AUSTRALIAN COUNCIL FOR EDUCATIONAL RESEARCH

Number 5
October 1973

ACER DATA PROCESSING BULLETIN

PREPARATION OF PUNCHED PROGRAM CARDS

ACER's data processing is mostly being carried out on three machines: IBM 7044 (Melbourne University), CDC CYBER (Melbourne University) and IBM 360/50 (ICIANZ). The preparation of program cards for these machines must follow certain punching conventions:-

(a) IBM 7044 (Melbourne University)

Cards for this machine must be punched in BCD format (i.e. using a 26 card punch or a 29 converted for 'scientific' usage as in the Melbourne University punch room).

Cards punched on the ACER 29 punch which punches in EBCDIC format must be punched with the following conversions

for character (left parenthesis, punch % percentage

character)right parenthesis, punch < less than

character = equal sign, punch # number symbol

character + plus, punch & ampersand

character ' inverted comma, punch @ at symbol

A card with these conversions is attached to the ACER punch.

(b) CDC CYBER 72 (Melbourne University)

Cards for this machine are routinely punched as for the 7044. However EBCDIC (29 punched) input may be accepted as follows:

(a) If the whole job (program & data) is punched on a 29 punch (as at ACER) punch 29 in columns 79-80 of the job card.

(b) If part of the job (e.g. just the data) is punched on a 29 punch (as at ACER) punch 29 in columns 79-80 of the preceding EOR (End-Of-Record) card.

(c) IBM 360/50 (ICIANZ)

Cards for this machine are routinely punched on a standard 29 (as at ACER). Cards punched as for the Melbourne University machine will be accepted with a modification to the EXEC card viz.

```
//  EXEC  FORTGCLG,PARM.FORT=BCD
```

OSIRIS PROGRAMS

(a) SUBMAT

Creates a subset of an OSIRIS matrix file. This program serves as a middle step between programs such as MDC, which produces a matrix of up to 200 variables, and programs such as PARCOR, REGRESSN and FACTAN, which can not accept matrices of that size.

Submat uses less than 100K, and hence runs as a class A job. For further information, see page C425 of the OSIRIS manual.

(b) SUMMAR
This program computes summary descriptive statistics for groups of cases. The user defines the groups, specifies the statistical operations, and supplies a list of input variables on which they are to be performed. An output file of the summarisation variables thus calculated can be produced.
Statistics which can be calculated include sums, means, standard deviations, variances, counts of cases, minima and maxima.
Summar runs as a class A job. For more information consult page C540 of the OSIRIS manual.

PROGRAM PRODUCTION/MODIFICATION

(a) WRDFRQ
This is a program for counting word frequencies in passages and sorting the words into alphabetic order.
Words can be grouped into several types and sorted independently within types.

Further work is in progress to enable the program to:

(a) print results after reading pre-determined numbers of words (eg 25, 50, 100).

(b) analyse mis-spellings.

WRDFRQ runs on the CYBER. It requires 60000 (octal) words of storage.

(b) ITANAL
Lindsay McKay's test scoring and item analysis program has been adapted to run on the CYBER, under the name of ITANAL. The program accepts alphabetic or numeric response data. At present it can handle up to 132 items, 13 scales, and 6 alternatives per item (not counting "no answer", which is treated as an additional alternative). These limits could be extended if necessary.

Means, standard deviations and KR20 reliabilities are calculated for each scale, as well as the scale intercorrelations.

For each item, the program gives the number and percentage choosing each alternative, and the point-biserial correlation between every scale and the choice of every alternative (not merely the correct alternative). When an item is correlated with a scale to which it belongs, the correlations are calculated by excluding the item from the scale total.

Core required is 60000 (octal).

(c) TESTAT
The Veldman TESTAT program has been modified to handle alphanumeric data. It now has 4 data input options.

(a) numeric data

(b) IBM 1230 special marking code data cards

(c) "normal" alphanumeric data (in which the set of response alternatives may consist of any set of alphabetic, numeric, or special characters, but with each question having the same set).

(d) "special" alphabetic data (in which some items have A,B,C,D,E, as possible alternatives, others have E,F,G,H,J, and so on).

The modified program runs on the IBM 360. It has been put in the OSIRIS program library (ISR.LIB) on disk in compiled form. To run it from disk, the following set-up should be used:

```
Job card
//JOBLIB  DD  DSN=ISR.LIB,DISP=OLD
//        EXEC  PGM=TESTAT
//FT06F001  DD  SYSOUT=A,DCB=(RECFM=FBA,LRECL=133,BLKSIZE=798)
//FT05F001  DD  *
data cards
/*
```

OTHER PROGRAMS

(a) DISCRIMINANT ANALYSIS

Two discriminant analysis programs are now available. One is from Veldman, the other is from Cooley and Lohnes.

Given some data in which several groups of subjects are measured on certain variables, discriminant analysis is a method for calculating a set of linear functions of the original variables which maximally discriminate between the groups, subject to the restriction that each function is orthogonal to all preceding functions.

The Veldman program requires raw data input, i.e. the scores of each subject on the original variables. The Cooley and Lohnes program requires data in a rather unusual form - namely the total sums of squares and cross products, the within groups s.s. and c.p. and the group means and grand means. Data in this form is punched out by Cooley and Lohnes' MANOVA and COVAR programs. These are not yet available, but it is planned to prepare a MANOVA deck shortly.

The Cooley and Lohnes program runs on the IBM 360, as a class A job.

The Veldman program runs on the Melbourne University IBM 7044.

(b) COMPONENT SCORES

This program is a modification of the BMD factor analysis program (03M). It performs factor analyses on one or more subsets of the variables in a given set of data, calculates the scores of subjects on the first principal component from each analysis, and punches out the component scores.

A compiled version has been stored in the ISR library (ISR.LIB) on disk under the name M03. A more detailed write-up and a deck for running the program from disk are available in the filing cabinets. Jobs must be run as class E, since the program requires more than 120K core.

(c) FUNSTAT LISTINGS

A set of program listings to accompany Roscoe's Fundamental Research Statistics for the Behavioral Sciences has been received by Richard Bell. Programs include simple analysis of variance (with options for Scheffe test for multiple comparisons, Helmert's orthogonal contrasts, Tukey's pairwise comparisons), analysis of variance with a nested variable, analysis of covariance.

FACTOR ANALYSIS PACKAGE : AGFAP

A program package for factor analysis is being obtained from the University of Alberta. This package has been mainly run from a terminal system and therefore some modification will be required to adapt the package to the IBM 360/50.

The AGFAP package features many options:

1. For raw data input it computes either product-moment or tetrachoric correlation coefficients.

2. There is a choice of principal components, image or common factor methods. For the "common" option there are three procedures available: unweighted least squares, generalised least squares, or maximum likelihood.

3. Many rotation options are available: blind orthogonal (orthomax, quartimax, equimax, varimax); blind oblique (Harris-Kaiser, promax, procrustes); Schonemann orthogonal; Mosier oblique.

4. The package also computes factor scores.