additional section of the data tape can be inserted in the reader; by raising the black switch, the problem is continued.

SCALING OF THE DATA: The data are read from tape as signed fractions. Locating the decimal point in the unstandardized numbers can be confusing unless all of the variables are scaled by the same constant.

If the scaling constant for all x_i is 10^{-z}, then the scaling in the results is as follows:

(1) The means, standard deviations, and the standard error of estimate are scaled by 10^{-z},

(2) The covariances are scaled by 10^{-2z}, and

(3) The remainder of the results are unaffected.

MATHEMATICAL FORMULAS: Let the jth observation of the ith variable be x_{ij}.

j. 1, 2, .. , s i, k. 1, 2, ..., n

The mean: $$\bar{x}_i = \sum_{j=1}^{s} \frac{x_{ij}}{s}$$

The Covariance: $$c_{ik} = \sum_{j=1}^{s} \frac{x_{ij}\, x_{kj}}{s} - \bar{x}_i\, \bar{x}_k$$

The Correlation Coefficient: $$r_{ik} = \frac{c_{ik}}{\sqrt{c_{ii}\, c_{kk}}}$$

The Standardized Regression Weight:

$$b^*_{ik} = \frac{-r^{ik}}{r^{ii}}, \quad k = 1, \ldots, n$$

where r^{ik} is the (i,k) element of the inverse of the correlation matrix.

The Standard Error of the Standardized Regression Weights:

$$S_{b^*_{ik}} = \sqrt{\frac{r^{ii}\, r^{kk} - (r^{ik})^2}{(s - n)\,(r^{ii})^2}}$$

The Multiple Correlation Coefficient:

$$R_i = \sqrt{1 - \frac{1}{r^{ii}}}$$

The Standard Error of Estimate:

$$S_i = \sqrt{c_{ii}\,(1 - R_i^2)}$$

The Unstandardized Regression Weights:

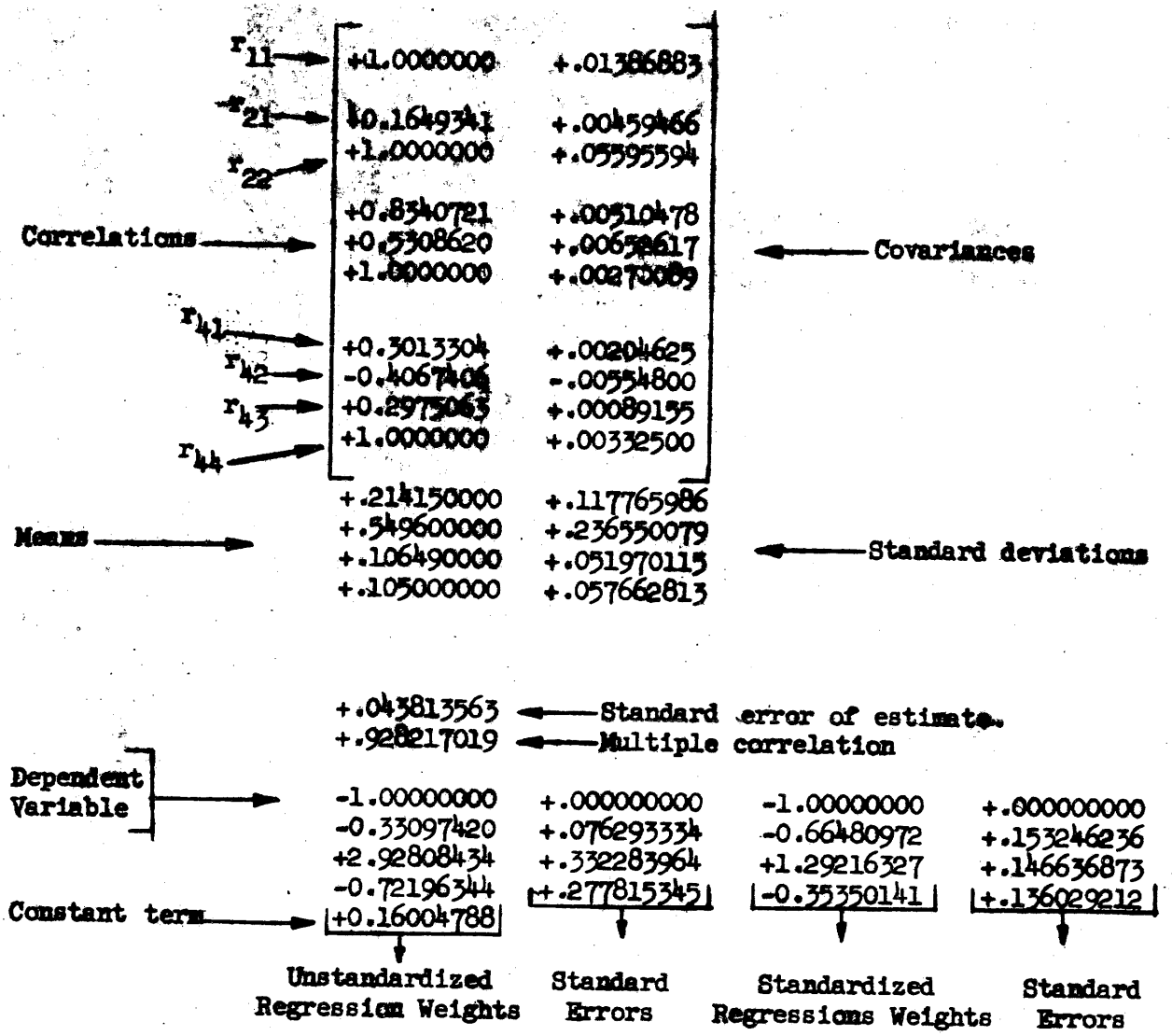
$$b_{ik} = \sqrt{\frac{c_{ii}}{c_{kk}}}\; b^*_{ik}$$

The Standard Error of the Unstandardized Regression Weights:

$$S_{b_{ik}} = \sqrt{\frac{c_{ii}}{c_{kk}}}\; S_{b^*_{ik}}$$

THE RESULTS:

The form of print-out of the results can be illustrated by an example. For a problem with four variables, the print-out could appear as follows. The mean and standard deviations are always printed to nine places.



DURATION:

The chart below may be helpful in making time estimates. These are average times for typical problems. If the number of variables, p, on the data tape, exceeds the number selected for the analysis, n, by a large amount, the reading time should be increased proportionately. Also the number of digits on input and output should be taken into account.

TIME IN SECONDS

Variables	Read Time (5 Digits)		Calculation	Printing Time (5 Digits)		
n	Type N	Type J	Invert R	R	Cov.	Reg. Weights
5	.25 s	.10 s	4	3	3	4
10	.52 s	.22 s	10	10	10	8
15	.88 s	.43 s	18	20	20	12
20	1.30 s	.70 s	35	35	35	17

NOTE 1: If the master tape stops on FF from location 378, the sum check has failed indicating the tape has been read incorrectly.

NOTE 2: If the number of variables, p, in any row of the data tape does not agree with the number of integers on the parameter tape, the computer will stop on FF from location 12J.

NOTE 3: If a drum error has been made, the computer will stop on FF from location 06N.

RT: 1/13/60

DATE April 6, 1959
ROUTINE K 14 BY G.H. Golub
REVISION BY V. West
WRITE-UP BY K.W. Dickman
APPROVED BY J. Snyder

lgr

COMPUTER ORDERS AND NOTES: This routine is a revision of K 14. Changes in K 14 are shown below.

LOCATION Abs.	LOCATION Rel.	ORDER		NOTES
		Subroutine B		Computes means and cross-products divided by s.
205	5	50L	267L	Remove stop
215	8	26371F	00F	Transfer to read final term. symbol
		Routine H		Read in data and perform transformations
275	0	5050SJ	50L	Same
276	1	26S4	26297F	Transfer to new tape stop
290	15	41F	22219F	Bypass old tape stop
291	16	2611L	0016L	Same
297	22	4013F	L012S5	
298	23	30298F	26366F	Stop on F; transfer to store on drum
299	24	L521S4	L06F	
300	25	40F	L3F	Stop on FF if p variables on data tape do not
301	26	36302F	FF18F	agree with integers on parameter tape
302	27	26372F	00F	
303	28	LJ12F	3637L	Same
		00366K		
366	0	L5328F	0020F	
367	1	464L	4615L	
368	2	J050SJ	502L	Store data on drum with Y 1
369	3	26S9	001080S7	
370	4	00F	26299F	Transfer to test number of variables read
371	5	4125L	2618L	
372	6	L5328F	L45S5	
373	7	F43L	403L	Increase drum address by p + 1
374	8	F525L	4025L	Count rows
375	9	L04F	36275F	Test: S rows?

LOCATION		ORDER		NOTES
Abs.	Rel.			
376	10	2210L	L512L	
377	11	40276F	22276F	
378	12	2613L	00F	
379	13	5050SJ	5013L	Read from drum using Y 1
380	14	2689	00108087	
381	15	00F	F5328F	
382	16	L4585	L414L	Increase drum address
383	17	4014L	26277F	Transform row
384	18	L523L	40215F	
385	19	814F	L024L	Read terminating symbol
386	20	3621L	248F	Stop: end of parameter tape
387	21	L5328F	0020F	
388	22	4615L	2210L	Set number of variables to be read
389	23	268F	001F	
390	24	00F	0013F	
		. . . . .		
		Sum Check Change		
		L50F6	89289	
		2418N		