U.S. DEPARTMENT OF THE INTERIOR

U.S. GEOLOGICAL SURVEY

NPBAS: A BASIC PROGRAM FOR NONPARAMETRIC COMPARISONS OF TWO OR MORE INDEPENDENT GROUPS OF DATA

By

M.L. Millgate[1/]

OPEN-FILE REPORT

92-369-A Program Documentation and Source Code, (Paper Copy)

92-369-B Program Documentation, Source Code, and Executable Program (360Kb, 5-1/4" Floppy Diskette)

Revised May 15, 1992

[1/] U.S. Geological Survey
Denver, Colorado

CONTENTS

FIGURES

NPBAS: A BASIC PROGRAM FOR NONPARAMETRIC COMPARISONS OF TWO OR MORE INDEPENDENT GROUPS OF DATA

By

M.L. Millgate

Abstract--The program NPBAS is translated and adapted from program NPSTAT (Rock, 1986), a FORTRAN-77 program, which compares the location, dispersion, and overall shape of two or more independent groups (samples) of data by use of the Mann-Whitney and Kolmogorov-Smirnov tests (two groups), the Kruskal-Wallis and Van der Waerden tests (three or more groups), and the squared-ranks tests (two or more groups). An executable program, source code and documentation are provided on 5-1/4 inch, 360Kb diskette and comments, documentation, and source code are provided in this report. Exact or approximate significance levels of test statistics (for two-tailed tests) are calculated and displayed for each applicable test. NPBAS is designed to be used particularly for comparison of non-normally distributed data sets (or distribution not known), data with missing values for more than one input variable, and for data measured on the ordinal and higher measurement scales. NPBAS duplicates numerous error free textbook examples, following the conventions contained in Rock's (1986) program.

INTRODUCTION

The program NPBAS is written and compiled in TURBOBASIC (Boreland International, Inc.) for use on IBM and compatible personal computers. The program is dimensioned and compiled using 640K memory but runs successfully as dimensioned on machines with as little as 512K available memory. Some users may not have access to compilers and must use the program with interpreters. Also, some may wish to modify the source code for their own convenience. Thus, full information is contained both in hardcopy and on disk, including a listing of program parameters and variables (Appendix A, below) to facilitate program adaptation to other user systems and preferences. Program description, documentation, and source code are contained in this report (paper copy). An executable file (NPBAS.EXE), source code (NPBAS.BAS), and documentation (NPBAS.DOC) are provided on 360Kb, 5-1/4 inch, diskette. The program (Appendix B, below) and input-output files are in American Standard Code (ASCII) following BASIC language conventions. The compiled version here presented does not require a coprocessor and the keyboard break (^C) is functional. A pretty-printed output file can be written to hard or floppy disk to maximize program portability such that hardcopy (for the parameters programmed) can be obtained from any printer.

NPBAS is essentially a translation of the program NPSTAT of Rock (1986), from FORTRAN-77 language to BASIC, retaining the existent methodology and construction of NPSTAT. NPBAS is written in BASIC as a matter of author convenience, making the essentials of the work of Rock (1986) available through the more commonly known language of BASIC and use of less expensive, more widely distributed compilers. In all instances the previous judgements of Rock (1986) as to statistical method and choice of critical values is retained. The names of NPSTAT program variables and parameters are retained to facilitate ease of program understanding and modification. Salient modifications of convenience made in NPBAS are (1) conversion from optional interactive data entry to batch processing, (2) inclusion of optional output of results (a read file which can be subsequently printed) at run time, and (3) control of intermediate sequential file (dump file of selected parameters) output exclusively by an input file parameter. Therefore, input files must be constructed prior to working entry in NPBAS (any line or full-screen editor or word processor will do), connected printers are not required for analysis; and the intermediate file, although readily decipherable after some experience, requires a separate print program to produce a formatted, pretty-printed product. Although single precision is adequate for all calculation except that in function PPND77 (Appendix B, below), double precision is conveniently used throughout the program. For small data arrays and small groups within arrays, most explanatory messages are not printed to output file, but rather are shown on screen at the time of activation of control parameters within the program.

NPSTAT was devised to provide two-tailed, nonparametric statistical comparisons for independent groups of geochemical data, having in mind the statistical analysis of large arrays of major, trace, and rare-earth elements (Rock, 1986, p. 757, 760, and 763). The nature of statistical analysis by standard hypothesis testing permits the NPBAS (NPSTAT) program to be used in the analysis of data arrays germane to any subject. The data arrays must be applicable to the subject variable and arrays must be adaptable to program and hardware limitations. The nature and detail of the data array(s), relation of data array(s) to subject matter, and validity and use of conclusions drawn from test statistics and significance levels provided by the program are entirely within the province of the program user.

NPBAS contains six commonly used nonparametric statistical tests (Mann-Whitney, Kolmogorov-Smirnov, Kruskal-Wallis, Van der Waerden and two squared-ranks tests) which can be used in a wide variety of situations, the only common denominators being the limitations of program and statistical method. The program parameters can be

redimensioned to accommodate larger (smaller) data arrays, can be modified to satisfy the different BASIC varieties (language modification for use within the BASIC interpreters or different compilers required), and other statistical tests can be programmed and appended to the existing program skeleton.

Observations must be amenable to the ordinal or higher level of measurement (Siegel, 1956). No provision is made for observations which do not admit of ranking relative to others within a data array. Rock (1986, p. 757) uses the word "group" in preference to "sample" because of the confusion arising between the geological sample and statistical sample. The word "group" here refers to an object or data set such that a "two-sample test" of statisticians becomes a "two-group test". The word "parameter(s)" is used to identify control parameters and program variables within the program environment, whereas the word "variable(s)" is reserved to identify the input (subject) variables(s) identified and named in the input, intermediate, and output files.

DATA INPUT AND INPUT FILE CONVENTION

NPBAS subjects each input variable to the statistical tests determined by parameters in the input file, whether several variables are input each in a single file or are input in the same file. Thus, large data arrays consisting of many input variables may be prepared for input in a number of smaller files, consisting of fewer variables, in order to stay within the memory limitations of the hardware employed.

As many as 15 input variables consisting of 15 groups of 500 observations in each group are allowed. A larger number of entries in any file requires program change and recompilation prior to data analysis. To analyze larger data arrays redimension: (1) for more than 15 groups increase all arrays of size 15 and 16 such as Chi999, R, Ngroup, Nvalue and Group and add higher values to the Data statements in subroutine Chisig; (2) for more than 500 observations increase arrays Rank, Calc, and the second dimension of array Ox(q,j); and (3) for more than 15 variables increase the first dimension of array Ox(q,j). If more than 200 tied values are encountered, NPBAS will abort but will continue with the next input variable. To calculate with more than 200 tied values the array Ntie must be redimensioned. The program has not been tested for data arrays having more than 222 observations for one input variable (in one file) but is thought to be dimensionally sound.

Missing data are identified at the outset by parameter in the input file (Fig. 1). Explanation of program parameters is contained in the parameter listing (below) and some are

explained in more detail in the print statements within the program which are displayed on screen at run time for each input file.

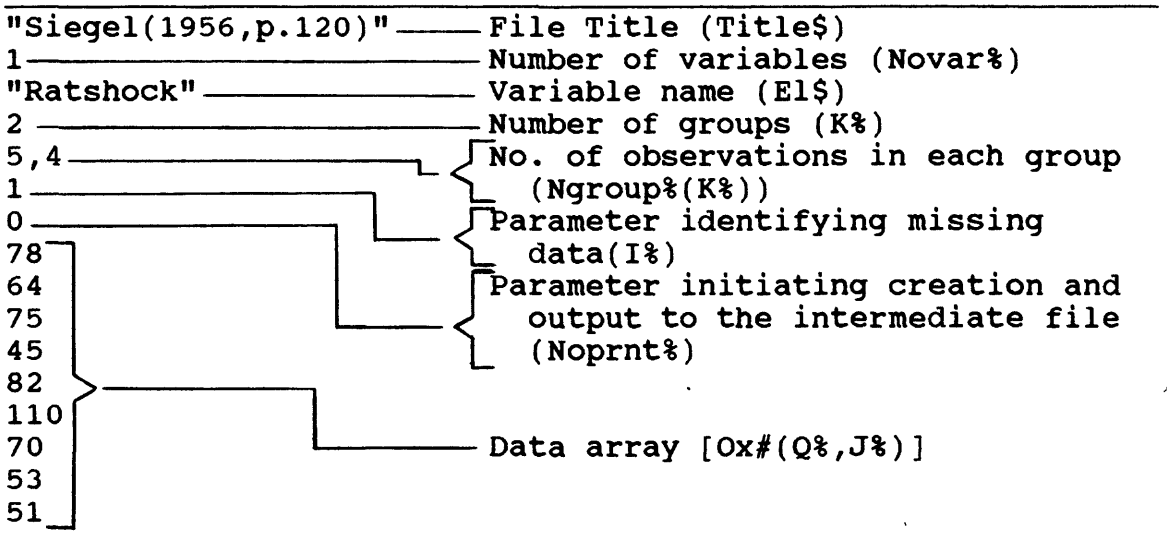

Figure 1.--Example of an input file with input parameters and input data array identified. Symbols (e.g., %) after parameters indicate data type for programming purposes.

Figure 1 shows an annotated input file for 1 variable in 2 groups. To obtain an intermediate file to disk code the "Noprnt" parameter other than zero. The "I" parameter is coded plus one (1) so that zeros, blanks (not spaces), and negative values in the input data array are treated as missing data within the program; coded zero (0) so that negative values are treated as missing data, but zeros and blanks are treated as real data values within the program; or coded minus one (-1) so that zeros, blanks (read as zeros), negative values, and positive values in the input data array are treated as real data within the program. Thus, the "I" parameter permits variations in the coding of data to allow for missing observations in the input data arrays. If very small positive or negative real numbers (numbers very close to zero less than plus or minus 0.0001 and thus smaller than the "Zero" parameter) are included in the input array, the real values of the parameter "Zero" must be revised to include input values within the specified range(s) and the program recompiled. Similar comments apply if very large negative numbers (greater than -10,000,000) are included in input data arrays.

Arrangement of parameters and data within the input file(s) follows the usual BASIC conventions--commas and carriage returns are read within the NPBAS program as delimiters. Other than the usual conventions, data may be arranged sequentially as the user wishes. Figure 2 shows an

annotated input file for 1 variable in 3 groups with the data array arranged somewhat differently than Figure 1.

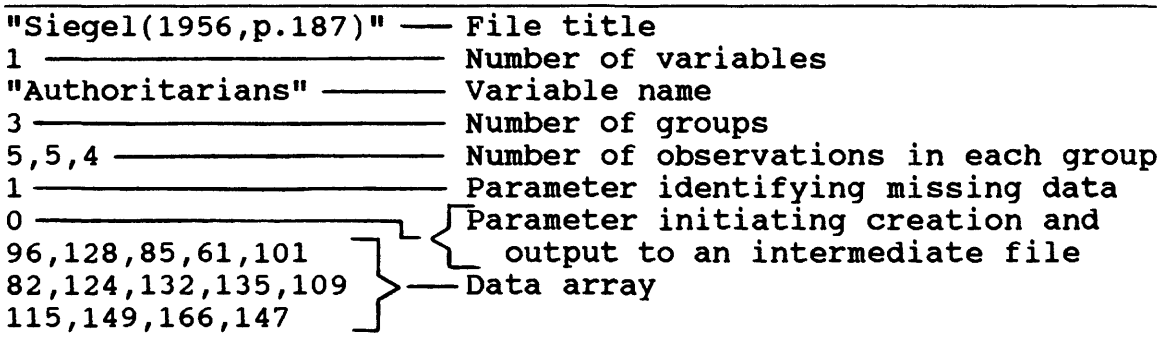

Figure 2.--Example of an input file with input parameters and input data array (variation) identified.

Multiple variables in one input file are similarly arranged, the only requirement being that all groups in the file must be in the same format. Figure 3 shows the first few lines of an input file for 8 variables in 3 groups each, consisting of 240 observations in the input file. The input file must bear a standard filename extension (.nps) because the program is designed to accept no other. If different filename extensions are desired the program must be changed and recompiled.

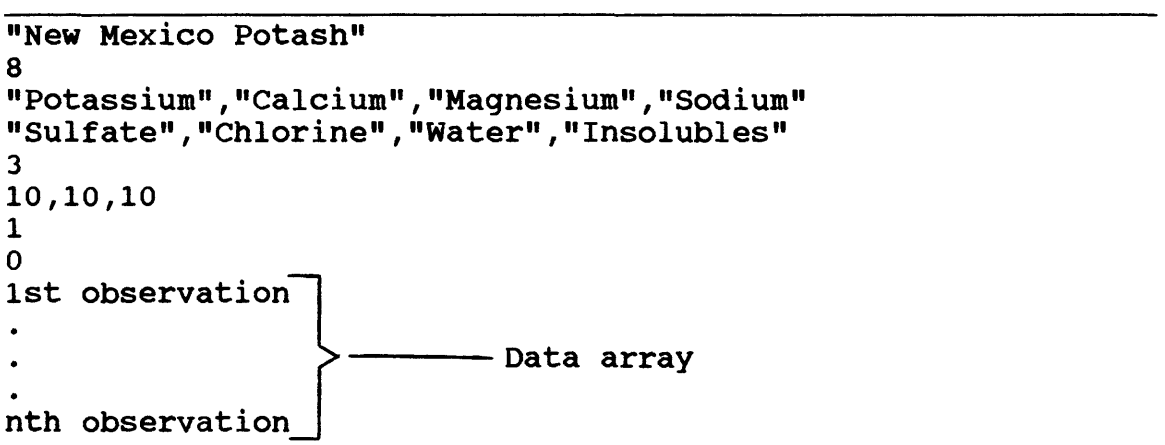

Figure 3.--Example of an input file (partial) showing file convention for multiple input variables and the input data array identified.

INTERMEDIATE FILE CONVENTION

Following the lead of Rock (1986), provision is made for an intermediate file in which some of the intermediate arrays are dumped for user convenience. The method of specifying the creation of an intermediate file is discussed above. The intermediate file is of value to users who desire more

information than is provided by the output file, which is largely summary, or have need of intermediate values for other input or further analysis. Figure 4 shows an annotated example of an intermediate file with parameters and arrays identified as created and used in the 3 nonparametric tests for 2 independent groups of data.

The intermediate filename is provided by program during file creation. The filename is the same as the input file and the filename extension is coded ".npi" to discriminate the intermediate file from input and output files. The same file title is continued to enhance parallelism with input and output. Selected NPBAS parameters are carried into the intermediate file, largely for convenience in the event of future programming. There is, of course, no barrier to a change in parameter designation in any future program.

Figure 5 shows an annotated example of an intermediate file with arrays identified as created and used in the 3 nonparametric tests for 3 or more groups. Parameters not identified in the first few lines of Figure 5 are the same as those in Figure 4.

OUTPUT FILE CONVENTION

Output files (designated by input filename plus the filename extension ".npo") are summary files of the results of analysis containing the test statistic(s) and the significance level of the statistic, derived from tables of significance within the program. For the Smirnov test, the exact probability for rejection of the null hypothesis is calculated, including those instances where decision rules are inconsistent or tables are incomplete. Available tables tend to be inconsistent or incomplete for small groups (Rock, 1986, p. 762). The Mann-Whitney test can assess outlying values because the test is executed where $n_1 = 1$ (the outlier) provided $n_2 > 18$ (the comparison group). Messages are printed to screen at run time if the input groups are too small to permit analysis. Figure 6 is an example of an output file for 1 variable in 2 groups. For more than three groups program execution proceeds provided n_1 is greater than 1. For three groups execution proceeds provided $n_1 + n_2 + n_3$ is greater than 5. Messages are printed to screen at run time if groups are too small to permit testing. Figure 7 is an example of an output file for 1 variable in 3 groups.

SIGNIFICANCE TESTS

The conventions of Rock (1986) are carried through the significance levels displayed on screen and in output files. Two-tailed significance levels are assigned through critical (99.99, 99.9, 99, 98, 97.5, 95, 90, or 80 percent) values of

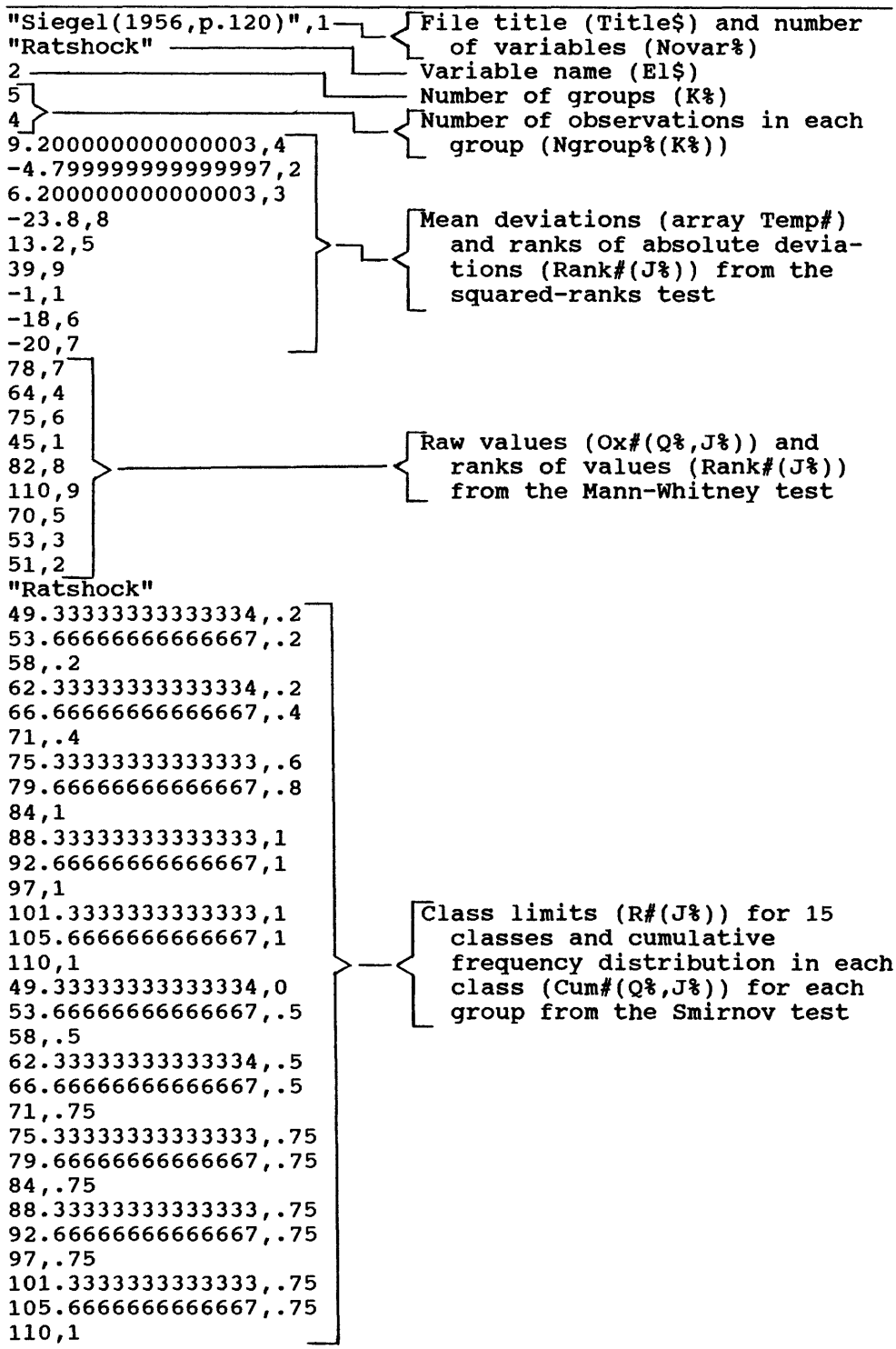

Figure 4.--Example of an intermediate file in two groups with parameters, arrays, and array origins identified. Symbols after parameters (e.g., %) indicate data type for programming purposes.

```
"Siegel(1956,p.187)",1
"Authoritarians"
3
5
5
4
2.200000000000003,1
34.2,13
-10.8,7
-32.8,12
7.200000000000003,4
-34.40000000000001,14
7.599999999999994,6
15.59999999999999,8
18.59999999999999,9
-7.400000000000006,5
-29.25,11
4.75,3
21.75,10
2.75,2
96,4
128,9
83,3
61,1
101,5
82,2
124,8
132,10
135,11
109,6
115,7
149,13
166,14
147,12
"Authoritarians"
4,.2666666666666667,-.6229332117872021,0,1
9,.6,.2533471745719368,0,2
3,.2,-.8416600569365983,0,3
1,6.666666666666667E-002,-1.501085945048739,0,4
5,.3333333333333333,-.4307283751114191,0,5
2,.1333333333333333,-1.110961771242894,0,6
8,.5333333333333333,8.365173320237468E-002,0,7
10,.6666666666666666,.430728375111419,0,8
11,.7333333333333333,.622933211787202,0,9
6,.4,-.2533471745719368,0,10
7,.4666666666666667,-8.365173320237468E-002,0,11
13,.8666666666666667,1.110961771242894,0,12
14,.9333333333333333,1.501085945048739,0,13
12,.8,.8416600569365985,0,14
```

Mean deviations (array Temp#) and ranks of absolute deviations (Rank#(J%)) from the squared-ranks test

Raw values (Ox#(Q%,J%)) and ranks of values (Rank#(J%)) from the Kruskal-Wallis Test

Arrays from the Van der Waerden test

Figure 5.--Example of an intermediate file in more than two groups with the arrays and array origins identified. Symbols after parameters (e.g., %) indicate data type for programming purposes. Arrays from the Van der Waerden test consist of ranks (T1#), quantiles (Dprob#), normal scores (Rank#(Q%)), operation parameters (Ifail%), and line counters (Q%). For "Ifail", a zero (0) indicates correct operation by function PPND77.

```
      NPBAS RESULTS FOR 2 INDEPENDENT GROUPS OF DATA

       FILE TITLE:  Siegel(1956,p.120)
       VARIABLE NAME: Ratshock
       Total Number of Observations: 9 in groups of 5
         and 4
       Number of Observations Ignoring Missing Data: 9

       MANN-WHITNEY U TEST:     Z = Undef.
         U =     9.00  U' =     11.00       %Sig. = <80.0%
       SMIRNOV TEST:
         Dmax = 0.350  Prob. = 12.698       %Sig. = Undef.
       SQUARED-RANKS TEST:
         T or T1 =  118.000                 %Sig. = <80%

     Wherever numbers of values in both groups <21, Z in
       Mann-Whitney test is undefined.

     In Kolmogorov-Smirnov test, %sig. undefined unless
       n1>79 and n2>99 or n1=n2; approximations then
       unavailable.
```

Figure 6.--Example of an output file for two groups showing summary results of analysis: Z, value of unit normal approximation; U and U', values of the Mann-Whitney U statistic; Dmax, maximum deviation for the Kolmogorov-Smirnov test; Prob, exact probability (in percent) for null hypothesis rejection based on Dmax; T or T1, squared ranks test statistic including arrays having tied (T) and non-tied (T1) values; %Sig., significance level for rejection of the null hypothesis; and some explanatory notes.

each statistic, depending on the availability of tables or numerical approximations. For example, a displayed value coded <90% indicates that the null hypothesis (of equal locations, dispersions, or distributions) cannot be rejected with 90% confidence, and implies that no 80% value is available. A displayed value coded >90% indicates the null hypothesis can be rejected with 90% but not 95% confidence. Messages are printed in most places in the output file and on demand to screen where significance levels cannot be output or are uncertain. Where arrays contain too few observations for testing, implausibly large values (e.g., 10,000) for some tests are simply carried forward from the program and printed to the test statistic; the significance level (%Sig) will indicate that no test was performed.

```
NPBAS RESULTS FOR 3 OR MORE INDEPENDENT GROUPS OF DATA

     FILE TITLE:  Siegel(1956,p.187)
     VARIABLE NAME: Authoritarians
     Total Number of Observations in 3 Groups:  14
     Number of Observations in 3 groups
       Ignoring Missing Data:  14

     KRUSKAL-WALLIS TEST:
       H =    6.406                    %Sig. = >95%
     VAN DER WAERDEN TEST:
       T1 =    6.480                   %Sig. = >95.0%
     SQUARED-RANKS TEST:
       TSQ =    0    T2 =    0.276     %Sig. = <80.0%
```

Figure 7.--Example of an output file for three or more groups showing results of analysis: H, value of the Kruskal-Wallis test statistic; T1, value of the Van der Waerden test statistic; TSQ and T2, value of sum of ties function and value of the squared-ranks test statistic; %Sig., significance level for rejection of the null hypothesis.

SUMMARY REMARKS

The program has not been extensively tested for files containing numerous input variables of large data arrays (large groups). However, the program yields values and conclusions in agreement with 24 error-free examples (small data sets) in the texts of Conover (1980), Siegel (1956), Till (1974), and Walpole and Myers (1978). Further, the NPBAS nonparametric tests yield results consistent with several of Till's (1974) examples which illustrate parametric and nonparametric methods of statistical testing.

The details of the statistical tests contained in program are not discussed here. Rather the reader is referred to Rock (1986) for brief descriptions of tests and applicability of nonparametric testing as well as a summary of the recent and classical references with respect to nonparametric methods.

References (below) are not all cited in text. Rather, some references (Beasley and Springer, 1977; Beyer, 1968; Kim and Jenrich, 1970; Kruskal and Wallis, 1952; and Steel and Torrie, 1980) are cited in the appendices (below) and are exclusively the citations of Rock (1986). These references are included here because some contain background material, methods, and tables of critical values that can by difficult to identify and obtain, but were not used in direct preparation of this report or the program. For citations in full context, the reader is referred to Rock (1986).

REFERENCES

Beasley, J.D. and Springer, S.G., 1977, The percentage points of the normal distribution: Applied Statistics, Journal of the Royal Statistical Society, Series C, v. 26, p. 118-120.

Beyer, W.H., 1968, Handbook of tables for probability and statistics, (2nd ed.): Chemical Rubber Company, Cleveland, Ohio, 642 p.

Conover, W.J., 1980, Practical non-parametric statistics, (2nd ed.): John Wiley & Sons, New York, 550 p.

Kim, P.J. and Jennrich, R.I., 1970, Tables of the exact and approximate sampling distribution of the two-sample Kolmogorov-Smirnov criterion, *in* Harter, H.L. and Owens, D.B., eds., Selected tables in mathematical statistics: Markham, Chicago, p. 79-170.

Kruskal, W.H. and Wallis, W.A., 1952, Use of ranks in one criterion variance analysis: Journal American Statistical Association, v. 47, no. 260, p. 583-621.

Rock, N.M.S., 1986, Npstat: a fortran-77 program to perform nonparametric variable-by-variable comparisons on two or more independent groups of data: Computers & Geosciences, v. 12, no. 6, p. 757-777.

Siegel, S.S., 1956, Nonparametric statistics for the behavioral sciences: McGraw-Hill, New York, 312 p.

Steel, R.G.D. and Torrie, J.H., 1980, Principles and procedures of statistics: a biometrical approach, (2nd ed.): McGraw-Hill, International Student Edition, New York, 633 p.

Till, Roger, 1974, Statistical methods for the earth scientist: Macmillan, London, 154 p.

Walpole, R.E., and Myers, R.H., 1978, Probability and statistics for engineers and scientists, (2nd ed.): Macmillan, New York, 580 p.

APPENDIX A

NPBAS Parameter Listing--Explanation of Arrays and Parameters in Alphabetical Order

Aibar---Carries mean normal scores for each group in Van der Waerden test.

Calc----Contains temporary values input to the ranking subroutine.

Chiapp--Trips a warning message that the Chi-square approximation may not be valid for sample numbers <5 in Kruskal-Wallis test.

Dcr-----Holds the two-tailed critical values of the Kolmogorov-Smirnov statistic at the 99, 95, 90, and 80% levels for $n_1 = n_2 < 41$ (from table A23A, Steel and Torrie, 1980).

Dmax----Carries maximum deviation of the two cumulative step functions for the Kolmogorov-Smirnov test.

Dprob---Carries quantiles for conversion to normal scores in the Van der Waerden test. (NOTE: This is the only double precision parameter required, found in Function PPND77 and equivalent NAG routine G01CEF).

El------Carries the variable (element) name or symbol.

Group---Carries numbers indicating where each group of analyses (data) starts and stops within the complete data file. (Group%(1) goes from 1 to Ngroup%(1), etc.).

H-------Eventually carries the value of the Kruskal-Wallis test statistic.

Hcr-----Carries the critical H values for 3 groups, <6 values in each group (from table A8 of Conover, 1980; note often very different values in table O of Siegel, 1956 and Beyer, 1968, p. 431-2, based on Kruskal and Wallis, 1952). Hcr5 refers to 5 values in largest group, Hcr4 to 4 values, Hcr3 to 3 values (for 2 values, test statistic is written into program). Hcri(j + 1, k) and Hcri(k,j) respectively contain the 95% and 90% critical statistic for i>=j>=k numbers of values.

I-------Identifies the smallest (largest negative) value recognized as real.

Indat---Parameter not used in NPBAS.

Infile--Carries the name of the input file.

Infmt---Parameter not used in NPBAS.

Intfile--Carries the name of the intermediate file.

K-------Carries the number of groups of analysis (data).

Koltes--Trips information messages where Kolmogorov-Smirnov test is inapplicable.

Krutes--Trips information messages where critical Kruskal-Wallis value not available, with small sample numbers.

Min, Max--Carry smaller and larger numbers of values for 2 groups.

Ngroup(K)--Carries the number of analyses (observations) in each K group.

N1n2----The product n_1 * n_2 of the numbers of real values in 2 groups (equals U1 + U2 in the Mann-Whitney test).
Nn1-----The product N(N+1)/2, the correct sum of ranks for N values.
Noprnt--Parameter controls output to the intermediate file.
Novar---Carries the number of variables being input to the program.
Ntie(J)-Carries the array numbers in Ox(M,J) of tied values.
Ntotal--Carries the total number of analyses (records).
Ntotvl--Carries the total number of values for each element (variable) in turn (same value in integer(%) as Totval is in real(#)).
Nul-----Trips various warning messages for small sample numbers.
Nvalue(K)--Carries the number of values for every element (observation) in each group (allowing for missing values). Nvalue%(1) = n_1 of textbook parlance, etc.
Outfile--Carries the name of the output file.
Ox(Q,J)-Carries the input data matrix (Q = variables, J = records).
Prob----Carries the exact probability of the Smirnov test statistic obtained.
R4------Carries the sum of ranks raised to the 4th powers in squared ranks test.
Rank----Carries the ranks of each data-value for each element (observation) in turn.
Ranksm--Carries the sum of ranks (should equal Nn1).
R(K)----Carries the sum of ranks for each group of analyses (observations) in the Kruskal-Wallis test, and class boundaries for determining cumulative step functions in the Kolmogorov-Smirnov test.
S2------Carries intermediate sums in Van der Waerden and squared ranks tests.
Sig-----Carries the significance level of Kolmogorov-Smirnov or Kruskal-Wallis statistic.
Sigt----Carries the significance level of the squared ranks test statistic.
Sigu----Carries the significance level of the Mann-Whitney or Van der Waerden statistic.
Sqrank--Allows or prevents execution of the squared ranks test depending on the number of data values.
Stan----Carries the 99.95, 99.5, 97.5, 95 and 90% percentage points of the standard normal distribution, to test correspondingly 99.9, 99, 95, 90 and 80% significance level for the squared ranks test statistic (two-tailed).
T-------Carries sum of tie functions to adjust test statistics for ties, in the Mann-Whitney and Kruskal-Wallis/Van der Waerden tests. Subroutine Rnk yields the sum of (t^3-t), where t is the number of tied values at each rank; actual tie factor is T/12 in the Mann-Whitney test, and (1-T(N^3-N)) in the Kruskal-Wallis test.

T1------Carries the value of test statistic in the squared ranks test.

T1cr----Carries two-tailed critical values of the squared ranks test statistic for 2 groups, <20 total data values (from table A9 of Conover, 1980). NB: T1cr99 and T1cr1 carry the upper and lower tails for a 99% significance test (corresponding to p values of .995 and .005 in Conover's table).

Temp----Used to carry temporary calculated values which are not retained.

Ties----Carries the number of ties for each stage of the ranking procedure.

Title---Title of file containing input data array.

Totval--Carries the total number of non-missing values for each element (variable).

Tsq-----Has the same function as T for the squared ranks test.

Ucr-----Carries two-tailed critical values of Mann-Whitney U statistic for n_2<21 (from Beyer, 1968, p. 406ff). Ucr995 carries 99% and 95% values; Ucr908 carries 90% and 80% values; in each case, Ucr(i+1,j) carries the higher critical level, and Ucr(j,i) carries the lower critical level, for i>=j numbers of values.

Z-------Carries the test statistic against the unit normal distribution for the Mann-Whitney test, with n_2>20 for 2 groups of data; for >2 groups Z ultimately carries the Van der Waerden test statistic.

Zero----Carries the minimum data-value accepted as real: (if Zero = +0.0001, zero and negative values are missing) (if Zero = -0.0001, zero values are real, negative values are missing) (if Zero = -10.00E6, positive and negative values are both taken as real).

Zval----Trips a message that Z is undefined where sample numbers <21 in the Mann-Whitney test.

APPENDIX B
NPBAS Source Code

```
'   Program NPBAS
'     Program to carry out nonparametric comparisons of 2 to
'15 independent groups of major, trace and REE element data.
'All data-groups must be in a SINGLE file in the same format.
DIM Ox#(15,500), Calc#(500), Rank#(500), R#(15), Nvalue%(15),_
    Ngroup%(15), Group%(16)

%FALSE =  0
%TRUE = NOT %FALSE

FileNameTrap:
IF ERR = 53 OR ERR = 64 OR ERR = 76 THEN
  BEEP
  LOCATE 10,25: PRINT "File not found, TRY AGAIN!"
  DELAY 2
  RESUME  FileNameTrap
END IF

CLS
PRINT "DO NOT ENTER the filename EXTENSION (.NPS); extension is"
PRINT "  provided by program then used to verify that the input"
PRINT "  file named is indeed an 'NPBAS file'!"
PRINT
INPUT "Enter FILENAME of INPUT FILE-->", File$
Infile$ = File$ +".NPS"

ON ERROR GOTO  FileNameTrap
 OPEN Infile$ for INPUT as #1
ON ERROR GOTO 0

'-----------------------------------------------------------------
'Read the input file (all groups of observations in one matrix).
'-----------------------------------------------------------------
INPUT #1, Title$, Novar%
  FOR Q% = 1 TO Novar%
    INPUT #1, El$(Q%)
  NEXT Q%
INPUT #1, K%
  FOR Q% = 1 TO K%
    INPUT #1, Ngroup%(Q%)
  NEXT Q%
INPUT #1, I%, Noprint%

'---Determine total number of observations (analyses) and work out
'start and stop values within overall data file for each group.
Ntotal% = 0
Group%(1) = 1
  FOR J% = 1 TO K%
      IF J% > 1 THEN
        Group%(J%) = Group%(J%-1) + Ngroup%(J%-1)
      END IF
    Ntotal% = Ntotal% + Ngroup%(J%)
  NEXT J%
Group%(K%+1) = Ntotal% + 1

FOR J% = 1 TO Ntotal%
```

```
  FOR Q% = 1 TO Novar%
    ON ERROR GOTO MoreM
  INPUT #1, Ox#(Q%,J%)
    ON ERROR GOTO 0
  NEXT Q%
NEXT J%
CLOSE #1

'-----------------------------------------------------------------
'Determine number of groups, write error messages for less than 2
'  or more than 15 groups.
'-----------------------------------------------------------------

IF K% < 2 THEN
  BEEP
  PRINT "        Number of groups of data in file ";Infile$
  PRINT "          less than two, CORRECT THE INPUT FILE !"
  PRINT
  INPUT "Press <ENTER> to Continue" ;Continue$
  GOTO MoreM
END IF
IF K% > 15 THEN
  BEEP
  PRINT "        Number of groups of data in file ";Infile$
  PRINT "greater than 15, RECONSTRUCT INPUT FILE(S) OR NPSTAT PROGRAM!"
  PRINT
  INPUT "Press <ENTER> to Continue";Continue$
  GOTO MoreM
END IF

'-----------------------------------------------------------------
'Abort operations if dimensioned size of array Ox is exceeded.
'-----------------------------------------------------------------
  IF Ntotal% > 500 THEN
    PRINT "More than 500 data entries; OX array in the program"
    PRINT "has been exceeded and must be increased."
    END
  END IF
'-----------------------------------------------------------------
'Determine what identifies missing data.
'-----------------------------------------------------------------
  IF I% > 0 THEN
       Zero# = 0.0001
    LOCATE 10,1
    PRINT "Zeros, blanks, and all negative values have been"
    PRINT "ignored (as missing data) in calculations."
    PRINT
  ELSEIF I% = 0 THEN
       Zero# = -0.0001
    LOCATE 10,1
    PRINT "Negative values have been ignored (as missing data),"
    PRINT "zeros and blanks included (as real data) in"
    PRINT "calculations ."
    PRINT
  ELSEIF I% < 0 THEN
       Zero# = -10.0E6
    LOCATE 10,1
    PRINT "Negative and positive values both treated as real data."
    PRINT
```

```
  END IF
INPUT "Press <ENTER> to Continue", Continue$
'-------------------------------------------------------------------
'Open and write heading to the intermediate file.
'-------------------------------------------------------------------
IF Noprint% <> 0 THEN
  Intfile$ = File$ + ".NPI"
  OPEN Intfile$ FOR OUTPUT AS #2
    WRITE #2, Title$, Novar%
  FOR Q% = 1 TO Novar%
    WRITE #2, El$(Q%)
  NEXT Q%
    WRITE #2, K%
  FOR Q% = 1 TO K%
    WRITE #2, Ngroup%(Q%)
  NEXT Q%
END IF
'===================================================================
'Calculate test statistic for each input variable in turn.
'===================================================================
Nul% = %False: Krutes% = %False: Zval% = %False: Chiapp% = %False
Koltes% = %False: Sqrank% = %False: FileOpen% = %False

FOR M% = 1 TO Novar%

'-------------------------------------------------------------------
'Initialize all summation parameters used only once in cycle.
'-------------------------------------------------------------------
Ranksm# = 0: Ntotvl% = 0: T# = 0: TSQ# = 0
'-------------------------------------------------------------------
'Determine total numbers of real values for all groups and assign
'  zero ranks to missing data.
'-------------------------------------------------------------------
  FOR Q% = 1 TO K%
  Nvalue%(Q%) = Ngroup%(Q%)
    FOR J% =  Group%(Q%) TO Group%(Q% + 1) - 1
      IF OX#(M%,J%) < Zero# THEN
        Rank#(J%) = 0
        Nvalue%(Q%) = Nvalue% (Q%) - 1
      END IF
    NEXT J%
    Ntotvl% = Ntotvl% + Nvalue%(Q%)
  NEXT Q%
'-------------------------------------------------------------------
'Abort calculation for any element (input variable) with no real
'  data-values in any group.
'-------------------------------------------------------------------
  IF Ntotvl% <= 0 THEN
    PRINT
    PRINT "FOR ";El$(M%);" -->: NO real values in input-data matrix."
    INPUT "Press <ENTER> to Continue", Continue$
       GOTO MoreM
  END IF

Totval# = CDBL (Ntotvl%)
'-------------------------------------------------------------------
'Calculate N(N+1)/2, correct sum of ranks for N values, to check
'  Ranksm.
'-------------------------------------------------------------------
```

```
Nn1# = Totval# * (Totval# + 1.0) / 2.0
'-------------------------------------------------------------
'Abort calculation for any element (input variable) with no
'  data-values in one or more groups.
'-------------------------------------------------------------
Max% = 0
Min% = 999

FOR Q% = 1 TO K%
  IF Nvalue%(Q%) <= 0 THEN
    PRINT
    PRINT "For";El$(M%);"  -->: NO values for one or more groups."
    INPUT "Press <ENTER> to Continue", Continue$
    GOTO MoreM
  END IF
'-------------------------------------------------------------
'Determine maximum and minimum numbers of values in all/both groups
'-------------------------------------------------------------
  IF Max% < Nvalue%(Q%) THEN Max% = Nvalue%(Q%)
  IF Min% > Nvalue%(Q%) THEN Min% = Nvalue%(Q%)
NEXT Q%
'-------------------------------------------------------------
'Avoid unnecessary calculations where numbers of values
'  insufficient to calculate test statistics.
'-------------------------------------------------------------
IF Max% <= 1 THEN
  PRINT
  PRINT "For ";El$(M%);" -->: Insufficient values to calculate"
  PRINT "                     any tests."
  BEEP
  INPUT "Press <ENTER> to Continue", Continue$
END IF
IF max% <= 1 THEN GOTO MoreM

IF K% > 2 OR Min% > 2 THEN GOTO Prelim
'-------------------------------------------------------------
'Abort Squared-Ranks and Kolmogorov-Smirnov calculations where
'  n2<3; (i.e., critical values indistinguishable from zero or
'  unity); assign dummy values to Dmax and Prob.
'-------------------------------------------------------------
T1# = 10.0E6
Sigt$ = "No test"
Sqrank% = %True
Dmax# = 10.0E6
Prob# = 10.0E6
Sig$ = "No test"
Koltes% = %True
'-------------------------------------------------------------
'Calculate Mann-Whitney test for any n1 provided n2>18.
'-------------------------------------------------------------
IF Max% > 18 THEN
  GOTO Squaranks
'-------------------------------------------------------------
'Otherwise abort Mann-Whitney calculations as well.
'-------------------------------------------------------------
ELSE
  PRINT "For ";El$(M%);" -->: Insufficient values to calculate any tests."
  GOTO MoreM
END IF
```

```
'-------------------------------------------------------------
'Preliminary calculations for Squared-Ranks tests.
'-------------------------------------------------------------
Prelim:
FOR Q% = 1 TO K%
'-------------------------------------------------------------
'Calculate mean for group Q in R(Q).
'-------------------------------------------------------------
      R#(Q%) = 0
        FOR J% = Group%(Q%) TO Group%(Q% + 1) - 1
          IF OX#(M%,J%) > Zero# THEN R#(Q%) = R#(Q%) + OX#(M%,J%)
        NEXT J%
      R#(Q%) = R#(Q%) / CDBL (Nvalue%(Q%))
'-------------------------------------------------------------
'Calculate absolute deviations from means for each group combined
'  and place into Calc ready to rank.
'-------------------------------------------------------------
        FOR J% = Group%(Q%) TO Group%(Q% + 1) - 1
          IF OX#(M%,J%) > Zero# THEN
            CALC#(J%) = ABS(OX#(M%,J%) - R#(Q%))
          ELSE
            CALC#(J%) = -10.0E6
          END IF
        NEXT J%
NEXT Q%
'-------------------------------------------------------------
'Rank mean deviations for all data groups combined.
'-------------------------------------------------------------
CALL Rnk (Calc#(), Rank#(), Tsq#, Clearnk%)
'-------------------------------------------------------------
'Increment Q% so that mean of corresponding group is subtracted.
'-------------------------------------------------------------
Q% = 1
FOR J% = 1 TO Ntotal%
  IF J% = Group%(Q% + 1) THEN Q% = Q% + 1
  IF OX#(M%,J%) > Zero# THEN
    TEMP# = OX#(M%,J%) - R#(Q%)
  ELSE
    TEMP# = -10.0E6
  END IF
'-------------------------------------------------------------
'Add mean deviations and ranks of absolute deviations (from the
'  squared-ranks test) to the intermediate file.
'-------------------------------------------------------------
  IF Noprint% <> 0 THEN
    WRITE #2, Temp#; Rank#(J%)
  END IF
NEXT J%
'-------------------------------------------------------------
'Put sum of squared ranks for group Q into R(Q) and return raw
'  data-values to array Calc in preparation for other rank tests.
'-------------------------------------------------------------
Squaranks:
FOR Q% = 1 TO K%
  R#(Q%) = 0
    FOR J% = Group%(Q%) TO Group%(Q% + 1) - 1
      R#(Q%) = R#(Q%) + Rank#(J%)^2
      Calc#(J%) = OX#(M%,J%)
    NEXT J%
```

```
NEXT Q%
'-------------------------------------------------------------------
'If there are ties, put sum of all ranks to 4th power into (R4)
'-------------------------------------------------------------------
IF Tsq# > 0 THEN
 R4# = 0
    FOR Q% = 1 TO Ntotal%
     R4# = R4# + Rank#(Q%)^4
    NEXT Q%
END IF
'-------------------------------------------------------------------
'Preliminary calculations for other ranking tests, rank raw data
'  values for all groups combined.
'-------------------------------------------------------------------
CALL Rnk (Calc#(), Rank#(), T#, Clearnk%)

'-------------------------------------------------------------------
'Add raw values of input array and ranks to the intermediate file.
'-------------------------------------------------------------------
IF Noprint% <> 0 THEN
  FOR J% = 1 to Ntotal%
    WRITE #2, Ox#(M%,J%); Rank#(J%)
  NEXT J%
END IF

IF K% = 2 THEN
'===================================================================
'         CALCULATE THE 3 TESTS APPLICABLE TO 2 GROUPS.
'===================================================================
  N1n2# = CDBL (Nvalue%(1) * Nvalue%(2))
'-------------------------------------------------------------------
'Calculate 2-group, squared-ranks statistic, put into T1#.
'-------------------------------------------------------------------
    IF Min% > 2 THEN
      T1# = R#(1)
      CALL Sq2 (Sigt$, R4#, T1#, R#(2), Tsq#, N1n2#)
'-------------------------------------------------------------------
'Calculate Smirnov test statistic, put in Dmax
'-------------------------------------------------------------------
      CALL Ks (Sig$, Prob#, Zero#, Dmax#, N1n2#, Group%(2), (M%), Nul%)
    END IF
'-------------------------------------------------------------------
'Calculate Mann-Whitney U statistic, put into R#(1), R#(2).
'-------------------------------------------------------------------
  CALL Mw (Zval%, Sigu$, Z#, Rank#(), T#, N1n2#, R#(), Ranksm#, Group%(2))

ELSE

'===================================================================
'         CALCULATE THE 3 TESTS APPLICABLE TO 3 OR MORE GROUPS
'===================================================================
  CALL Sqk (Sigt$, T1#, Tsq#, R4#, R#())
'-------------------------------------------------------------------
'Calculate Kruskal-Wallis statistic, put into H
'  Trip warning messages about chi-square approximation where
'  one or more groups has <6 values.
'-------------------------------------------------------------------
    IF K% > 3 AND Min% < 5 THEN Chiapp% = %True
    IF K% = 3 AND Min% < 5 AND Max% > 5 THEN Chiapp% = %True
```

```
  CALL Kw (Sig$, Krutes%, Rank#(), Ranksm#, H#, T#)
'-----------------------------------------------------------------
'Calculate the Van der Waerden statistic, put into Z
'-----------------------------------------------------------------
  N1n2# = Totval# + 1.0
  S2# = 0
  CALL Vdw (El$(M%), Sigu$, Z#, Rank#())

END IF

'-----------------------------------------------------------------
'Output test results to system monitor and output files.
'-----------------------------------------------------------------
IF K% <= 2 THEN
  CALL Result2scrn (Z#, R#(1), R#(2), Sigu$, Dmax#, Sig$,_
                    Prob#, T1#, Sigt$, Title$, (M%), Ngroup%(),_
                    Koltes%, Nul%, Zval%, Sqrank%)
ELSEIF K% > 2 THEN
  CALL ResultKscrn (Tsq#, T1#, Sigt$, H#, Sig$, Z#, Sigu$, Title$,_
                    (M%), Chiapp%, Nul%, Krutes%)
END IF

IF M% = 1 THEN
  INPUT "Do you want results in an OUTPUT FILE? ENTER (Y/N)-->",Y$
    IF Y$ = "Y" OR Y$ = "y" THEN
      FileOpen% = %TRUE
      Outfile$ = File$ + ".NPO"
      OPEN Outfile$ for OUTPUT as #3
    END IF
END IF

IF FileOpen% = %True AND K% <= 2 THEN
  CALL Result2disk (Z#, R#(1), R#(2), Sigu$, Dmax#, Sig$,_
                  Prob#, T1#, Sigt$, Title$, (M%), Ngroup%(),_
                  Koltes%, Nul%, Zval%, Sqrank%)
ELSEIF FileOpen% = %True AND K% > 2 THEN
  CALL ResultKdisk (Tsq#, T1#, Sigt$, H#, Sig$, Z#, Sigu$, Title$,_
                    (M%), Chiapp%, Nul%, Krutes%)
END IF

MoreM:
  IF ERR = 62 THEN
    BEEP
    LOCATE 10,25: PRINT "Input past end error, CORRECT THE FILE!"
    DELAY 2
  END IF

Clearnk% = 1
CALL Rnk (Calc#(), Rank#(), T#, Clearnk%)
Clearnk% = 0
'-----------------------------------------------------------------
'End of calculations for one input parameter.
'-----------------------------------------------------------------
NEXT M%

IF Noprint% <> 0 THEN CLOSE #2
IF Fileopen% = %True THEN CLOSE #3

PRINT
```

```
INPUT "Do you wish to examine ANOTHER FILE? Enter (Y/N)-->", Y$
IF Y$ = "Y" OR Y$ = "y" THEN FileNameTrap

END
'================================================================
'                    END OF MAIN PROGRAM
'There is no order to subroutine and function listing, pointers
'  are set to properly reposition read routines regardless of
'  position in the program.
'================================================================

'****************************************************************
'Subroutine RESULTKDISK to print results for 3 or more groups to
'  disk file.
'****************************************************************
SUB ResultKdisk (Tsq#, T1#, Sigt$, H#, Sig$, Z#, Sigu$, Title$,_
                 M%, Chiapp%, Nul%, Krutes%)
SHARED El$(), Novar%, Ntotal%, Ntotvl%, K%

Patkw$ = "              H = ####.###                    %Sig_. = \     \"
Patvdw$ = "             T1 = ####.###                    %Sig_. = \     \"
Patsqk$ = "             TSQ = ####    T2 = ####.###     %Sig_. = \     \"
IF M% = 1 THEN
 PRINT #3, "         NPBAS RESULTS FOR 3 OR MORE INDEPENDENT GROUPS OF DATA"
END IF
PRINT #3,
IF M% = 1 THEN PRINT #3, "            FILE TITLE:  "; Title$
PRINT #3, "            VARIABLE NAME: ";El$(M%)
PRINT #3, "            Total Number of Observations in"; K%;"Groups: ";Ntotal%
PRINT #3, "            Number of Observations in"; K%;"groups"
PRINT #3, "                      Ignoring Missing Data: ";Ntotvl%
PRINT #3,
PRINT #3, "            KRUSKAL-WALLIS TEST:"
PRINT #3, USING Patkw$;H#,Sig$
PRINT #3, "            VAN DER WAERDEN TEST:"
PRINT #3, USING Patvdw$;Z#,Sigu$
PRINT #3, "            SQUARED-RANKS TEST:"
PRINT #3, USING Patsqk$;Tsq#,T1#,Sigt$

IF M% = Novar% THEN
  IF Chiapp% = %True THEN
    PRINT #3,
    PRINT #3,"     Where one or more group values <5, chi-square approx-"
    PRINT #3,"       imations doubtful and significance levels uncertain."
  END IF
  IF Krutes% = %True THEN
    PRINT #3,
    PRINT #3,"     Critical H value not available at >80% level for"
    PRINT #3,"       variables with <6 total values; refer to Siegel"
    PRINT #3,"       (1956, Table O)."
  END IF
  IF Nul% = %True THEN
    PRINT #3,
    PRINT #3,"     Too few values for one or more variables to execute"
    PRINT #3,"       any of the tests for >2 groups."
  END IF
END IF

END SUB
```

```
'*****************************************************************
'Subroutine RESULT2DISK to print results for 2 groups to disk file.
'*****************************************************************
SUB Result2disk (Z#, R1#, R2#, Sigu$, Dmax#, Sig$, Prob#, T1#, Sigt$,_
                 Title$, M%, Ngroup%(1), Koltes%, Nul%, Zval%, Sqrank%)
SHARED El$(), Novar%, Ntotal%, Ntotvl%
Patmw$ = "              U =##,###.##  U' =##,###.##       %Sig_. = \     \"
Patks$ = "              Dmax = #.###  Prob_. = ##.###     %Sig_. = \     \"
Patsq2$ = "              T or T1 = ####.###               %Sig_. = \     \"
  IF Z# < 10000 THEN
    Z$ = STR$(Z#)
  ELSE
    Z$ = "Undef."
  END IF
IF M% = 1 THEN_
PRINT #3, "                NPBAS RESULTS FOR 2 INDEPENDENT GROUPS OF DATA"
PRINT #3,
IF M% = 1 THEN PRINT #3, "           FILE TITLE:  "; Title$
PRINT #3, "           VARIABLE NAME: ";El$(M%)
PRINT #3, "           Total Number of Observations:";Ntotal%;
PRINT #3, "in groups of";Ngroup%(1);"and";Ngroup%(2)
PRINT #3, "           Number of Observations Ignoring Missing Data:";Ntotvl%
PRINT #3,
PRINT #3, "           MANN-WHITNEY U TEST:    Z = ";Z$
PRINT #3, USING Patmw$;R1#,R2#,Sigu$
PRINT #3, "           SMIRNOV TEST:"
PRINT #3, USING Patks$;Dmax#,Prob#,Sig$
PRINT #3, "           SQUARED-RANKS TEST:"
PRINT #3, USING Patsq2$;T1#,Sigt$

IF M% = Novar% THEN
  IF Zval% = %True THEN
    PRINT #3,
    PRINT #3,"     Wherever numbers of values in both groups <21, Z in"
    PRINT #3,"       Mann-Whitney test is undefined."
  END IF
  IF Nul% = %True THEN
    PRINT #3,
    PRINT #3,"     In Kolmogorov-Smirnov test, %sig. undefined unless n1>79"
    PRINT #3,"       and n2>99 or n1=n2; approximations then unavailable."
  END IF
  IF Koltes% = %True THEN
    PRINT #3,
    PRINT #3,"     Kolmogorov-Smirnov test inapplicable to some comparisons be-"
    PRINT #3,"     cause critical value is unavailable (small sample numbers)."
  END IF
  IF Sqrank% = %True THEN
    PRINT #3,
    PRINT #3,"     Squared ranks test inapplicable with <3 values in"
    PRINT #3,"       either group of data."
  END IF
END IF
END SUB
'*****************************************************************
'Subroutine RESULTKSCRN to print results for 3 or more groups to
'  the system monitor.
'*****************************************************************
SUB ResultKscrn (Tsq#, T1#, Sigt$, H#, Sig$, Z#, Sigu$, Title$,_
```

```
                M%, Chiapp%, Nul%, Krutes%)
SHARED El$(), Novar%, Ntotal%, Ntotvl%, K%
Patkw$ = "              H = ####.###                    %Sig_. = \     \"
Patvdw$ = "             T1 = ####.###                    %Sig_. = \     \"
Patsqk$ = "             TSQ = ####    T2 = ####.###     %Sig_. = \     \"
CLS
IF M% = 1 THEN
 PRINT "  NPBAS RESULTS FOR 3 OR MORE INDEPENDENT GROUPS OF DATA"
END IF
PRINT
IF M% = 1 THEN PRINT "           FILE TITLE:  "; Title$
PRINT "           VARIABLE NAME: ";El$(M%)
PRINT "           Total Number of Observations in"; K%;"Groups: ";Ntotal%
PRINT "           Number of Observations in"; K%;"groups"
PRINT "                     Ignoring Missing Data: ";Ntotvl%
PRINT
PRINT "           KRUSKAL-WALLIS TEST:"
PRINT USING Patkw$;H#,Sig$
PRINT "           VAN DER WAERDEN TEST:"
PRINT USING Patvdw$;Z#,Sigu$
PRINT "           SQUARED-RANKS TEST:"
PRINT USING Patsqk$;Tsq#,T1#,Sigt$
PRINT

IF M% = Novar% THEN
  INPUT "Do you want MESSAGES shown ON SCREEN? ENTER (Y/N)-->",Y$
  IF Y$ = "Y" or Y$ = "y" THEN
    LOCATE 16,1
    PRINT STRING$(53," ")
    LOCATE 15,1
    IF Chiapp% = %True THEN
      PRINT
      PRINT "Where one or more group values <5, chi-square approx-"
      PRINT "  imations doubtful and significance levels uncertain."
    END IF
    IF Krutes% = %True THEN
      PRINT
      PRINT "Critical H value not available at >80% level for"
      PRINT "  variables with <6 total values; refer to Siegel"
      PRINT "  (1956, Table O)."
    END IF
    IF Nul% = %True THEN
      PRINT
      PRINT "Too few values for one or more variables to execute"
      PRINT "  any of the tests for >2 groups."
    END IF
  PRINT
  END IF
END IF
END SUB
'*****************************************************************
'Subroutine RESULT2SCRN to print results for 2 groups to the
'  system monitor.
'*****************************************************************
SUB Result2scrn (Z#, R1#, R2#, Sigu$, Dmax#, Sig$, Prob#, T1#, Sigt$,_
                 Title$, M%, Ngroup%(1), Koltes%, Nul%, Zval%, Sqrank%)
SHARED El$(), Novar%, Ntotal%, Ntotvl%
Patmw$ = "              U =##,###.##  U' =##,###.##       %Sig_. = \     \"
Patks$ = "              Dmax = #.###  Prob_. = ##.###     %Sig_. = \     \"
```

```
Patsq2$ = "              T or T1 = ####.###                %Sig_. = \     \"
  IF Z# < 10000 THEN
    Z$ = STR$(Z#)
  ELSE
    Z$ = "Undef."
  END IF
CLS
IF M% = 1 THEN
  PRINT "              NPBAS RESULTS FOR 2 INDEPENDENT GROUPS OF DATA"
END IF
PRINT
IF M% = 1 THEN PRINT "           FILE TITLE:  "; Title$
PRINT "           VARIABLE NAME: ";El$(M%)
PRINT "           Total Number of Observations:";Ntotal%;
PRINT "in groups of";Ngroup%(1);"and";Ngroup%(2)
PRINT "           Number of Observations Ignoring Missing Data:";Ntotvl%
PRINT
PRINT "           MANN-WHITNEY U TEST:    Z = ";Z$
PRINT USING Patmw$;R1#,R2#,Sigu$
PRINT "           SMIRNOV TEST:"
PRINT USING Patks$;Dmax#,Prob#,Sig$
PRINT "           SQUARED-RANKS TEST:"
PRINT USING Patsq2$;T1#,Sigt$
PRINT
IF M% = Novar% THEN
  INPUT "Do you want MESSAGES shown ON SCREEN? ENTER (Y/N)-->",Y$
  IF Y$ = "Y" or Y$ = "y" THEN
    LOCATE 15,1
    PRINT STRING$(53," ")
    LOCATE 14,1
    IF Zval% = %True THEN
      PRINT
      PRINT "Wherever numbers of values in both groups <21, Z in"
      PRINT "  Mann-Whitney test is undefined."
    END IF
    IF Nul% = %True THEN
      PRINT
      PRINT "In Kolmogorov-Smirnov test, %sig. undefined unless n1>79"
      PRINT "  and n2>99 or n1=n2; approximations then unavailable."
    END IF
    IF Koltes% = %True THEN
      PRINT
      PRINT "Kolmogorov-Smirnov test inapplicable to some comparisons be-"
      PRINT "  cause critical value is unavailable (small sample numbers)."
    END IF
    IF Sqrank% = %True THEN
      PRINT
      PRINT "Squared ranks test inapplicable with <3 values in"
      PRINT "  either group of data."
    END IF
  PRINT
  END IF
END IF
END SUB
'*****************************************************************
'Subroutine VDW to calculate Van der Waerden test statistic and
'  its significance level.
'*****************************************************************
SUB Vdw (El$, Sigu$, T1#, Rank#(1))
```

```
SHARED N1n2#, K%, Group%(), Nvalue%(), Ntotal%, Totval#,_
       Noprint%
LOCAL Aibar#(), Dprob#, S2#
DIM Aibar#(15)
'-----------------------------------------------------------------
'Write variable name to the intermediate file.
'-----------------------------------------------------------------
IF Noprint% <> 0 THEN WRITE #2, El$
'-----------------------------------------------------------------
'Convert each rank into a standard normal score and sum squares to
'  get S2 (Equation 3, Conover, 1980, p. 318), using NAG routine
'  G01CEF.
'-----------------------------------------------------------------
FOR Q% = 1 TO Ntotal%
  IF Rank#(Q%) = 0 THEN GOTO Moreq
  Dprob# = Rank#(Q%) / N1n2#
  T1# = Rank#(Q%)

    IF Dprob# <= 0 OR Dprob# >= 1.0 THEN
      PRINT "CALCULATIONS ABORTED--Probability < 0 or > 1 encountered"
      PRINT "  in Van der Waerden test"
      PRINT
      INPUT "Press <ENTER> to continue"; continue$
      STOP
    END IF
  Ifail% = 0
'-----------------------------------------------------------------
'Call Function PPND77 to calculate "normal scores" from normal
'  deviate corresponding to lower tail area of Dprob.
'-----------------------------------------------------------------
  Rank#(Q%) = FN Ppnd77# (Dprob#, Ifail%)
'-----------------------------------------------------------------
'Add rank, quantile, normal score, and Ifail and Q paramenters to
'  the intermediate file.
'-----------------------------------------------------------------
  IF Noprint% <> 0 THEN
    WRITE #2, T1#; Dprob#; Rank#(Q%); Ifail%; Q%
  END IF

  S2# = S2# + Rank#(Q%)^2
Moreq:
NEXT Q%

S2# = S2# / (Totval# - 1.0)
T1# = 0
'-----------------------------------------------------------------
'Sum Ai parameter for each group, put into Aibar(Q%) and
'  calculate the mean.
'-----------------------------------------------------------------
FOR Q% = 1 TO K%
  Aibar#(Q%) = 0
  FOR J% = Group%(Q%) TO Group%(Q% + 1) - 1
    Aibar#(Q%) = Aibar#(Q%) + Rank#(J%)
  NEXT J%
  Aibar#(Q%) = Aibar#(Q%) / CDBL(Nvalue%(Q%))
'-----------------------------------------------------------------
'Total all Ai parameters, put into T1 (Equation 4 of Conover,
'  1980, p. 318).
'-----------------------------------------------------------------
```

```
  T1# = T1# + CDBL(Nvalue%(Q%)) * (Aibar#(Q%)^2)
NEXT Q%
T1# = T1# / S2#
'-------------------------------------------------------------------
'Calculate significance level of T1 from Chi-square approximation.
'-------------------------------------------------------------------
CALL Chisig (T1#, K%, Sigu$)
END SUB
'*******************************************************************
'Subroutine SQK to calculate Squared Ranks statistic for K
'  (>2 groups) and its significance level.
'*******************************************************************

SUB Sqk (Sigt$, T1#, Tsq#, R4#, R#(1))
SHARED K%, Nvalue%(), Totval#
LOCAL Temp#, S2#
'-------------------------------------------------------------------
'Calculate sum of squares of mean square ranks (equation 7 of
'  Conover, 1980, p.241).
'-------------------------------------------------------------------
T1# = 0
FOR Q% = 1 TO K%
  T1# = T1# + R#(Q%)^2 / CDBL(Nvalue%(Q%))
NEXT Q%
IF Tsq# = 0 THEN
'-------------------------------------------------------------------
'Calculate test statistic using equations 8 & 9 of Conover (1980,
'  p.242), where no ties occur.  Temp# holds (N + 1)(2N + 1) / 6
'  in Conover's symbols.
'-------------------------------------------------------------------
  Temp# = (Totval# + 1.0) * (2.0 * Totval# + 1.0) / 6.0
  T1# = T1# - Totval# * Temp#^2
  T1# = T1# * 30.0 / (Temp# * Totval# * (8.0 * Totval# + 11.0))
ELSE
'-------------------------------------------------------------------
'Calculate test statistic from equation 7 where ties are present.
'  Calculate average of squared ranks for all groups, put into S2.
'-------------------------------------------------------------------
  S2# = 0
  FOR Q% = 1 TO K%
    S2# = S2# + R#(Q%)
  NEXT Q%
  S2# = S2# / Totval#
  Temp# = Totval# * (S2#^2)
  T1# = (T1# - Temp#) * (Totval# - 1.0) / (R4# - Temp#)
END IF
CALL Chisig (T1#, K%, Sigt$)
END SUB
'*******************************************************************
'Subroutine SQ2 to calculate Squared ranks test for 2 groups
'  and its significance level.
'*******************************************************************
SUB Sq2 (Sigt$, R4#, T1#, T2#, Tsq#, N1n2#)
SHARED  Nvalue%(), Totval#, Max%
LOCAL Q%, J%, T1cr99%(),T1cr90%(),T1cr80%(),T1cr20%(),T1cr10%(),T1cr01()
DIM  Stan#(5), Crval#(10), T1cr99%(3:10,3:10), T1cr90%(3:10,3:10),_
T1cr80%(3:10,3:10), T1cr20%(3:10,3:10), T1cr10%(3:10,3:10),_
T1cr01%(3:10,3:10)
RESTORE Sq2data
```

```
FOR Q% = 3 TO 10
  FOR J% = 3 TO 10
    READ T1cr99%(Q%,J%)
  NEXT J%
NEXT Q%
FOR Q% = 3 TO 10
  FOR J% = 3 TO 10
    READ T1cr90%(Q%,J%)
  NEXT J%
NEXT Q%
FOR Q% = 3 TO 10
  FOR J% = 3 TO 10
    READ T1cr80%(Q%,J%)
  NEXT J%
NEXT Q%
FOR Q% = 3 TO 10
  FOR J% = 3 TO 10
    READ T1cr20%(Q%,J%)
  NEXT J%
NEXT Q%
FOR Q% = 3 TO 10
  FOR J% = 3 TO 10
    READ T1cr10%(Q%,J%)
  NEXT J%
NEXT Q%
FOR Q% = 3 TO 10
  FOR J% = 3 TO 10
    READ T1cr01%(Q%,J%)
  NEXT J%
NEXT Q%
FOR Q% = 1 TO 5
  READ Stan#(Q%)
NEXT Q%
'-----------------------------------------------------------------
'Calculate test statistic in no-ties situation. For large samples
'  (n,m > 10) use approximation p.455 of Conover (1980).
'  Temp = n(N+1)(2N+1)/6 in Conover's symbols.
'-----------------------------------------------------------------

IF Tsq# = 0 THEN
  IF Max% > 10 THEN
    Temp1# = CDBL(Nvalue%(1)) * (Totval# + 1.0) * (2 * Totval# + 1.0) / 6.0
    Temp2# = SQR(CDBL(Nvalue%(2)) * (8.0 * Totval# + 11.0) / 30.0)
'-----------------------------------------------------------------
'Use quantiles of standard normal population to test significance
'  and approximation, bottom of p.455 of Conover (1980).
'-----------------------------------------------------------------

    FOR Q% = 1 TO 5
      Temp2# = Stan#(Q%) + Temp2#
      Crval#(Q%) = Temp1# + Temp2#
      Crval#(Q% + 5) = Temp1# - Temp2#
    NEXT Q%
    CALL Sig4 (T1#, Sigt$, Crval#())
'-----------------------------------------------------------------
'For m and n <10 use exact values in Table A9 of Conover (1980).
'-----------------------------------------------------------------
  ELSE
    IF (T1# > CDBL(T1cr99%(Nvalue%(2),Nvalue%(1))) OR_
```

```
       T1# < CDBL(T1cr01%(Nvalue%(2), Nvalue%(1)))) THEN
          Sigt$ = ">99%  "
       ELSEIF (T1# > CDBL(T1cr90%(Nvalue%(2),Nvalue%(1))) OR_
            T1# < CDBL(T1cr10%(Nvalue%(2),Nvalue%(1)))) THEN
          Sigt$ = ">90%  "
       ELSEIF (T1# > CDBL(T1cr80%(Nvalue%(2),Nvalue%(1))) OR_
             T1# < CDBL(T1cr20%(Nvalue%(2),Nvalue%(1)))) THEN
          Sigt$ = ">80%  "
       ELSE
          Sigt$ = "<80%  "
    END IF
  END IF
ELSE
'-----------------------------------------------------------------
'If there are ties, use equation 4, p.240 of Conover (1980);
'  Temp = nm/N-1 in Conover's symbols.
'-----------------------------------------------------------------
  Temp1# = N1n2# / (Totval# - 1.0)
  Temp2# = (T1# + T2#) / Totval#
  T1# = (T1# - CDBL(Nvalue%(1)) * Temp2#)
  T1# = T1# / SQR(Temp1# * (R4#/Totval# - (Temp2#)^2))
'-----------------------------------------------------------------
'Then determine significance from two-tailed quantiles of standard
'  normal variables; reject at 90% if T1 <95% quantiles, etc.
'-----------------------------------------------------------------

  FOR Q% = 1 TO 5
    Crval#(Q%) = Stan#(Q%)
    Crval#(Q% + 5) = -Stan#(Q%)
  NEXT Q%

  CALL Sig4 (T1#, Sigt$, Crval#())
END IF

Sq2data:
'=============Data for T1cr99%
DATA 77,110,149,194,245,302,346,413,126,174,230,281,351,414,494,567,_
     190,255,319,391,478,559,654,754,271,346,431,526,624,731,847,970,_
     371,467,571,683,803,929,1067,1212,492,604,731,863,1005,1156,1319,1489,_
     629,769,916,1073,1239,1417,1601,1798,798,961,1130,1314,1505,1708,1921,_
     2145
'=============Data for T1cr90%
DATA 70,101,129,161,197,238,285,333,119,154,197,246,294,350,413,476,_
     178,228,282,342,410,479,558,639,255,319,386,463,545,634,730,831,_
     347,428,515,608,707,814,929,1051,464,560,664,776,896,1023,1159,1303,_
     601,717,840,972,1112,1261,1420,1587,765,901,1045,1197,1360,1533,_
     1715,1907
'=============Data for T1cr80%
DATA 65,90,117,149,182,221,260,305,111,142,182,222,270,321,375,435,_
     169,214,264,319,379,445,514,591,243,300,364,435,511,592,679,772,_
     335,407,487,572,665,764,871,984,447,536,632,735,846,965,1091,1224,_
     581,689,803,925,1056,1195,1343,1498,742,866,1001,1144,1296,1457,_
     1627,1806
'=============Data for T1cr20%
DATA 26,29,35,42,50,59,69,77,50,62,71,85,99,114,130,149,_
     87,103,121,142,163,187,212,239,136,163,187,215,247,280,315,352,_
     203,236,271,308,350,394,440,489,285,329,374,423,476,531,590,652,_
     390,444,501,561,625,694,766,843,514,580,649,724,801,885,972,1064
'=============Data for T1cr10%
```

```
DATA 21,21,26,30,38,42,49,54,39,50,57,66,78,90,102,114,_
     75,88,103,120,135,155,175,195,124,139,164,187,211,239,268,299,_
     188,212,240,274,308,344,384,425,268,300,340,381,426,473,524,576,_
     365,406,457,510,567,626,689,755,486,539,601,665,734,806,883,963
'=============Data for T1cr01%
DATA 14,14,14,14,14,14,21,21,30,30,30,39,39,46,50,54,_
     55,55,66,75,79,88,99,110,91,104,115,124,136,152,167,182,_
     140,155,172,195,212,235,257,280,204,236,260,284,311,340,368,401,_
     304,325,361,393,429,466,508,549,406,448,486,526,573,620,672,725
'=============Data for Stan#
DATA 3.2905,2.5758,1.9600,1.6449,1.2816

END SUB
'*****************************************************************
'Subroutine SIG4 to compare squared-ranks test statistic with 99.9,
'  99, 95, and 80% values of standard normal random variable
'  (note for significance level A, test statistic T or T1 must
'  exceed A/2 quantile or be less than 1-A/2 quantile)
'*****************************************************************
SUB Sig4 (Testat#, Sig$, Crval#(1))

IF (Testat# > Crval#(1) OR Testat# < Crval#(6)) THEN
  Sig$ = ">99.9%"
ELSEIF (Testat# >  Crval#(2) OR Testat# < Crval#(7)) THEN
  Sig$ = ">99%  "
ELSEIF (Testat# > Crval#(3) OR Testat# < Crval#(8)) THEN
  Sig$ = ">95%  "
ELSEIF (Testat# > Crval#(4) OR Testat# < Crval#(9)) THEN
  Sig$ = ">90%  "
ELSEIF (Testat# > Crval#(5) OR Testat# < Crval#(10)) THEN
  Sig$ = ">80%  "
ELSE
  Sig$ = "<80%  "
END IF

END SUB
'*****************************************************************
'Subroutine KS to calculate Smirnov (2-group Kolmogorov-Smirnov)
'  statistic and its significance level.
'*****************************************************************
SUB Ks (Sig$, Prob#, Zero#, Dmax#, N1n2#, Group2%, M%, Nul%)
SHARED Ntotal%, Nvalue%(), Min%, Max%, El$(), Ox#(),_
       Totval#, Noprint%
LOCAL P%, Q%, J%, R#()
DIM Cum#(2,15), Dcr99%(40), Dcr95%(40), Dcr90%(40),_
    Dcr80%(40), R#(15)
RESTORE Ksdata

FOR P% = 1 TO 40: READ Dcr99%(P%): NEXT P%
FOR P% = 1 TO 40: READ Dcr95%(P%): NEXT P%
FOR P% = 1 TO 40: READ Dcr90%(P%): NEXT P%
FOR P% = 1 TO 40: READ Dcr80%(P%): NEXT P%
Ksdata:
'============Data for Dcr99
DATA 0,0,0,4,4,5,5,6,6,7,7,7,8,8,8,9,9,9,9,10,10,10,10,11,11,11,11,_
     12,12,12,12,12,13,13,13,13,13,14,14,14
'============Data for Dcr95
DATA 0,0,0,3,4,4,5,5,5,6,6,6,6,7,7,7,7,8,8,8,8,8,9,9,9,9,9,_
     10,10,10,10,10,11,11,11,11,11,11,11,12
```

```
'============Data for Dcr90
DATA 0,0,2,3,3,4,4,4,5,5,5,5,6,6,6,6,7,7,7,7,7,8,8,8,8,8,8,_
     9,9,9,9,9,9,10,10,10,10,10,10,10
'============Data for Dcr80
DATA 0,0,2,3,3,3,4,4,4,4,5,5,5,5,5,6,6,6,6,6,6,7,7,7,7,7,7,_
     8,8,8,8,8,8,8,8,9,9,9,9,9
'-----------------------------------------------------------------
'Initialize counting arrays.
'-----------------------------------------------------------------
FOR Q% = 1 TO 2
  FOR J% = 1 TO 15
    CUM#(Q%,J%) = 0
  NEXT J%
NEXT Q%
'-----------------------------------------------------------------
'Determine minimum and maximum data-values for both groups to-
'  gether and put into R(1) and R(15) .
'-----------------------------------------------------------------
R#(1) = 10.0E6
R#(15) = 0
FOR J% = 1 TO Ntotal%
  IF Ox#(M%,J%) > R#(15) THEN R#(15) = Ox#(M%,J%)
  IF Ox#(M%,J%) > Zero# AND Ox#(M%,J%) < R#(1) THEN R#(1) = Ox#(M%,J%)
NEXT J%
'-----------------------------------------------------------------
'Determine class interval to divide range into 15 classes.
'-----------------------------------------------------------------
Temp# = (R#(15) - R#(1)) / 15.0
'-----------------------------------------------------------------
'Determine class boundaries for calculating cumulative step func-
'  tions (top class to include maximum, bottom to include minimum
'  value).
'-----------------------------------------------------------------
R#(1) = R#(1) + Temp#
FOR J% = 2 TO 14
  R#(J%) = R#(1) + CDBL(J% - 1) * Temp#
NEXT J%
'-----------------------------------------------------------------
'Determine cumulative step functions for the 2 groups,
'  put numbers of real values in Cum.
'-----------------------------------------------------------------
FOR Q% = 1 TO 15
  FOR J% = 1 TO Group2% - 1
    IF Ox#(M%,J%) > Zero# AND Ox#(M%,J%) <= R#(Q%) THEN_
       Cum#(1,Q%) = Cum#(1,Q%) + 1.0
  NEXT J%
  FOR J% = Group2% TO Ntotal%
    IF Ox#(M%,J%) > Zero# AND Ox#(M%,J%) <= R#(Q%) THEN_
       Cum#(2,Q%) = Cum#(2,Q%) + 1.0
  NEXT J%
NEXT Q%
'-----------------------------------------------------------------
'Determine step functions as proportions of total real values
'  and determine maximum deviation between the two step functions.
'-----------------------------------------------------------------
Dmax# = 0
FOR J% = 1 TO 15
  Cum#(1,J%) = Cum#(1,J%)/CDBL(Nvalue%(1))
  Cum#(2,J%) = Cum#(2,J%)/CDBL(Nvalue%(2))
```

```
  Temp# = ABS(Cum#(1,J%) - Cum#(2,J%))
  IF Temp# > Dmax# THEN Dmax# = Temp#
NEXT J%
'-------------------------------------------------------------------
'Add raw data values and cumulative frequencey distributions to
'  the intermediate file.
'-------------------------------------------------------------------
IF Noprint% <> 0 THEN
  WRITE #2, El$(M%)
  FOR Q% = 1 TO 2
    FOR J% = 1 TO 15
      WRITE #2, R#(J%), CUM#(Q%,J%)
    NEXT J%
  NEXT Q%
END IF
'-------------------------------------------------------------------
'Determine exact probability of Dmax value using function AKSCDF.
'-------------------------------------------------------------------
Prob# = FN Akscdf# (Min%, Max%, Dmax#)
'-------------------------------------------------------------------
'Use equal-sample approximations if numbers of values in both
'  groups are equal.
'-------------------------------------------------------------------
IF Nvalue%(1) = Nvalue%(2) THEN
  IF Nvalue%(1) < 41 THEN
'-------------------------------------------------------------------
'For n1=n2<41, compare Dmax with critical values from Steel
'  and Torrie (1980, table A23A).
'-------------------------------------------------------------------

    Temp# = Dmax# * CDBL(Nvalue%(1))
    CALL Sig2 (Temp#, Sig$, CDBL(Dcr99%(Nvalue%(1))),_
              CDBL(Dcr95%(Nvalue%(1))), CDBL(Dcr90%(Nvalue%(1))),_
              CDBL(Dcr80%(Nvalue%(1))))
  ELSE
'-------------------------------------------------------------------
'For n1=n2>40, use large-sample approximation (Table A23A of
'  Steel and Torrie, 1980).
'-------------------------------------------------------------------
    Temp# = SQR(CDBL(Nvalue%(1)))
    CALL Sig2 (Dmax#, Sig$, 2.3018/Temp#, 1.9206/Temp#, 1.7308/Temp#,_
              1.5174/Temp#)
  END IF
'-------------------------------------------------------------------
'For n2>99>n1>80: determine significance level using Smirnov
'  approximation with critical values X*SQRT(n1+n2)/n1n2 from
'  Siegel (1956, Table M), Steel and Torrie (1980, Table A23B),
'  Beyer (1968, p.429), etc. as recommended by Kim and Jennrich
'  (1970, p. 84).
'-------------------------------------------------------------------
ELSEIF (Min% > 79 AND Max% > 99) THEN
  Temp# = SQR(Totval#/N1n2#)
  CALL Sig1 (Dmax#, Sig$, 1.95*Temp#, 1.73*Temp#, 1.6276*Temp#,_
            1.5174*Temp#, 1.48*Temp#, 1.3581*Temp#, 1.2239*Temp#,_
            1.073*Temp#)
'-------------------------------------------------------------------
'For other n1, n2, leave significance level approximation undefined.
'-------------------------------------------------------------------
ELSE
```

```
  NUL% = %True
  SIG$ = "Undef. "
END IF

END SUB
'*****************************************************************
'Function AKSCDF, to calculate exact probability of a given
'  difference between 2 cumulative frequency distributions, in the
'  Smirnov test (algorithm adapted to GEC4090 from p.88 of Kim
'  and Jennrich, 1970); adapted to TURBOBASIC from FORTRAN-77
'  after Rock (1986).
'*****************************************************************
DEF FN Akscdf# (M%, N%, D#)
LOCAL U#(), K%, W#, Q%, J%
DIM U#(500)
'-----------------------------------------------------------------
'Array U must have dimensions of at least Totval+1 in main program.
'  NOTE: Next statement altered from that in Kim and Jennrich,
'  (1970) to produce same result on GEC 4090 computer; may require
'  adjustment on other systems.
'-----------------------------------------------------------------
K% = CINT(CDBL(M%*N%)*D#) - 1
U#(1) = 1.0
FOR J% = 1 TO N%
  U#(J% + 1) = 1.0
  IF (M%*J% > K%) THEN U#(J% + 1) = 0
NEXT J%

FOR Q% = 1 TO M%
  W# = CDBL(Q%)/CDBL(Q% + N%)
  U#(1) = W#*U#(1)
  IF (N%*Q% > K%) THEN U#(1) = 0
    FOR J% = 1 TO N%
      U#(J% + 1) = U#(J%) + U#(J% + 1) * W#
      IF ABS(N%*Q% - M%*J%) > K% THEN U#(J% + 1) = 0
    NEXT J%
NEXT Q%

FN Akscdf# = 100.0 * U#(N% + 1)
END DEF
'*****************************************************************
'Subroutine MW to calculate Mann-Whitney U statistic and
'  significance level.
'*****************************************************************
SUB Mw (Zval%, Sigu$, Z#, Rank#(1), T#, N1n2#, U#(1), Ranksm#, Group2%)
SHARED Ntotal%, Nvalue%(), Totval#, Max%, Min%
LOCAL Q%, J%
DIM Ucr995%(21,20), Ucr908%(21,20)
RESTORE Mwdata

FOR Q% = 1 TO 20
  FOR J% = 1 TO 21
    READ Ucr995%(J%,Q%)
  NEXT J%
NEXT Q%
FOR Q% = 1 TO 20
  FOR J% = 1 TO 21
    READ Ucr908%(J%,Q%)
  NEXT J%
```

```
NEXT Q%

Mwdata:
'============Data for Ucr995%
DATA 0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,_
     0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,1,1,2,2,2,2,_
     3,3,0,0,0,0,0,0,0,0,1,1,2,2,3,3,4,5,5,6,6,7,8,0,0,0,1,2,0,_
     1,2,3,3,4,5,6,7,7,8,9,10,11,12,13,0,0,1,2,3,5,2,3,4,5,6,7,9,_
     10,11,12,13,15,16,17,18,0,0,1,3,5,6,8,4,6,7,9,10,12,13,15,16,_
     18,19,21,22,24,0,0,2,4,6,8,10,13,8,9,11,13,15,17,18,20,22,24,_
     26,28,30,0,0,2,4,7,10,12,15,17,11,13,16,18,20,22,24,27,29,31,_
     33,36,0,0,3,5,8,11,14,17,20,23,16,18,21,24,26,29,31,34,37,39,_
     42,0,0,3,6,9,13,16,19,23,26,30,21,24,27,30,33,36,39,_
     42,45,48,0,1,4,7,11,14,18,22,26,29,33,37,27,31,34,37,41,44,47,_
     51,54,0,1,4,8,12,16,20,24,28,33,37,41,45,34,38,42,45,49,53,56,_
     60,0,1,5,9,13,17,22,26,31,36,40,45,50,55,42,46,50,54,58,63,67,_
     0,1,5,10,14,19,24,29,34,39,44,49,54,59,64,51,55,60,64,69,73,_
     0,1,6,11,15,21,26,31,37,42,47,53,59,64,70,75,60,65,70,74,79,0,2,_
     6,11,17,22,28,34,39,45,51,57,63,67,75,81,87,70,75,81,86,_
     0,2,7,12,18,24,30,36,42,48,55,61,67,74,80,86,93,99,81,87,92,_
     0,2,7,13,19,25,32,38,45,52,58,65,72,78,85,92,99,106,113,93,99,_
     0,2,8,13,20,27,34,41,48,55,62,69,76,83,90,98,105,112,119,127,105

'============ Data for Ucr908%
DATA  0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,0,1,_
      1,1,1,2,2,2,3,3,3,4,4,4,0,0,1,0,0,1,2,2,3,3,4,5,5,6,7,7,8,9,_
      9,10,11,0,0,1,3,1,2,3,4,5,6,7,8,9,10,11,12,14,15,16,17,18,0,_
      1,2,4,5,4,5,6,8,9,11,12,13,15,16,18,19,20,22,23,25,0,1,3,5,7,_
      9,7,8,10,12,14,16,17,19,21,23,25,26,28,30,32,0,1,4,6,8,11,13,11,13,_
      15,17,19,21,24,26,28,30,33,35,37,39,0,2,5,7,10,13,16,19,15,18,20,23,26,_
      28,31,33,36,39,41,44,47,0,2,5,9,12,15,18,22,25,21,24,27,_
      30,33,36,39,42,45,48,51,54,0,3,6,10,13,17,21,24,28,32,27,_
      31,34,37,41,44,48,51,55,58,62,0,3,7,11,15,19,23,27,31,36,_
      40,34,38,42,46,50,54,57,61,65,69,0,4,8,12,17,21,26,30,35,_
      39,44,49,42,47,51,55,60,64,68,72,77,0,4,9,13,18,23,28,33,_
      38,43,48,53,58,51,56,61,65,70,75,80,84,0,4,10,15,20,25,31,_
      36,41,47,52,58,63,69,61,66,71,77,82,87,92,0,5,10,16,22,27,_
      33,39,45,51,57,63,68,74,80,72,77,83,88,94,100,0,5,11,17,23,_
      29,36,42,48,54,61,67,74,80,86,93,83,89,95,101,107,0,6,12,18,_
      25,31,38,45,52,58,65,72,79,85,92,99,106,96,102,109,115,0,6,13,_
      20,27,34,41,48,55,62,69,77,84,91,98,106,113,120,109,116,123,1,7,_
      14,21,28,36,43,51,58,66,73,81,89,97,104,112,120,128,135,123,130,1,_
      7,15,22,30,38,46,54,62,70,78,86,94,102,110,119,127,135,143,151,138

Ranksm# = 0: U#(1) = 0: U#(2) = 0
'-----------------------------------------------------------------
'Sum ranks in the two groups.
'-----------------------------------------------------------------
FOR J% = 1 TO Group2% - 1
  U#(1) = U#(1) + Rank#(J%)
NEXT J%

FOR J% = Group2% TO Ntotal%
  U#(2) = U#(2) + Rank#(J%)
NEXT J%
'-----------------------------------------------------------------
'Calculate U and U'.
'-----------------------------------------------------------------
FOR Q% = 1 TO 2
```

```
  Ranksm# = Ranksm# + U#(Q%)
  U#(Q%) = N1n2# + CDBL(Nvalue%(Q%) * (Nvalue%(Q%) + 1)) / 2.0 - U#(Q%)
NEXT Q%
'-------------------------------------------------------------------
'Calculate tie adjustment from sums of tie functions.
'-------------------------------------------------------------------

T# = T# / 12.0
IF Max% > 20 THEN
'-------------------------------------------------------------------
'Determine signigicance level of U for N>20 by comparison with Z
'  unit normal approximation.
'-------------------------------------------------------------------
  Z# = ABS(U#(1) - (N1n2#/2.0))
    IF T# = 0 THEN
      Z# = Z# / SQR(N1n2# * (Totval# + 1.0) / 12.0)
    ELSE
      Z# = Z# / SQR(N1n2# * ((Totval#^3 - Totval#) / 12.0 - T#)_
           / (Totval# * (Totval# - 1.0)))
    END IF
  CALL Sig1 (Z#, Sigu$, 3.30, 2.879, 2.5758, 2.324, 2.24, 1.96,_
            1.6449, 1.2816)

ELSE
'-------------------------------------------------------------------
'Where max. number of values <21, compare minimum R(U)  value
'  with critical U values from Beyer (1968, p. 406ff). NOTE: nega-
'  tive values are used in CALL statements since subroutine uses >
'  comparisons whereas critical relationship is <.
'-------------------------------------------------------------------
  Zval% = %True
  Z# = 10000.0
  Temp# = -U#(1)

  IF U#(2) < U#(1) THEN Temp# = -U#(2)
  CALL Sig2 (Temp#, Sigu$, -CDBL(Ucr995%(Max% + 1,Min%)),_
              -CDBL(Ucr995%(Min%,Max%)), -CDBL(Ucr908%(Max% + 1,Min%)),_
              -CDBL(Ucr908%(Min%,Max%)))
END IF

END SUB
'*******************************************************************
'Subroutine SIG1 to compare large-sample Smirnov and Mann-Whitney
'  Z approximations with 99.9, 99.5, 99, 98, 97.5, 95, 90, and
'  80% values.
'*******************************************************************
Sub Sig1 (Testat#, Sig$, V999#, V995#, V99#, V98#, V975#, V95#,_
        V90#, V80#)
IF Testat# > V999# THEN
  Sig$ = ">99.9% "
ELSEIF Testat# > V995# THEN
  Sig$ = ">99.5% "
ELSEIF Testat# > V99# THEN
  Sig$ = ">99.0% "
ELSEIF Testat# > V98# THEN
  Sig$ = ">98.0% "
ELSEIF Testat# > V975# THEN
  Sig$ = ">97.5% "
ELSEIF Testat# > V95# THEN
```

```
  Sig$ = ">95.0% "
ELSEIF Testat# > V90# THEN
  Sig$ = ">90.0% "
ELSEIF Testat# > V80# THEN
  Sig$ = ">80.0% "
ELSE
  Sig$ = "<80.0% "
END IF
END SUB
'*****************************************************************
'Subroutine SIG2 to compare equi-sample Smirnov and small-sample
'  Mann-Whitney statistics with 99, 95, 90, and 80% critical values.
'*****************************************************************
SUB Sig2 (Testat#, Sig$, V99#, V95#, V90#, V80#)
IF Testat# > V99# THEN
  Sig$ = ">99.0% "
ELSEIF Testat# > V95# THEN
  Sig$ = ">95.0% "
ELSEIF Testat# > V90# THEN
  Sig$ = ">90.0% "
ELSEIF Testat# > V80# THEN
  Sig$ = ">80.0% "
ELSE
  Sig$ = "<80.0% "
END IF
END SUB
'*****************************************************************
'Subroutine KW to calculate Kruskal-Wallis statistic and
'  significance level.
'*****************************************************************
SUB Kw (Sig$, Krutes%, Rank#(1), Ranksm#, H#, T#)
SHARED Group%(), Ntotvl%, Max%, Min%, K%, Nvalue%(), Totval#
LOCAL Q%, J%, R#(), Hcr5#(), Hcr4#(), Hcr3#()
DIM R#(15), Hcr5#(6,5), Hcr4#(5,4), Hcr3#(4,3)
RESTORE Kwdata

FOR Q% = 1 TO 6: FOR J% = 1 TO 5: READ Hcr5#(6,5): NEXT J%: NEXT Q%
FOR Q% = 1 TO 5: FOR J% = 1 TO 4: READ Hcr4#(5,4): NEXT J%: NEXT Q%
FOR Q% = 1 TO 4: FOR J% = 1 TO 3: READ Hcr3#(4,3): NEXT J%: NEXT Q%
Kwdata:
'--------Data for Hcr5
DATA 0,0,4.45,4.8711,4.86,4.9091,4.05,4.2933,5.04,5.1055,_
     5.2682,5.2462,3.84,4.4946,4.4121,5.5152,5.6308,5.6264,_
     3.96,4.5182,4.5231,4.6181,5.6176,5.6429,4.0364,4.5077,_
     4.5363,4.52,4.5,5.66
'--------Data for Hcr4
DATA 0,0,4.8214,5.0,4.8667,4.0179,4.1667,5.1250,5.4,5.2364,_
     3.8889,4.4444,4.7,5.7273,5.5758,4.0667,4.4455,4.773,4.5,5.6538
'--------Data for Hcr3
DATA 0,0,4.2857,4.5714,3.8571,4.4643,4.5,5.1389,4.0,4.25,_
     4.6,5.0667
'-----------------------------------------------------------------
'Sum ranks for each group in turn, put into R, i.e., from 1 to end
'  of group 1, 1+end of group 1 to end of group 2, etc.
'-----------------------------------------------------------------
H# = 0
FOR Q% = 1 TO K%
  R#(Q%) = 0
    FOR J% = Group%(Q%) TO Group%(Q% + 1) - 1
```

```
      R#(Q%) = R#(Q%) + Rank#(J%)
    NEXT J%
  Ranksm# = Ranksm# + R#(Q%)
  R#(Q%) = R#(Q%)^2 / CDBL(Nvalue%(Q%))
  H# = H# + R#(Q%)
NEXT Q%
'-----------------------------------------------------------------
'Calculate value of H statistic (uncorrected for ties).
'-----------------------------------------------------------------
H# = H# * 12.0 / (Totval# * (Totval# + 1.0)) - 3.0 * (Totval# + 1.0)
'-----------------------------------------------------------------
'Adjust H value for ties.
'-----------------------------------------------------------------
IF T# > 0 THEN
  T# = 1.0 - T# / (Totval# * (Totval#^2 - 1.0))
  H# = H# / T#
END IF
'-----------------------------------------------------------------
'Find intermediate number of values J (Max >= J >= Min).
'-----------------------------------------------------------------
J% = 0
J% = Ntotvl% - Max% - Min%
'-----------------------------------------------------------------
'For >3 groups, or > 5 samples in one or more groups, calculate
'  significance level of H statistic from chi-square approximation.
'-----------------------------------------------------------------
IF K% > 3 OR Max% > 5 THEN
  CALL Chisig (H#, K%, Sig$)
'-----------------------------------------------------------------
'Ignore cases with <6 total values, or with J (and hence Min)
'  = 1, as no significance values >80% available.
'-----------------------------------------------------------------
ELSEIF Ntotvl% < 6 OR J% = 1 THEN
  Sig$ = "Undef. "
  Krutes% = %True
'-----------------------------------------------------------------
'For 3 groups, < 6 values in each group, compare calculated H
'  value with exact probabilities in Seigel (1956, Table O).
'  Determine critical level with 2 values in largest group.
'-----------------------------------------------------------------
ELSE
  IF Max% = 2 THEN CALL Sig3 (H#, Sig$, 4.5714, 3.7143)
'-----------------------------------------------------------------
'Determine critical level with 3 values in largest group.
'-----------------------------------------------------------------
  IF Max% = 3 THEN CALL Sig3 (H#, Sig$, Hcr3#(J% + 1,Min%), Hcr3#(Min%,J%))
'-----------------------------------------------------------------
'Determine critical level with 4 values in largest group.
'-----------------------------------------------------------------
  IF Max% = 4 THEN CALL Sig3 (H#, Sig$, Hcr4#(J% + 1,Min%), Hcr4#(Min%,J%))
'-----------------------------------------------------------------
'Determine critical level with 5 values in largest group.
'-----------------------------------------------------------------
  IF Max% = 5 THEN CALL Sig3 (H#, Sig$, Hcr5#(J% + 1,Min%), Hcr5#(Min%,J%))
END IF

END SUB
'*****************************************************************
'Subroutine SIG3 to compare small-sample, 3-group Kruskal-Wallis
```

```
'  statistic with 95 and 90% critical values.
'*********************************************************************
SUB Sig3 (Testat#, Sig$, V95#, V90#)
IF Testat# > V95# THEN
  Sig$ = ">95%  "
ELSEIF Testat# > V90% THEN
  Sig$ = ">90%  "
ELSE
  Sig$ = "<90% "
END IF

END SUB
'*********************************************************************
'Subroutine CHISIG determines approximate significance levels of
'  Kruskal-Wallis, Van der Waerden and Squared Ranks test statis-
'  tics, by comparing with chi-square for (K-1) degrees of freedom,
'  where K is the number of groups of data.
'*********************************************************************
SUB Chisig (Testat#, K%, Sig$)
DIM Chi999#(15), Chi99#(15), Chi95#(15), Chi90#(15), Chi80#(15)
RESTORE Chisigdata

FOR Q% = 1 TO 15: READ Chi999#(Q%): NEXT Q%
FOR Q% = 1 TO 15: READ Chi99#(Q%): NEXT Q%
FOR Q% = 1 TO 15: READ Chi95#(Q%): NEXT Q%
FOR Q% = 1 TO 15: READ Chi90#(Q%): NEXT Q%
FOR Q% = 1 TO 15: READ Chi80#(Q%): NEXT Q%
Chisigdata:
     DATA 0.0,0.0,13.82,16.27,18.46,20.52,22.46,24.32,26.12,27.88,_
          29.59,31.26,32.91,34.52,36.12
     DATA 0.0,0.0,9.21,11.34,13.28,15.09,16.81,18.48,20.09,21.67,_
          23.21,24.72,26.22,27.69,29.14
     DATA 0.0,0.0,5.99,7.82,9.49,11.07,12.59,14.07,15.51,16.91,_
          18.31,19.68,21.03,22.36,23.68
     DATA 0.0,0.0,4.61,6.25,7.78,9.24,10.64,12.02,13.36,14.68,_
          15.99,17.28,18.55,19.81,21.06
     DATA 0.0,0.0,3.22,4.64,5.99,7.29,8.56,9.80,11.03,12.24,_
          13.44,14.63,15.81,16.98,18.15
'---------------------------------------------------------------
'Chi999 to Chi80 carry values of Chi-squared for particular numbers
'  of groups at significance levels from 99.9 to 80% (from table C of
'  Siegel, 1956). Note: Chi(K) corresponds to K-1 degrees of freedom.
'---------------------------------------------------------------
IF Testat# > Chi999#(K%) THEN
  Sig$ = ">99.9% "
ELSEIF Testat# > Chi99#(K%) THEN
  Sig$ = ">99.0% "
ELSEIF Testat# > Chi95#(K%) THEN
  Sig$ = ">95.0% "
ELSEIF Testat# > Chi90#(K%) THEN
  Sig$ = ">90.0% "
ELSEIF Testat# > Chi80#(K%) THEN
  Sig$ = ">80.0% "
ELSE
  Sig$ = "<80.0% "
END IF
END SUB
'*********************************************************************
'Function PPND77 to calculate "normal scores" in Van der Waerden
```

```
'  test.  Algorithm as ASlll of Beasley and Springer (1977),
'  adapted to TURBOBASIC from FORTRAN-77 after Rock (1986).
'*****************************************************************
DEF FN Ppnd77# (P#, Ifail%)
LOCAL Q#, R#
  Zero# = 0.0: Half# = 0.5: One# = 1.0: Split# = 0.42
'-----------------------------------------------------------------
'HASH SUMS; the sums of the moduli of the coefficients (included
'  by Rock, 1986, p. 777; for use in checking transcriptions)
'  are not included in this program.
'-----------------------------------------------------------------
  A0# = 2.50662823884: A1# = -18.61500062529: A2# = 41.39773534
  A3# = -25.44106049637: B1# = -8.4735109309: B2# = 23.08336743743
  B3# = -21.06224101826: B4# = 3.13082909833: C0# = -2.78718931138
  C1# = -2.29796479134: C2# = 4.85014127135: C3# = 2.32121276858
  D1# = 3.54388924762: D2# = 1.63706781897
'-----------------------------------------------------------------
'Function produces normal deviate corresponding to lower tail area
'  of probability value P (all input in DOUBLE PRECISION).  Refer
'  to Rock (1986) for notes on FORTRAN-77 standard functions and
'  system dependent statements for this Function.
'-----------------------------------------------------------------
IF P# = 0.5 THEN
  FN Ppnd77# = 0.0
ELSE
  Ifail% = 0
  Q# = P# - Half#
'-----------------------------------------------------------------
'">", in original algorithm becomes "<" for FORTRAN-77 version of
'  the next statement, carried on in this version, in BASIC.
'-----------------------------------------------------------------
  IF ABS(Q#) < Split# THEN
    R# = Q# * Q#
    FN Ppnd77# = (Q# * (((A3# * R# + A2#) * R# + A1#) * R# + A0#)_
                 / ((((B4# * R# + B3#) * R# + B2#) * R# + B1#) * R# + One#))
  ELSE
    R# = P#
    IF Q# > Zero# THEN R# = one# - P#
    IF R# <= Zero# THEN
      PRINT "WARNING:  Failure in FN Ppnd77 during VDW test."
      Ifail% = 1
'-----------------------------------------------------------------
'Return dummy value in PPND77 to output.
'-----------------------------------------------------------------
      FN Ppnd77# = 10.0E-6
    ELSE
      R# = SQR(-LOG(R#))
      FN Ppnd77# = ((((C3# * R# + C2#) * R# + C1#)_
                 * R# + C0#) / ((D2# * R# + D1#) * R# + One#))
      IF Q# < Zero# THEN FN Ppnd77# = -((((C3# * R# + C2#) * R# + C1#)_
                   * R# + C0#) / ((D2# * R# + D1#) * R# + One#))
    END IF
  END IF
END IF
END DEF
'*****************************************************************
'Subroutine RNK to rank values while leaving in original order.
'*****************************************************************
SUB Rnk (Calc#(1), Rank#(1), T#, Clearnk%)
```

```
SHARED Zero#, Ntotal%, Ntotvl%
Local Ties#, Ntie%(), Valmin#, Q%, J%
DIM Ntie%(200)

IF Clearnk% <> 0 THEN
  FOR J% = 1 TO Ntotal%
    Rank#(J%) = 0
  NEXT J%
  EXIT SUB
END IF

T# = 0: Ties# = 0
For Q% = 1 TO Ntotvl%
  Ties# = Ties# - 1.0
  IF Ties# > 0 THEN GOTO SomeQ
'-----------------------------------------------------------------
'Assign current minimum value to Valmin.
'-----------------------------------------------------------------
  Valmin# = 99999
    FOR J% = 1 TO Ntotal%
      IF Calc#(J%) > Zero# AND Calc#(J%) < Valmin# THEN
         Valmin# = Calc#(J%)
      END IF
    NEXT J%
'-----------------------------------------------------------------
'Assign next lowest rank to current minimum value, count number of
'  ties and convert already ranked data-values into missing values
'  so that they will subsequently be ignored.
'-----------------------------------------------------------------
  Ties# = 0
    FOR J% = 1 TO Ntotal%
      IF Calc#(J%) = Valmin# THEN
         Rank#(J%) = CDBL(Q%)
         Ties# = Ties# + 1.0
'-----------------------------------------------------------------
'Identify current minimum value by number J in array Calc(J) and
'  allow for declared size of array Ntie.
'-----------------------------------------------------------------
         IF Ties# > 200 THEN
           PRINT "Number of ties > 200; the declared size of array"
           PRINT "Ntie%: the program must be recompiled."
           STOP
         END IF
         Ntie%(CINT(Ties#)) = J%
         Calc#(J%) = -10.0E6
      END IF
    NEXT J%

  IF Ties# < 2 THEN GOTO SomeQ
'-----------------------------------------------------------------
'Increase rank values where ties occur and sum T (t^cubed - t).
'-----------------------------------------------------------------
  T# = T# + ((Ties#)^3 - Ties#)
    FOR J% = 1 TO (CINT(Ties#))
      Rank#(Ntie%(J%)) = Rank#(Ntie%(J%)) + 0.5 * (Ties# - 1.0)
    NEXT J%

SomeQ:
NEXT Q%
END SUB
```