DOCUMENT RESUME

ED 064 959 24 EM 010 209

TITLE	Documentation Standards Handbook for EIN Software Catalog, October 1970.
INSTITUTION	Interuniversity Communications Council (EDUCOM), Boston, Mass.
SPONS AGENCY	National Science Foundation, Washington, D.C.; Office of Education (DHEW), Washington, D.C.
BUREAU NO	BR-8-0697/Attachment No. 2
PUB DATE	Oct 70
GRANT	OEG-0-8-080697-4468(007)
NOTE	17p.; See also EM 010 206, EM 010 207, EM 010 208, EM 010 210, EM 010 211, EM 010 212
EDRS PRICE	MF-$0.65 HC-$3.29
DESCRIPTORS	Abstracts; *Computer Programs; *Documentation; *Information Networks; *Interinstitutional Cooperation; *Program Descriptions
IDENTIFIERS	*Educational Information Network; EIN

ABSTRACT

The Educational Information Network (EIN) enables member colleges, universities, and educational service organizations to share their computational resources among interested constituents. To accomplish this objective, the EIN Software Catalog is published, describing the available resources. The purpose of this handbook is to define the documentation for programs and systems of programs available through EIN. By following the guidelines in the handbook, participating institutions facilitate the listing of their programs in the Catalog. (JY)

EDUCOM

EIN PROJECT

DOCUMENTATION STANDARDS HANDBOOK FOR EIN SOFTWARE CATALOG

October 1970

Prepared under a
Joint Grant from the
U.S. Office of Education
and the
National Science Foundation
Grant No. OEG-0-8-080697-4468 (007)

100 CHARLES RIVER PLAZA

BOSTON, MASS. 02114 617-227-1805

Edition of June 1969 prepared by:

Thomas A. Keenan, Executive Director
Richard D. Ferguson, Research Associate/Editor

Edition of October 1970 augumented by:

John C. LeGates, Executive Director
Wayne A. Zafft, Editor

The work presented or reported herein was performed pursuant to a Grant from the U.S. Office of Education, Department of Health, Education, and Welfare. However, the opinions expressed herein do not necessarily reflect the position or policy of the U.S. Office of Education, and no official endorsement by the U. S. Office of Education should be inferred.

TABLE OF CONTENTS

DOCUMENTATION STANDARDS HANDBOOK FOR THE EIN SOFTWARE CATALOG

Introduction

The Educational Information Network (EIN) is an EDUCOM project, jointly sponsored by the National Science Foundation (NSF) and the U.S. Office of Education. The major objective of the project is to enable member colleges, universities, and educational service organizations to share their computational resources among interested constituents. To accomplish this objective, the *EIN Software Catalog* is published, describing the available resources.

The catalog listings provide the information necessary to permit a user to identify and locate those computational resources of potential utility. After selecting the resource that best answers a particular need, the user will then prepare data according to the required format and send the input to the selected resource center for processing. A record-keeping system will assure the resource center of reciprocal credit or payment for its services and will, in addition, monitor the progress of resource utilization throughout the network.

To expedite the technical aspects of this process, it is necessary for EDUCOM to receive properly prepared documentation from participating institutions through their designated technical representatives. The purpose of this handbook is to define the documentation for programs and systems of programs available through EIN.

The EIN staff recognizes that present documentation at participating institutions does not exist in any common format, and that, in many instances, the documentation submitted will not easily meet the format specified in this handbook. However, it is imperative that all information herein requested be included, regardless of the format chosen.

Documentation Criteria

Documentation that meets the EDUCOM standards must include the amount of information necessary to inform a prospective user of the precise problem that the program was designed to solve, and to enable him to prepare the required input for the desired results. For extended systems of programs, the catalog entry should identify the capabilities of the system so that the reader can judge the value of studying the references in more detail.

The central item in the catalog entry is a 300-word program abstract that summarizes the functional features of a program and clearly states those processing options which make it useful. While abstracting is known to be somewhat of an art, we believe that careful attention to detail in the content of the abstract is more important that strict attention to form.

Each catalog entry will be carefully checked for the inclusion of the information items specified in this handbook. Those deficient in terms of content will be returned to the submitter so that he can supply the missing items. In addition, all entries will be returned to the resource institution for a final examination before printing.

Constructive suggestions for improvement of the format and/or the definitions are welcome. Any suggested changes will be placed before the EIN Executive Committee.

Model Catalog Entry

A model catalog entry for a program submitted by the University of Iowa has been reproduced in the following pages. The entry is representative of the type that we hope to receive from submittors.

EIN DOCUMENTATION STANDARDS FOR PROGRAMS

Field	*Definition*
DESCRIPTIVE TITLE	This title is the name that describes what the program does.
CALLING NAME	This item provides the calling name or loading name of the program, including any symbols needed for identification in the computer center's library.
INSTALLATION NAME	Enter the name of the university computer center at which the program is presently operating and its location (if different from the university).
AUTHOR(S) AND AFFILIATION(S)	The author of the program is usually the person who did the actual programming and design work. If these tasks were separate, please list both parties and their present affiliations (spefific office or particular laboratory). An author may request that his name not appear in the published catalog.
LANGUAGE	The programming language in which the program is written.
COMPUTER	The computer on which the program executes.
PROGRAM AVAILABILITY	Specify whether program decks and/or listings are available on user request.
CONTACT	Person(s) to contact at the institution for further information.

FUNCTIONAL ABSTRACT

This information is central to the catalog entry. It should contain the following.

a. Description of the Problem
The problem that the program is designed to solve should be described in such a way as to help the reader to identify elements that may be analogous to his own problem. While this description must be brief, it should clearly identify the conceptual limitations of the program.

Field & Definition (for Programs; continued)

b. Method of Solution
When the method of solution is well known or documented in standard publications, it should be identified by references. Modifications to well-known methods, new methods, or novel combinations of methods should be fully described to indicate their applicability.

c. Special Features of the Program
Processing features and options that contribute to the uniqueness of the program should be summarized. Types of input and output should be discussed in terms of their potential value for solution of problems.

d. References
Books, periodicals, and other supporting literature related to the program and its methodology should be listed.

USER INSTRUCTIONS

This item will be the basis for data preparation. Hence, instructions regarding data requirements should be clearly and explicitly stated as follows.

a. Input Preparation Formats and Options
These instructions provide the user with the information necessary to prepare his data for input to the program in terms of:
(1) the precise definition of all variables;
(2) the exact format and arrangement of input parameters;
(3) the required card or tape format for all input data;
(4) the sequence of control statements, if required.

b. Output Formats and Options
These instructions will clearly explain all output variables and the choices of formats available for selection. Some note regarding the accuracy of results also should be included.

c. Data Restrictions
The user must be provided with a full explanation of any data restrictions such as those constituting illegal input, numerical or data-set limitations, and the number or size of the data sets that can be handled by the program.

d. Procedural References
Manuals, detailed documentation, etc. required to use the program are listed.

SAMPLE INPUT/OUTPUT MODELS

This item will help the user to understand all of the aforementioned information quickly and easily. The sample test case

Field & Definition (for Programs; continued)

demonstrates the main features of the program through graphic illustrations (printouts) of real results.

COST ESTIMATE

A statement of the number of computational units used in processing the sample problem should be provided to give the user an idea of the potential time and costs for using the program. Actual or approximate dollar value should be stated where possible.

EIN DOCUMENTATION STANDARDS FOR SYSTEMS SOFTWARE

The description of systems follows the guidelines for Programs. However, the documentation should reference operational details rather than describe them in full. The fields should include the following information.

Field	*Definition*
DESCRIPTIVE TITLE	As for Programs, including the specification of the type of software described. For example, Remote Interactive System or Operating System.
CALLING NAME	As for Programs.
INSTALLATION NAME	As for Programs.
AUTHOR(S) AND AFFILIATION(S)	As for Programs; this may include the source of the system.
LANGUAGE	As for Programs, if applicable.
COMPUTER	As for Programs.
PROGRAM AVAILABILITY	As for Programs. Specify whether or not the availability is without charge.
CONTACT	As for Programs. This person will be used more heavily than for user programs.

Field & Definition (for Systems; continued)

FUNCTIONAL ABSTRACT

The functional abstract should include a brief summary of the software: what it does, how it does it, and any other information a person may require to determine whether it can be helpful to his needs. The purpose of the abstract is to inform of the uses to which the system is suited.

USER INSTRUCTIONS

This item will be the basis for a user's decision as to whether the constraints of the system are acceptable to his purpose. It should include the data restrictions, as well as the hardware, software, and interface constraints, if applicable. Other special requirements should be mentioned.

References should be cited, including manuals, articles, and all other material required for user understanding of the system. Location and prices of this material should be given. EIN will maintain a copy of each item in the central office as a backup reference.

SAMPLE INPUT/OUTPUT MODELS

Exclude except for writeups of specific programs, where desired.

COST ESTIMATE

The pricing algorithm for use of the system should be given. Examples, where applicable, are to be included.

EIN DOCUMENTATION STANDARDS FOR COMPUTER FACILITIES

The description of facilities should follow the guidelines for Programs. However, rather than including full operational details, the documentation should reference them. The fields should include the following information.

Field	*Definition*
DESCRIPTIVE TITLE	As for Programs.
INSTALLATION NAME	As for Programs.
LANGUAGE	As for Programs.

Field & Definition (for Facilities; continued)

COMPUTER — As for Programs.

AVAILABILITY — Include access information.

CONTACT — As for Systems Software

FUNCTIONAL ABSTRACT

The purpose of the functional abstract is to describe the potential utility of the facility in broad terms. The abstract should include a description of the hardware and special software, the executive system and special peripheral equipment available. Access information should also be given, including type of access and any special constraints (type of terminals needed for remote entry, length or width of tapes, etc.).

USER INSTRUCTIONS

As for Systems Software, include a preliminary bibliography of system references and any special local references. The times the system is available should be listed. The services provided by the installation should be described (consulting, programming, debugging, keypunching, etc.), as well as any special procedures or options that should be noted.

SAMPLE INPUT/OUTPUT MODELS

Exclude except for writeups of specific programs, if desired.

COST ESTIMATE

This should include the rates for consulting, postage and handling, and other services, as well as actual computer charges. Also, indicate any differential between educational and noneducational rates.

000 0107

DESCRIPTIVE TITLE	General Multiple Regression Analysis
CALLING NAME	REGAN1
INSTALLATION NAME	The University of Iowa University Computer Center
AUTHOR(S) AND AFFILIATION(S)	Stan Walljasper, IBM 7044 version Converted to IBM 360 by Louise R. Levine The University of Iowa Computer Center
LANGUAGE	FORTRAN IV (G)
COMPUTER	IBM 360/65
PROGRAM AVAILABILITY	Decks and listings presently available
CONTACT	Mrs. Louise R. Levine, Program Librarian, Applications Programming, University Computer Center, The Univ. of Iowa, Iowa City, Iowa 52240 Tel.: (319) 353-5580

FUNCTIONAL ABSTRACT

REGAN1 computes the Pearson Product Moment correlation matrix, mean, standard deviation, and population standard deviation. If desired, the multiple R, standard error of the estimates, F ratio, degrees of freedom, regression coefficients, partial coefficients and F ratio associated with each individual variate, the intercept constant, and residuals also may be calculated. There is *no* allowance for missing data. The basic equations and a simple explanation may be found in Cooley and Lohnes.[1]

REFERENCES

1. Cooley, W.W., and Lohnes, P.R., *Multivariate Procedures for the Behavioral Sciences*, (John Wiley & Sons, Inc., New York, 1962), pp. 31–45.

000 0107

000 0107

USER INSTRUCTIONS

Input Deck

The input deck consists of the following cards in the order described below.

Title Card

This card gives any alphanumeric information to be printed as a heading on the output.

Problem Card

Columns	*Contents*
4– 5	Number of variables (independent + dependent $\leq$ 50)
6–10	Total number of observations
11–15	1: No data error check is desired 2: Data error check is desired
16–20	1: No residuals are to be calculated 2: Residuals are to be calculated
21–30	If data error check is requested, punch the lower bound for allowable data. The assumed format is F10.5. Leave blank if no data error check is requested.
31–40	If data error check is requested, punch the upper bound for allowable data. The assumed format is F10.5.
41	Number of Format Cards ($\leq$ 3)
42	blank: card input 2,3,4,8, or 9: unit number of disc or tape input Appropriate DD cards for this file should be placed before the //GO.SYSIN card.

Format Cards

Columns	*Contents*
1–80	The data should be read by subject for all variables with an E or F type FORTRAN format. A maximum of 3 Format Cards is allowed.

Data Cards

The Data Cards contain the records of observations. Any uniform

continued

000 0107

number of cards per observations may be used. Each observation must start on a new card.

/* Card

//GO.FT05F00n DD * Card

This card must be present for each problem, with n taking the value of 2 for the first problem and increasing by one for each additional problem.

Permutation Cards

Columns	*Contents*
5	1: No residuals are desired 2: Residuals are desired
6–10	Number of variables to be used for the multiple R (independent + dependent variables)
11–12	Dependent variable number
13–14	Independent variables by number. If there are more than 34 independent variables, continue on another card beginning in Cols. 1–2.
⋮	⋮
79–80	

Only one problem can be run with as many Permutation Cards (max. 20) as necessary.

Program Deck

```
//name JOB (project no., etc.)
//step EXEC STAT,PARM.GO='REGAN1'
//GO.SYSIN DD *
     Title Card
     Problem Card
     Format Card
     Data Cards
/*
//GO.FT05F002  DD *
     Permutation Cards
/*
```

Method

The data are read and the summation on matrices are updated using an auxiliary index matrix.[1] Means, deviations, and product-moment

continued

000 0107

000 0107

correlations are computed by a triangular method.[2] The simple correlation matrix is transferred to a work area from which the multiple R, Beta weights, etc., are calculated. The basic equations and a simple explanation may be found in Cooley and Lohnes.[3]

REGAN1 runs on a 360/65 under HASP and OS with FORTRAN IV (G) and uses 70K bytes of core during execution with the source deck. A work area for data set reference number 1 is required.

REFERENCES

1. Caffrey, J. "CRAM, Algorithm 67", *Collected Algorithms from CACM*, (A.C.M., New York).

2. Hafley, W.L., and Lewis, J.S., "Triangular Regression, Algorithm 142", *Collected Algorithms from CACM*, (New York, A.C.M.).

3. Cooley, W.W., and Lohnes, P.R., *Multivariate Procedures for the Behavioral Sciences*, (John Wiley & Sons, Inc., New York, 1962).

000 0107

SAMPLE INPUT

```
TESTING REGAN)
    4   21    1    2                    1
(4F10.0)
      68.0      2.53      30.6      3.8
      77.8      2.61      34.0      3.8
      83.8      2.82      46.0      3.8
      83.0      2.66      35.7      5.9
      83.8      2.62      54.1      5.3
      90.5      2.86      59.3      3.3
      92.5      2.96      51.9      3.0
      93.2      3.20      52.6      2.9
      93.6      2.90      51.7      5.5
      93.3      3.06      67.4      4.4
      94.7      3.36      70.0      4.1
      98.0      3.89      67.8      4.3
     100.7      3.79      60.9      6.8
     101.5      4.38      75.3      5.5
     103.1      4.41      74.8      5.8
     104.2      4.35      71.7      6.7
     105.4      4.33      83.0      5.5
     106.7      4.26      87.1      5.7
     108.1      4.40      94.0      5.2
     109.9      4.49     107.4      4.5
     113.1      5.13     118.0      3.8
/*
//GO.FT05F002 DD *
   2    4 4 1 2.3
/*
//
```

SAMPLE OUTPUT

000 0107

```
TESTING REGAN1

ECHO CHECK
 NO.OF VARIABLES =        4
 NO.OF OBSERVATIONS =    21
 MS1                      1
 MS2                      2
 ER1                     0.0
 ER2                     0.0
 (4F10.0)

    68.000     2.530    30.600     3.800

ACTUAL NOB =   21

   VAR     MEAN          STD. DEV.        POP. S.D.
    1    0.954713E 02   0.115109E 02   0.112334E 02
    2    0.357190E 01   0.812613E 00   0.793029E 00
    3    0.663475E 02   0.229587E 02   0.224054E 02
    4    0.472857E 01   0.113981E 01   0.111234E 01

CORR. MATRIX   R(I, 1)       R(I, 2)       R(I, 3)       R(I, 4)       R(I, 5)       R(I, 6)       R(I, 7)

R( 2,J)        0.919635

R( 3,J)        0.915466      0.901840

R( 4,J)        0.342833      0.355816      0.164767

 DEPENDENT VARIABLE IS NUMBER  4,   INDEPENDENT VARIABLES ARE  1, 2, 3,

     MLT.R.           STD.ERROR            F                DF
0.56936973E 00    0.10163383E 01    0.27182302E 01    0.17000000E 02

        B              BETA              F
 0.75735092E-01    0.76484263E 00    0.16857405E 01
 0.10165939E 01    0.72476649E 00    0.17452154E 01
-0.59031434E-01   -0.11890440E 01    0.49302368E 01

INTERCEPT      -0.22165442E 01


  TABLE OF RESIDUALS

     Y              Y          RESIDUAL      RESIDUAL
  ESTIMATE         VALUE                      STE
   3.69906        3.80000      0.10094       0.09931        1
   4.32189        3.80000     -0.52189      -0.51350        2
   4.28140        3.80000     -0.48140      -0.47367        3
   4.66619        5.90000      1.23381       1.21398        4
   3.59993        5.30000      1.70007       1.67274        5
   4.04438        3.30000     -0.74438      -0.73241        6
   4.73434        3.00000     -1.73434      -1.70646        7
   4.99001        2.90000     -2.09001      -2.05641        8
   4.76846        5.50000      0.73154       0.71978        9
   3.98160        4.40000      0.41840       0.41168       10
   4.23912        4.10000     -0.13912      -0.13689       11
   5.15771        4.30000     -0.85771      -0.84393       12
   5.66786        6.80000      1.13214       1.11394       13
   5.47818        5.50000      0.02182       0.02147       14
   5.65937        5.50000     -0.15937      -0.15681       15
   5.86468        6.70000      0.83532       0.82189       16
   5.26818        5.50000      0.23182       0.22810       17
   5.05344        5.70000      0.64656       0.63616       18
   4.89448        5.20000      0.30552       0.30061       19
   4.33127        4.50000      0.16873       0.16602       20
   4.59851        3.80000     -0.79851      -0.78567       21

0.17558823E 02
```

000 0107

000 0107

COST ESTIMATE

The costing algorithm is based on the amount of computing time, operator time, core storage, cards in and out, and lines and pages printed. The total cost for the Sample Output above (before a slight change in the pricing schedule) was $0.56.

Charge to user = computer costs + postage + network overhead
= $0.56 + postage + network overhead

CONTENTS—REGAN]

000 0107