DECUS
PROGRAM LIBRARY

DECUS NO.	8-495
TITLE	CORRELATION ANALYSIS
AUTHOR	Ralf Beyer
COMPANY	DFVLR, Institut fuer Flugfuehrung Flughafen, Germany
DATE	January 8, 1971
SOURCE LANGUAGE	PAL-D

CORRELATION ANALYSIS

DECUS Program Library Write-up

DECUS NO. 8-495

ABSTRACT

This program is a patch to be used in connection with the "FORT" routine of DECUS NO. 8-137 "Programs for Storage, Manipulation and Calculation of Data Using DECtape." In addition to the standard analysis of variance procedure the significance of first and higher order correlations of the data samples with respect to the sequence of their input and the linear correlation coefficient are calculated.

REQUIREMENTS

Storage

The routine fits within the "FORT" routine of DECUS NO. 8-137 and occupies memory locations 31-37, 176-177 and 24Ø1-3771.

USAGE

Loading

The routine "FORT" together with the routine "SUBS" (see DECUS NO. 8-137) are loaded first followed by the patch described.

Calling Sequence

The program cannot be called as a subroutine.

Switch Settings

The switch register is used to enter the starting address (2Ø4) only.

Start up and/or Entry

Assuming that "FORT," "SUBS" and the patch have been loaded properly in that order set the switch register to 2Ø4, press LOAD ADDRESS and START. The program responds with the initial dialog of "FORT". Proceed as described in DECUS NO. 8-137. Make sure that a data tape with data written according to the format described under "DATRIT" of DECUS NO. 8-137 has been mounted on DECtape unit no. 1 with the remote control switch on and "WRITE" locked.

RESTRICTIONS

Status Core

By insertion of the patch into the FORT routine of DECUS NO. 8-137 the maximum number of lines and blocks allowed on input had to be reduced to 100_8 each. However, this reduction

normally does not present a problem. If these limits are exceeded parts of the program are overwritten by data and the program shall be destroyed.

DESCRIPTION

Discussion

The program performs the analysis of variance as described in DECUS NO. 8-137 first and in addition calculates the significance of linear and higher order correlations of the data samples with respect to the sequence of their input after output of the analysis of variance table. The linear correlation coefficient is calculated too.

During the analysis of variance "FORT" is linked to the begin of the patch (START) to check that:

a) All data samples have the same size.

b) Not more than 8 data samples are entered.

Otherwise the program proceeds as normal but after the output of the AOV-table the diagnostic of the error is typed which prevented the calculation of the correlations and the typeout of the correlation analysis table. Furthermore the sample means are stored by the patch in a table (MEANTB).

After the last sample has been entered the analysis of variance table is typed. If at least 3 data samples have been entered the significance of linear and higher order correlations of the samples with respect to the sequence of their input is calculated. Otherwise the program types the error diagnostic "LESS THAN 3 SAMPLES" and restarts.

The calculation of the correlations is based on the following considerations:

Assume that X_i is the independent variable with constant intervals (e.g. 1, 2, 3, 4 ...). Between the sample means

$$Y_i = \frac{\sum Y}{n} \qquad Y = f(X_i)$$

"contrasts" can be defined which are special forms of so-called linear functions. A contrast C is a linear function with the weighing factors c_i:

$$C = c_1 Y_1 + c_2 Y_2 + \ldots c_k Y_k$$

where

$$\sum_{o}^{k} c_i = 0$$

By means of these contrasts the total sum of squares between samples (from the analysis of variance) can be represented by the sum of its components which have the following form:

$$SBS_o = \frac{(c_1 Y_1 + c_2 Y_2 + \ldots c_k Y_k)^2}{n (c_1^2 + c_2^2 + \ldots c_k^2)}$$

$$SBS_o = \frac{C^2}{m \sum c_i^2}$$

The index o designates the order of the correlation for which a component SBS_o is calculated. The corresponding ci for each order of calculation may be found in a table in the appendix to this program. All components SBS_o add up to the total sum of squares between samples if they are independent, i.e.,

$$c_{I_1} c_{II_1} + c_{I2} c_{II_2} + \ldots c_{I_k} c_{II_k} = 0$$

The total sum of squares between samples then is

$$SBS_{tot} = SBS_{lin} + SBS_{square} + SBS_{cubic} + \ldots SBS_{nth\ order}$$

The statistical significance of each component of SBS_o is tested by means of the F-test:

$$F = \frac{SBS_o}{s^2_{within}}$$

s^2_{within} - Variance within samples (from the analysis of variance table)

The required F is found in a table under 1 and (n-1)k degrees of freedom

n - sample size
k - number of samples

The linear correlation coefficient is calculated according to

$$r_{lin} = \sqrt{\frac{SBS_{lin}}{SBS_{tot}}}$$

FORMAT

Input Data

Floating-point numbers according to "DATRIT" of DECUS NO. 8-137.

REFERENCES

Textbooks

Mittenecker, Erich: "Planning und statistische Auswertung von Experimenten" (Published by Franz Deuticke Book Company, Vienna, Austria, 1970)

Table of ci's

A table of ci's may be found in the appendix for k samples of size n.

APPENDIX

Table of ci's for k samples of size n up to the fourth order of correlation

k	order	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8
3	linear	-1	0	1					
	square	1	-2	1					
4	linear	-3	-1	1	3				
	square	1	-1	-1	1				
	cubic	-1	3	-3	1				
5	linear	-2	-1	0	1	2			
	square	2	-1	-2	-1	2			
	cubic	-1	2	0	-2	1			
	4th	1	-4	6	-4	1			
6	linear	-5	-3	-1	1	3	5		
	square	5	-1	-4	-4	-1	5		
	cubic	-5	7	4	-4	-7	5		
	4th	1	-3	2	2	-3	1		
7	linear	-3	-2	-1	0	1	2	3	
	square	5	0	-3	-4	-3	0	5	
	cubic	-1	1	1	0	-1	-1	1	
	4th	3	-7	1	6	1	-7	3	
8	linear	-7	-5	-3	-1	1	3	5	7
	square	7	1	-3	-5	-5	-3	1	7
	cubic	-7	5	7	3	-3	-7	-5	7
	4th	7	-13	-3	9	9	-3	-13	7