



# DECUS

## PROGRAM LIBRARY

DECUS NO.	FOCAL8-171
TITLE	MINNESOTA SOCIOLOGY STATISTICS PROGRAMS
AUTHOR	Philip M. Voxland
COMPANY	Department of Sociology University of Minnesota Minneapolis, Minnesota
DATE	April 20, 1971
SOURCE LANGUAGE	FOCAL-69

### ATTENTION

This is a USER program. Other than requiring that it conform to submittal and review standards, no quality control has been imposed upon this program by DECUS.

The DECUS Program Library is a clearing house only; it does not generate or test programs. No warranty, express or implied, is made by the contributor, Digital Equipment Computer Users Society or Digital Equipment Corporation as to the accuracy or functioning of the program or related material, and no responsibility is assumed by these parties in connection therewith.



# MINNESOTA SOCIOLOGY STATISTICS PROGRAMS

DECUS Program Library Write-up

DECUS NO. FOCAL8-171

## ABSTRACT

The program package consists of a series of small statistical analysis programs of interest to behavioral science researchers. Various parametric and non-parametric statistics are calculated for nominal, ordinal, interval, and ratio level measurements, for discrete and continuous data and for raw data, grouped data and tabular data.

## LANGUAGE

## FOCAL

## CORE MEMORY

4K- All of the programs operate within 4K with the extended functions loaded. Certain programs will be able to accommodate more data if either the extended functions are removed or if 8K of memory are available.

## MINIMUM MACHINE REQUIREMENTS

Digital Equipment Corporation PDP-8 family computers

## GENERAL LOADING INSTRUCTIONS

All of the programs are loaded by simply placing the program tape into the paper tape reader and reading in the program tape. Although all of the programs can be stored on a LIBRA disk system or used with the QUAD multi-user system, none of them require the presence of either system.

## COMMENTS ON RUNNING THE PROGRAMS

To save valuable core space, instructions to the users by the programs are often terse. A number of the table analysis programs begin with asking the number of rows and number of columns in the table being analyzed. Other programs require typing in a -9999 to indicate that all the data has been typed into the computer and that analysis should begin.

PROGRAM NUMBER	PROGRAM NAME	COMMENTS
FOCAL8-171.1	BINO	(Binomial Distribution) The exact and cumulative probability for x number of successes in a sample of size n for a given probability of success is calculated according to the binomial distribution.
FOCAL8-171.2	CHIS	(Table percentages and Chi-Square) In addition to calculating a Chi-Square value and degrees of freedom for an r by c table, the program provides row, column, and base n percents and row and column totals and percents.
FOCAL8-171.3	FISH	(Fisher's Exact Probabilities) When use of the chi-square test is inappropriate for a 2 by 2 table, Fisher's exact test can be used for small sample tests of significance. Exact and cumulative probabilities of the occurrence of the observed table are provided.
FOCAL8-171.4	GATD	(Gamma, Taus, Sommer's D's) Various ordinal bivariate measures of association and Pearson's r are calculated for an r by c table.
FOCAL8-171.5	KOLS	(Kolmogorov-Smirnov tests) The Kolmogorov-Smirnov tests are non-parametric tests for measuring the significance of difference between two ordinal level distributions. In the one-sample case, the observed distribution is compared with a uniform frequency distribution.
FOCAL8-171.6	LAMB	(Lambda) Lambda is an asymmetric non-parametric measure of relationship between two nominal-level variables which have been grouped into an r by c table.
FOCAL8-171.7	MOVE	(Moving Averages) Analysis of the trend exhibited by a set of data over a period of time is often done by means of a moving average. The size (n) of a sample is indicated; interim means and standard deviations are produced.
FOCAL8-171.8	PEAR	(Pearson's r and regression) This program computes the means and standard deviations, regression equation constants, and the correlation between interval data pairs.



PROGRAM NUMBER	PROGRAM NAME	COMMENTS
FOCAL8-171.9	RANT	(Randomization test) For extremely small samples of interval level data obtained from two groups, the randomization test is a powerful non-parametric test. This program calculates all the possible mean score differences for the data.
FOCAL8-171.10	RUNS	(Runs test) The runs test is used to determine whether an ordered series of dichotomous data differs significantly from a random set of observations of the same size.
FOCAL8-171.11	STPC	(Standardized percentages) One technique of multivariate analysis for tabular distributions involves computing the average amount of association in a set of equal sized tables. The program will combine and weight a set of tables, printing out standardized percents and frequencies.
FOCAL8-171.12	DESC	(Descriptive Statistics - Univariate Distributions) The mean and standard deviation, variance, measures of skewness and kurtosis, sums of scores, sums of squares, and sample size are given for either grouped or ungrouped data.
FOCAL8-171.13	ETAN	(Eta and Analysis of Variance) The Eta statistic is used for measuring the association between nominal and interval variables, and also for determining whether a curvilinear relationship is present. In addition to eta, category means and standard deviations are calculated. A simple one-way analysis of variance is produced. Grouped or ungrouped data may be analyzed.
FOCAL8-171.14	TTST	For two samples of interval data, grouped or ungrouped, sample means and overall means and standard deviations are calculated. The F-ratio of variances, t-score for significance of difference between means, degrees of freedom and standard error are calculated.

# BINO

DECUS Program Library Write-up

DECUS NO. FOCAL8-171.1

## ABSTRACT

The BINO program calculates cumulative and exact probabilities for samples of size  $N$  where a level of success has been specified. Binomial probabilities are frequently used in statistical hypothesis testing. Probabilities less than .00000 are not printed.

## LANGUAGE

FOCAL

CORE MEMORY

4K

## MINIMUM MACHINE REQUIREMENTS

A Digital Equipment Corporation PDP-8 computer with teletype.

## INPUT REQUIRED

Sample size, proportions of 'successes'.

## OUTPUT

Cumulative and exact probability of  $X$  successes where  $X$  ranges from zero to the size of the sample.

## CHIS

DECUS Program Library Write-up

DECUS NO. FOCAL8-171.2

### ABSTRACT

Chi square is a program that is used to determine whether or not there is a significant difference between the observed number of cases falling into a particular category and the expected number, given independence.

Chi square can deal effectively with only two or more nominal categories. Although it can be used for other levels of data, there are usually other tests which can be used to better advantage. Chi square does not apply when the expected frequencies are less than five. In this case, however, by manipulating the size of the table it can be made suitable for this test.

### LANGUAGE

### FOCAL

### CORE MEMORY

4K

### MINIMUM MACHINE REQUIREMENTS

A Digital Equipment Corporation PDP-8 computer with teletype.

### RESTRICTION

Table may not have more than six columns.

### INPUT

The program will start by asking how many rows and columns are in the matrix. It will continue by asking for the values of each cell of the matrix.

### OUTPUT

After responses are made the program will print out the following information:

1. The observed frequencies in each cell.
2. The row percents (using marginal totals).
3. The column percents (using marginal totals).
4. The total percents (using the total of the entire matrix).
5. The degrees of freedom.
6. The chi square value.

7. The marginal row percents.
8. The marginal column percents

#### REFERENCES

Hubert Blalock, Jr., Sociel Statistics, McGraw-Hill, (1960). Lillian Cohen, Statistical Methods for Social Scientists, Prentice-Hall (1954). Ted Anderson and M. Zelditch, A Basic Course in Statistics, Holt, Rinehard (1968).

## ABSTRACT

Fisher's exact probability test is an extremely powerful non-parametric technique for analyzing discrete data (either nominal or ordinal) when the sample size is too small for using chi-square, ( $N \leq 20$  or when the smallest expected frequency is  $\leq 5$ ). Fisher's test is appropriate when the data is represented by frequencies in a  $2 \times 2$  contingency table. The scores must fall into one or the other of two mutually exclusive groups from two independent random samples. The groups may be any two independent groups, such as experimentals and controls, Democrats and Republicans, males and females, etc.

The test determines whether the two groups differ in the proportion with which they fall into two classifications.

The probability figures (region of rejection) is one-tailed. If the user desires a two-tailed test, the probability level must be doubled.

LANGUAGE - FOCAL

CORE MEMORY - 4K

## MINIMUM MACHINE REQUIREMENTS

A Digital Equipment Corporation PDP-8 computer with teletype.

## RESTRICTIONS AND LIMITATIONS

The user may enter no more than four (4) cell frequencies.

## INPUT REQUIRED

The user should enter raw data, cell by cell, one row at a time. After entering each cell, the user should press either the ALT MODE key or the SPACE bar.

## OUTPUT

The program will print the exact and cumulative probability of obtaining any matrix where the marginal frequencies are considered to remain constant for the user's cell frequencies. Since in a  $2 \times 2$  table there is only one degree of freedom, once cell 'a' is assigned a value for the user's problem the remaining cell frequencies can be derived. Thus, the program prints all possible values that cell 'a' may take for the user's problem and computes the exact and cumulative probability of obtaining all possible matrices.

The original table (the matrix entered by the user) is identified and the sample size is printed.

## REFERENCE

T.R. Anderson and M. Zelditch, Jr. A Basic Course in Statistics, Holt, Rinehard, and Winston, (1968).



## GATD

DECUS Program Library Write-up

DECUS NO. FOCAL8-171.4

### ABSTRACT

Given the cell frequencies in an R by C table, GATD calculates measures of association and correlation:

1. Pearson's R (using column and row indices as category endpoints)
2. Gamma
3. Tau and Tau'
4. Sommer's D Y.X and D X.Y

### LANGUAGE

### FOCAL

### CORE MEMORY

4K

### MINIMUM MACHINE REQUIREMENTS

A Digital Equipment Corporation PDP-8 computer with teletype.

### RESTRICTIONS AND LIMITATIONS

Program will accept larger tables if extended functions or extra core is available.

## ABSTRACT

The Kolmogorov-Smirnov Test measures the extent to which two distributions differ and the statistical significance of the difference. It is primarily used when the independent variable is of the ordinal level of measurement. However, it may also be employed as a univariate descriptive measure.

When used as a bivariate measure the following procedure is followed: First, the conditional distributions for both categories of the dichotomous variable are formed. Next, the proportion of cases in each conditional category of the dependent variable is calculated and these proportions are then cumulated. Thirdly, the difference between each of the cumulative proportions for each category of the conditional distributions is calculated. From these differences the difference with the largest absolute value is selected and compared to a probability distribution of differences. If the difference selected is greater than the difference from the probability distribution at a given level of probability then we may conclude that the two conditional distributions are not statistically independent - i.e. they are related in some manner.

When used as a univariate descriptive measure, the above procedure is followed with the exception that instead of comparing two conditional distributions a single distribution of variable is compared with a hypothetical uniform distribution of that variable. When used in this manner, the Kolmogorov-Smirnov Test produces a measure equal to the chi-square measure and its significance may be measured by using the chi-square distribution.

## LANGUAGE

## FOCAL

## CORE MEMORY

4K

## MINIMUM MACHINE REQUIREMENTS

This program requires a Digital Equipment Corporation PDP-8 computer with 4K of Core Memory.

## USING THE PROGRAM

The program may be read-in using the normal procedures for loading a FOCAL program tape. To begin executing the program simply type the FOCAL command GO and the RETURN key. The program begins by asking whether you have one or two samples. If you wish to use the program as a univariate measure, type 1. If you wish to use the program as a bivariate

measure, type 2. It will then ask for the number of categories of the dependent variable and the frequencies in each of the categories; first for the first category of the independent variable and secondly for the second category of the independent variable.

## OUTPUT

Frequencies for each cell of the two conditional distributions (in the case of the univariate measure for the single distribution), cumulative proportions for each cell of the conditional distributions, and the difference between the cumulative proportions are printed. Also, the largest difference between the cumulative proportions is given and the values of the probability distribution of differences for the .10, .05, and .01 levels of significance are given.

## REFERENCE

Theodore R. Anderson and Morris C. Zelditch, A Basic Course in Statistics, Holt, Rinehart, and Winston, New York (1968), pps 269-272.

# KOLMOGOROV-SMIRNOV TESTS

ONE OR TWO SAMPLES (TYPE '1' OR '2'):2  
HOW MANY CATEGORIES:3

TYPE IN THE OBSERVED FREQUENCIES

FOR SAMPLE= 1:10 :20 :30

FOR SAMPLE= 2:30 :20 :10

	CELL	SAMPLE 1 FREQ	SAMPLE 2 FREQ	SAMPLE 1 CUM PROP	SAMPLE 2 CUM PROP	DIFFERENCE
=	1=	10=	30=	0.167=	0.500=-	0.333
=	2=	20=	20=	0.500=	0.833=-	0.333
=	3=	30=	10=	1.000=	1.000=	0.000

THE LARGEST DIFFERENCE (D) IS= 0.333

2-TAILED CRITICAL D VALUES FOR LARGE N

P= .10 P= .05 P= .01  
D= 0.223 D= 0.248 D= 0.298

1-TAILED CHI SQUARE= 13.33 D.O.F.=2  
\*GO

# KOLMOGOROV-SMIRNOV TESTS

ONE OR TWO SAMPLES (TYPE '1' OR '2'):1  
HOW MANY CATEGORIES:3

TYPE IN THE OBSERVED FREQUENCIES:30 :30 :30

	CELL	OBSERVED FREQ	UNIFORM FREQ	OBSERVED CUM PROP	UNIFORM CUM PROP	DIFFERENCE
=	1=	30=	30.0=	0.333=	0.333=	0.000
=	2=	30=	30.0=	0.667=	0.667=	0.000
=	3=	30=	30.0=	1.000=	1.000=	0.000

THE LARGEST DIFFERENCE (D) IS= 0.000

2-TAILED CRITICAL D VALUES FOR LARGE N

P= .10 P= .05 P= .01  
D= 0.129 D= 0.143 D= 0.172

\*

ABSTRACT

Lambda is an asymmetric measure of association between two nominal level variables which are arranged in a bivariate table distribution. Lambda is based on percentage improvement in prediction.

LANGUAGE

FOCAL

CORE MEMORY

4K. Program could accept a larger table if additional core is available.

MINIMUM MACHINE REQUIREMENTS

A Digital Equipment Corporation PDP-8 computer with teletype.

INPUT REQUIRED

Number of Rows and Number of Columns in the table, and the cell frequency counts.

OUTPUT

Lambda using rows and columns as independent variables.



## MOVE

DECUS Program Library Write-up

DECUS NO. FOCAL8-171.7

### ABSTRACT

The MOVE program calculates, at a user set interval, interim means and standard deviations in addition to a grand mean and a grand standard deviation of a set of data.

### LANGUAGE

### FOCAL

### CORE MEMORY

4K

### MINIMUM MACHINE REQUIREMENTS

A Digital Equipment Corporation PDP-8 computer with teletype.

### RESTRICTIONS AND LIMITATIONS

Entering a value for "number of observations per average" greater than 22 results in an error message indicating that the variables used have filled the storage register.

### INPUT REQUIRED

1. M, "number of observations per average."
2. First M data values.
3. After first M data values, one new data value at a time.
4. Typing -9999 for a new data value indicates end-of-data.

### OUTPUT

1. After first M data values have been entered, mean and standard deviation of that group are calculated and printed in columns.
2. When a single new data value is entered, a new standard deviation, based on the last M values, is calculated and printed.
3. Upon completion of entering data, sum scores, sum of the squares, grand mean, standard deviation and size of the set of data are calculated and printed.

## ABSTRACT

This program performs simple linear correlation and regression and provides all regression equation terms. Additionally, it computes the standard error of the estimate and provides options for entering further data to be processed, predicting an X value given a Y value, and predicting a Y value given an X value.

## LANGUAGE

## FOCAL

## CORE MEMORY

4K

## MINIMUM MACHINE REQUIREMENTS

A Digital Equipment Corporation PDP-8 computer with teletype and punch paper.

## RESTRICTIONS AND LIMITATIONS

The only limit on the amount of data which can be entered for processing is that imposed by the computer's memory capacity. Additionally, no value of -9999 may be entered as a datum.

## INPUT REQUIRED

Raw paired data are required.

## OUTPUT

The basic output includes: Pearson's  $r$ ,  $r^2$ , the mean of X, the mean of Y, the standard deviation of X, the standard deviation of Y,  $b_{X.Y}$ ,  $b_{Y.X}$ ,  $a_{X.Y}$ ,  $a_{Y.X}$ , the standard error of the estimate of X on Y, and the standard error of the estimate of Y on X. Three optional output features exist: the user's ability to feed more data for processing at the end of each computation of the above features, the ability to predict a Y value given an X value, and the ability to predict an X value given a Y value.

## REFERENCES

Hubert M. Blalock, Social Statistics, McGraw-Hill (1960); Dennis J. Palumbo, Statistics in Political and Behavioral Science, Appleton-Century-Crofts (1969).

Chapters 17 and 18 in Blalock; Chapter 9 in Palumbo.

## ABSTRACT

For extremely small sample interval data obtained from two different groups, the randomization test is a powerful nonparametric test. This program calculates all the possible mean score differences.

## LANGUAGE

## FOCAL

## CORE MEMORY

## 4K

## MINIMUM MACHINE REQUIREMENTS

A Digital Equipment Corporation PDP-8 series computer with 4K memory; list teletype; punch paper tape.

## RESTRICTIONS AND LIMITATIONS

The major limitation of the program is that it is only capable of handling small samples. Input containing more than 26 elements will take a large amount of computer time. The program may be used with confidence only for interval data, this is a restriction of the test itself not the program.

## INPUT REQUIRED

The input for the RANT program consists of two parts. The first portion is for the total number of elements in each of two groups of data. For example, if there were 3 members in group 1 and 3 members in group 2, the operator would punch the 3 key when he is asked to give the number of elements in group 1 and the 3 key, again, when he is asked "and group 2?" The second portion consists of the operator "punching in" elements of each group. The input is totally numeric except for a possible plus or minus sign.

## OUTPUT

RANT will print out each of the possible mean score differences of your data with four elements to a row and as many columns as is required to complete the output. The number of outputs and the overall mean are also printed with their respective labels.

## REFERENCE

Siegel, Nonparametric Statistics.

## RUNS TEST

DECUS Program Library Write-up

DECUS NO. FOCAL8-171.10

### ABSTRACT

RUNS TEST performs the Wald-Wolfowitz runs test on user's data. The runs test tests the hypothesis that a sample has been randomly drawn from an ordered population against the hypothesis that they are drawn from a randomly ordered different population.

### LANGUAGE

### FOCAL

### CORE MEMORY

4K

### MINIMUM MACHINE REQUIREMENTS

A Digital Equipment Corporation PDP-8 with 4K memory, teletype terminal, punched paper tape reader.

### RESTRICTIONS AND LIMITATIONS

Machine: Maximum array length, virtually unlimited, program processes one case at a time.  
Program: Approximate Z score printed in output as reflecting the difference between sample and a normal approximation to the sampling distribution of the expected number of runs is calculated on the basis of the assumption of a large (N greater than 30) sample size.

### INPUT REQUIRED

User must order the cases in his data on the basis of either increasing or decreasing score on the test variate prior to running RUNS TEST. All one or two character alphanumeric codes that uniquely define the two samples are legal input data. If input data is numerical codes, program will class all codes at or above a cutpoint defined by user into one sample, and those below the cutpoint into the second sample.

### OUTPUT

RUNS TEST prints the number of cases, number of runs, mean length of runs, and standard deviation of run lengths for each of both samples. For the combined samples the program prints the total number of observations input, the total number of runs, the mean of the sampling distribution, and a Z score reflecting the difference between sample data and a normal approximation to the sampling distribution of the number of runs.

### REFERENCE

Sidney Siegel, Nonparametric Statistics for the Behavioral Sciences, McGraw-Hill (1956).



## ABSTRACT

Calculation of standardized percentages is a statistical technique used to determine the average amount of association in several same-sized tables. The program combines the tables into one table with each of the contributing control tables contributing proportionately to the standardized table according to the number of observations per table.

## LANGUAGE

## FOCAL

## CORE MEMORY

4K Availability of additional core enlarges the number of size of tables which can be standardized.

## RESTRICTIONS AND LIMITATIONS

Program does not require extended functions, eliminating them increases the number or size of the tables which can be standardized.

## INPUT REQUIRED

Cell frequencies for each table are input after indicating the number of tables, and the rows and columns in the tables.

## OUTPUT

Standardized percentages and frequencies for the standardized table are printed out row by row.

## REFERENCE

Anderson and Zelditch, A Basic Course in Statistics, Holt, (1968).



## ABSTRACT

Computation of the mean and the standard deviation; measures of skewness and kurtosis, sums of scores, sums of squares and sample size are given for either grouped or ungrouped data.

## LANGUAGE

## FOCAL

## CORE MEMORY

4K

## MINIMUM MACHINE REQUIREMENTS

A Digital Equipment Corporation PDP-8 computer with teletype.

## RESTRICTIONS AND LIMITATIONS

There is no limit to the number of values which may be entered.

## INPUT

At the start of program execution, a question is asked if the data are in the form of grouped frequencies. After typing N or NO (for ungrouped frequencies) and YES, the complete word (for grouped frequencies), instructions on how to proceed will appear on the teletype.

Simply pressing the space-bar after every entry automatically sets up the machine for the next entry. Either a terminator (,) or a space entered after X automatically registers a colon for the entry of the particular value's f.

## REFERENCE

Cristina P. Parel, et al, Introduction to Statistical Methods, Quezon City: University of the Philippines, (1965).

## ETAN (Eta/analysis of Variance)

DECUS Program Library Write-up

DECUS NO. FOCAL8-171.13

### ABSTRACT

Etan calculates as a percentage the variation within each variable X against the grand total variation of Y. Also, means and standard deviations are calculated, and a simple one-way analysis of variance table.

### LANGUAGE

### FOCAL

### CORE MEMORY

4K

### MINIMUM MACHINE REQUIREMENTS

A Digital Equipment Corporation PDP-8 computer with teletype.

### RESTRICTION

With extended functions deleted additional categories may be entered.

### INPUT

The machine will ask, "How many categories:". Type a number (e.g. 4). The machine will then print that many times, "Type in data for category= X." Type in the numbers, leaving a space at the end of each number or a RETURN if you get too close to the edge of the paper. Type -9999 to end each category. It is not counted in the tabulations.

### OUTPUT

The program will, for each category, print N, mean, standard deviation. Next it will type Eta-squared and an Anova table. The Anova table gives sum squares and degrees of freedom for the total, between, and within. There is also an estimated variation for the between and the within categories. After all this, F is printed.

### REFERENCES

Anderson and Zelditch, Jr., A Basic Course in Statistics, Holt, Rinehart, and Winston, Inc., (1968). Blalock, Social Statistics, McGraw-Hill Book Company, (1960).

## ABSTRACT

The program calculates t-scores for testing significance of difference between means for two groups. Two t-scores and degrees of freedom are calculated depending on whether equal variances are assumed. Mean standard deviations and sample sizes for each group can be input directly or calculated by the program. If the parameters are not known, the original observations may be from either grouped or ungrouped data.

## LANGUAGE

## FOCAL

## CORE MEMORY

## 4K

## MINIMUM MACHINE REQUIREMENTS

A Digital Equipment Corporation PDP-8 computer with teletype.

## REFERENCE

Hubert M. Blalock, Social Statistics, McGraw-Hill (1960).