PB86-203783

User-Friendly IBM PC (Personal Computer) Computer Programs for Solving Sampling and Statistical Problems

(U.S.) Environmental Monitoring and Support Lab.
Cincinnati, OH

U.S. DEPARTMENT OF COMMERCE
National Technical Information Service
NTIS

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
OFFICE OF RESEARCH AND DEVELOPMENT
ENVIRONMENTAL MONITORING AND SUPPORT LABORATORY
CINCINNATI, OHIO 45268

Gentlemen:

Enclosed is a copy of a diskette of the "User-Friendly IBM PC Computer Programs for Solving Sampling and Statistical Problems" as requested.

The program's menu will automatically appear on the screen when the diskette is inserted and the computer is turned on. The programs on the diskette can also be copied to the hard disk. In this case, the user just types "EMSLSTAT" on the C drive to run the program's menu.

If there are any suggestions on the programs, please do not hesitate to let me know.

Your personal comments would also be appreciated.

Sincerely yours,

Philip C. L. Lin

Philip C. L. Lin, PH.D.
Mechanical Engineer
Sampling and Field Measurements Section
Physical and Chemical Methods Branch

EPA/600/4-86/023
May 1986

USER-FRIENDLY IBM PC COMPUTER PROGRAMS FOR SOLVING SAMPLING AND STATISTICAL PROBLEMS

PB86-203783

BY

PHILIP C. L. LIN

ENVIRONMENTAL MONITORING AND SUPPORT LABORATORY

OFFICE OF RESEARCH AND DEVELOPMENT

U. S. ENVIRONMENTAL PROTECTION AGENCY

CINCINNATI, OHIO 45268

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before completing)

1. REPORT NO. EPA/600/4-86/023	2	3 RECIPIENT'S ACCESSION NO
4. TITLE AND SUBTITLE User-Friendly IBM PC Computer Programs for Solving Sampling and Statistical Problems		5 REPORT DATE May 1986
		6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Philip C. L. Lin		8. PERFORMING ORGANIZATION REPORT NO
9. PERFORMING ORGANIZATION NAME AND ADDRESS Sampling and Field Measurements Section, Physical and Chemical Methods Branch, Environmental Monitoring and Support Laboratory, USEPA, Cincinnati, Ohio 45268		10. PROGRAM ELEMENT NO.
		11. CONTRACT/GRANT NO.
12. SPONSORING AGENCY NAME AND ADDRESS Environmental Monitoring and Support Laboratory, Office of Research and Development, U. S. Environmental Protection Agency, Cincinnati, Ohio 45268		13. TYPE OF REPORT AND PERIOD COVERED
		14. SPONSORING AGENCY CODE EPA 600/6
15. SUPPLEMENTARY NOTES		

16. ABSTRACT

User friendly IBM personal computer programs for solving sampling and related statistical problems have been prepared. The programs are designed so that persons without an in-depth understanding of statistics can easily use them. Specific, detailed, written instructions for application of the programs are provided in the report. The computer disc containing the programs will be made available on request to the Environmental Monitoring and Support Laboratory - Cincinnati (EMSL-Cincinnati).

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

18. DISTRIBUTION STATEMENT Distribute to Public	19. SECURITY CLASS (This Report) Unclassified	21 NO OF PAGES 74
	20 SECURITY CLASS (This page) Unclassified	22. PRICE

EPA Form 2220-1 (Rev. 4-77) PREVIOUS EDITION IS OBSOLETE

DISCLAIMER

This report has been reviewed by the Environmental Monitoring and Support Laboratory - Cincinnati, U.S. Environmental Protection Agency, and approved for publication. Mention of trade names or commercial products does not constitute endorsement of recommendation for use.

FOREWORD

Environmental measurements are required to determine the quality of ambient waters and the character of waste effluents. The Environmental Monitoring and Support Laboratory - Cincinnati conducts research to:

Develop and evaluate techniques to measure the presence and concentration of physical, chemical, and radiological pollutants in water, wastewater, bottom sediments, and solid waste.

Investigate methods for the concentration, recovery, and identification of viruses, bacteria, and other microbiological organisms in water, and determine the responses of aquatic organisms to water quality.

Develop and operate an Agency-wide quality assurance program to assure standardization and quality control of systems for monitoring water and wastewater.

The function of the Sampling and Field Measurement Section of the Physical and Chemical Methods Branch in the Environmental Monitoring and Support Laboratory is to provide field measurement and sampling techniques relating to water quality sampling programs. This report provides user-friendly IBM PC computer programs for solving sampling and statistical problems so that an individual may use the programs and obtain the benefits of

the statistical package without an in-depth understanding of statistics employed. Descriptions of basic statistics are also presented for those who wish to know more of the details of the statistics.

Robert L. Booth

Director

Environmental Monitoring and Support Laboratory - Cincinnati

ABSTRACT

User friendly IBM personal computer programs for solving sampling and related statistical problems have been prepared. The programs are designed so that persons without an in-depth understanding of statistics can easily use them. Specific, detailed, written instructions for application of the programs are provided in the report. The computer disc containing the programs will be made available on request to the Environmental Monitoring and Support Laboratory - Cincinnati (EMSL-Cincinnati).

CONTENTS

	Page
Foreword	iii
Abstract	v
Figures	viii
1. Introduction	1
2. Instructions for Using Sampling Programs on the IBM PC	4
3. Examples of Sampling Programs	6
Appendix A. Definitions of Basic Statistics	A-1
Appendix B. Descriptions of Statistical Sampling Program on the Disk	B-1
B.1 Curve Fitting With a Linear Regression	B-1
B.2 Normal Deviate Z	B-2
B.3 Percentage Area Under the Normal Curve	B-3
B.4 Student t	B-5
B.5 Percentage Area Under the Student t	B-5
B.6 Chi Square	B-6
B.7 Sample Mean, Standard Deviation, and Confidence Intervals for the Mean and Variance	B-6
B.8 Determination of the Number of Samples	B-8
B.9 Probability of Exceeding a Standard	B-9
B.10 Hypothesis Testing	B-10
B.11 Power Spectrum Analysis	B-11
B.12 Comparing Two Means	B-18

CONTENTS (Cont'd.)

Page

B.13 Percentage Area Under the F distribution B-19

B.14 F Distribution B-19

B.15 Significant Test between Variabilities of Two Samples. B-19

B.16 Significant Test between the Population Variability and the Sample Variability B-21

Appendix C. Nomeclature. C-1

FIGURES

Number		Page
A-1	Normal distribution	A-2
A-2	Distribution of student t with β=4 degrees of freedom	A-7
A-3	Chi square distribution	A-7
B-1	Standard normal distribution	B-4
B-2	Time record of TOC of municipal wastewater at Racine, Wisconsin	B-17
B-3	Power spectrum of TOC concentration of municipal wastewater at Racine, Wisconsin	B-17

SECTION 1

INTRODUCTION

Statistical techniques are useful in assessing the quality of a sampling program. Frequently, field persons engaged in sample collection do not have the time to thoroughly study and understand all the statistics required to take a representative sample. The computer programs described herein were developed for those people and are designed so that an individual may use the programs and obtain the benefits of the statistical package without an in-depth understanding of the statistics employed. A disc containing the programs will be provided by the Environmental Monitoring and Support Laboratory - Cincinnati (EMSL-Cincinnati) upon request.

For those persons who wish to know more of the details of the statistical package, descriptions are presented in the Appendices. Those who wish to proceed directly to the computer portion will find the programs in Sections 2 and 3. The programs are user-friendly to those familiar with the IBM PC.

Typical Examples for Use of the Programs

In order to assist the user in working the computer programs, a series of questions and answers have been developed. Questions that those designing field sampling programs may wish to have answered are listed below, together with the names of the computer programs designed to answer the questions:

Question - How many samples must be taken to reduce the anticipated error to some reasonably fixed value?

Answer - Use program No. 8 "Determination of Sample Number" if the reduction of the anticipated error is based on the accuracy of the sample variance. Use program No. 9 "Determination of Sample Number" if the reduction of the anticipated error is based on the accuracy of the mean.

Question - What is the probability of an effluent exceeding a standard?

Answer - Use program No. 10 "Probability of Exceeding the Standard."

Question - How does one test whether a sample belongs in a particular distribution?

Answer - Use program No. 11, "Hypothesis Testing."

Question - What is the sampling frequency required to capture a significant event in a long-term monitoring program?

Answer - Use program No. 12, "Power Spectrum Analysis."

Question - How does one determine the sample mean, standard deviation, and confidence intervals for the mean and variance?

Answer - Use program No. 7, "Sample Mean, Standard Deviation, and Confidence Intervals for Population Mean and Variance."

Question - Which program should one use to correlate observed data in a linear manner?

Answer - Use program No. 1, "Linear Regression" to determine the linear relationship and its correlation coefficient.

Question - A material is treated by two different processes. Would there be any justification for saying there was a difference between the two processes? Which program should one use to answer this question?

Answer - Use program No. 13, "Comparing Two Means."

Question - New equipment is used to measure a compound and it is expected that the measurement uniformity would improve. The question to ask is

whether the improvement (more uniformity) really exists or has that occurred by chance. Which program should one use to test for the significant difference between variances of two samples?

Answer - Use program No. 16, "Test for Significant Difference between Variabilities of Two Samples."

SECTION 2

INSTRUCTIONS FOR USING SAMPLING PROGRAMS ON THE IBM PC

Some individuals, especially those that have had extensive computer experience, will be at ease in a few minutes with these programs. In those cases the instructions may be bypassed, and the reader may begin to run the programs immediately. For those who need additional assistance, the following instructions are provided to assist the reader to "boot up" the programs and make logical selections.

Instructions to load and use the disk

1. Place the program disk in Disk Drive A and close the door.
2. Turn on the power of each instrument beginning with the printer, monitor, and, finally, the computer. After a brief warm-up, you will see the program menu:

```
*****************************************
*     PROGRAM MENU ........PAGE 1       *
*****************************************

 1. LINEAR REGRESSION
 2. CALCULATION OF NORMAL DEVIATE Z
 3. CALCULATION OF THE PERCENTAGE AREA OF NORMAL DISTRIBU-
    TION (FROM MINUS INFINITY TO NORMAL DEVIATE Z)
 4. CALCULATION OF STUDENT T
 5. CALCULATION OF THE PERCENTAGE AREA OF STUDENT T DIST-
    RIBUTION (FROM MINUS INFINITY TO STUDENT T)
 6. CALCULATION OF CHI SQUARE
 7. CALCULATION OF SAMPLE MEAN,STANDARD DEVIATION, AND CON-
    FIDENCE INTERVALS FOR THE POPULATION MEAN AND VARIANCE
 8. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
    THE VARIANCE
 9. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
    THE SAMPLE MEAN
10. CALCULATION OF THE PROBABILITY OF EXCEEDING A STANDARD
11. HYPOTHESIS TESTING
12. POWER SPECTRUM ANALYSIS
13. PROCEED TO NEXT PAGE
14. QUIT
TYPE THE DESIRED OPTION NUMBER AND PRESS ENTER ?
```

3. Type an option number after the question mark (?) and press ENTER. The desired program will be loaded into the computer.
4. After you run the desired program, you have several choices:
   (a) go back to program menu,
   (b) do another calculation,
   (c) quit,

   by typing the requested option number and press ENTER.
5. If you want to abort program calculation, press CONTROL-BREAK key. If you want to start over again, type "A:EMSLSTAT" and press ENTER.

SECTION 3

EXAMPLES OF SAMPLING PROGRAMS

```
*****************************************
*      PROGRAM MENU ........PAGE 1       *
*****************************************

 1. LINEAR REGRESSION
 2. CALCULATION OF NORMAL DEVIATE Z
 3. CALCULATION OF THE PERCENTAGE AREA OF NORMAL DISTRIBU-
    TION (FROM MINUS INFINITY TO NORMAL DEVIATE Z)
 4. CALCULATION OF STUDENT T
 5. CALCULATION OF THE PERCENTAGE AREA OF STUDENT T DIST-
    RIBUTION (FROM MINUS INFINITY TO STUDENT T)
 6. CALCULATION OF CHI SQUARE
 7. CALCULATION OF SAMPLE MEAN,STANDARD DEVIATION, AND CON-
    FIDENCE INTERVALS FOR THE POPULATION MEAN AND VARIANCE
 8. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
    THE VARIANCE
 9. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
    THE SAMPLE MEAN
10. CALCULATION OF THE PROBABILITY OF EXCEEDING A STANDARD
11. HYPOTHESIS TESTING
12. POWER SPECTRUM ANALYSIS
13. PROCEED TO NEXT PAGE
14. QUIT
TYPE THE DESIRED OPTION NUMBER AND PRESS ENTER ? 1


************************
* 1. LINEAR REGRESSION *
************************


THIS PROGRAM ESTIMATES A LINE, Y=A+BX, WHERE X IS THE INDEPEN-
DENT VARIABLE AND Y IS THE DEPENDENT VARIABLE.

ANSWER EACH QUESTION AFTER A QUESTION MARK (?) AND THEN PRESS ENTER.

DEFINE X(INDEPENDENT VARIABLE SUCH AS BOD, ETC)=? BOD
DEFINE Y(DEPENDENT VARIABLE SUCH AS TOC, ETC)=? TOC
YOUR DATA STORED IN A FILE MUST BE IN X(INDEPENDENT VAR)
AND Y(DEPENDENT VAR) FORMAT (FOR EXAMPLE, 30.1,100.3).
AN EXAMPLE FILE TEST1.DAT IS ON THIS DISK WHICH YOU CAN USE FOR
A TEST RUN.
IS YOUR DATA STORED IN A FILE (Y/N) ? Y
INPUT FILENAME(NO MORE THAN 8 CHARACTERS)
DATA FROM DISK A, TYPE A:  DATA FROM HARD DISK, TYPE C: FIRST
AND THEN TYPE FILENAME.


DO YOU WISH TO LIST THE FILENAME BEFORE YOU PROCEED(Y/N) ? N
 TYPE FILENAME? A:TEST1.DAT
```

```
SELECT NUMBER OF OPTION:
          1. LIST INPUT DATA
          2. MODIFY OR ADD INPUT DATA
          3. DELETE SOME OF THE DATA
          4. PERFORM REGRESSION ANALYSIS
          5. STORE DATA
          6. GO TO PROGRAM MENU
          7. DO ANOTHER REGRESSION
OPTION ? 1
LISTING OF DATA
DATA POINT        X              Y
 1               10             115
 2               31             249
 3               17             208
 4               42             374
 5               36             307
 6               33             299

SELECT NUMBER OF OPTION:
          1. LIST INPUT DATA
          2. MODIFY OR ADD INPUT DATA
          3. DELETE SOME OF THE DATA
          4. PERFORM REGRESSION ANALYSIS
          5. STORE DATA
          6. GO TO PROGRAM MENU
          7. DO ANOTHER REGRESSION
OPTION ? 4

REGRESSION EQUATION:
Y= 55.95309 + 7.196932 X

COEFFICIENT OF CORRELATION= .9712766

ACTUAL VERSUS ESTIMATED VALUES
X=BOD  Y=TOC
X               Y               ESTIMATED Y    ERROR
 10              115             127.9224      -12.92239
 31              249             279.058       -30.05798
 17              208             178.3009       29.69908
 42              374             358.2243       15.77576
 36              307             315.0426      -8.042633
 33              299             293.4519       5.548157

SELECT NUMBER OF OPTION:
          1. LIST INPUT DATA
          2. MODIFY OR ADD INPUT DATA
          3. DELETE SOME OF THE DATA
          4. PERFORM REGRESSION ANALYSIS
          5. STORE DATA
          6. GO TO PROGRAM MENU
          7. DO ANOTHER REGRESSION
OPTION ? 6
```

```
*****************************************
*     PROGRAM MENU ........PAGE 1       *
*****************************************

 1. LINEAR REGRESSION
 2. CALCULATION OF NORMAL DEVIATE Z
 3. CALCULATION OF THE PERCENTAGE AREA OF NORMAL DISTRIBU-
     TION (FROM MINUS INFINITY TO NORMAL DEVIATE Z)
 4. CALCULATION OF STUDENT T
 5. CALCULATION OF THE PERCENTAGE AREA OF STUDENT T DIST-
     RIBUTION (FROM MINUS INFINITY TO STUDENT T)
 6. CALCULATION OF CHI SQUARE
 7. CALCULATION OF SAMPLE MEAN,STANDARD DEVIATION, AND CON-
     FIDENCE INTERVALS FOR THE POPULATION MEAN AND VARIANCE
 8. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
     THE VARIANCE
 9. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
     THE SAMPLE MEAN
10. CALCULATION OF THE PROBABILITY OF EXCEEDING A STANDARD
11. HYPOTHESIS TESTING
12. POWER SPECTRUM ANALYSIS
13. PROCEED TO NEXT PAGE
14. QUIT
TYPE THE DESIRED OPTION NUMBER AND PRESS ENTER ? 2


                    ************************
                    * 2. NORMAL DEVIATE Z *
                    ************************


Z IS THE DISTANCE FROM THE POPULATION MEAN IN UNITS OF THE STANDARD
DEVIATION IN A NORMAL DISTRIBUTION CURVE. THE CREATION OF THE CONFIDENCE
INTERVAL FOR THE MEAN AT A CERTAIN CONFIDENCE LEVEL REQUIRES THE VALUE OF Z.
TO USE THIS PROGRAM TO CALCULATE THE Z VALUE REQUIRES THE USER TO PROVIDE
THE CONFIDENCE LEVEL (TWO-TAILED TEST). INPUT A VALUE LESS THAN 99.997 %.

ANSWER EACH QUESTION AFTER A QUESTION MARK (?) AND THEN PRESS ENTER.

INPUT CONFIDENCE LEVEL %  ? 95


*********************************************
     CONFIDENCE LEVEL =              95  %
     THE NORMAL DEVIATE Z =          1.959961
*********************************************


DO YOU WISH TO DO ANOTHER CALCULATION (Y/N)? N
```

```
*****************************************
*     PROGRAM MENU ........PAGE 1       *
*****************************************

 1. LINEAR REGRESSION
 2. CALCULATION OF NORMAL DEVIATE Z
 3. CALCULATION OF THE PERCENTAGE AREA OF NORMAL DISTRIBU-
     TION (FROM MINUS INFINITY TO NORMAL DEVIATE Z)
 4. CALCULATION OF STUDENT T
 5. CALCULATION OF THE PERCENTAGE AREA OF STUDENT T DIST-
     RIBUTION (FROM MINUS INFINITY TO STUDENT T)
 6. CALCULATION OF CHI SQUARE
 7. CALCULATION OF SAMPLE MEAN,STANDARD DEVIATION, AND CON-
     FIDENCE INTERVALS FOR THE POPULATION MEAN AND VARIANCE
 8. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
     THE VARIANCE
 9. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
     THE SAMPLE MEAN
10. CALCULATION OF THE PROBABILITY OF EXCEEDING A STANDARD
11. HYPOTHESIS TESTING
12. POWER SPECTRUM ANALYSIS
13. PROCEED TO NEXT PAGE
14. QUIT
TYPE THE DESIRED OPTION NUMBER AND PRESS ENTER ? 3


*****************************************************************
* 3. CALCULATION OF THE PERCENTAGE AREA % OF NORMAL DISTRIBUTION *
*****************************************************************


THIS IS A NORMAL DISTRIBUTION PROGRAM TO CALCULATE THE PROBABILITY
INTEGRATED FROM MINUS INFINITY TO A NORMAL DEVIATE Z.
THE USER HAS TO INPUT A VALUE OF NORMAL DEVIATE Z.
DO NOT EXCEED A Z VALUE OF 4.12 WHICH GENERATES AN AREA OF 99.99901 %

ANSWER EACH QUESTION AFTER A QUESTION MARK (?) AND THEN PRESS ENTER.
IF YOU WANT TO START OVER AGAIN, PRESS FUNCTION KEY 1 AND RETURN.

INPUT A Z VALUE =? 1.96

*****************************************************
AREA OF NORMAL DISTRIBUTION (FROM Z=MINUS INFINITY TO
Z= 1.96 ) = 97.50023  %
*****************************************************


 DO YOU WISH TO DO ANOTHER CALCULATION (Y/N)? N
```

```
*****************************************
*     PROGRAM MENU ........PAGE 1       *
*****************************************

 1. LINEAR REGRESSION
 2. CALCULATION OF NORMAL DEVIATE Z
 3. CALCULATION OF THE PERCENTAGE AREA OF NORMAL DISTRIBU-
     TION (FROM MINUS INFINITY TO NORMAL DEVIATE Z)
 4. CALCULATION OF STUDENT T
 5. CALCULATION OF THE PERCENTAGE AREA OF STUDENT T DIST-
     RIBUTION (FROM MINUS INFINITY TO STUDENT T)
 6. CALCULATION OF CHI SQUARE
 7. CALCULATION OF SAMPLE MEAN,STANDARD DEVIATION, AND CON-
     FIDENCE INTERVALS FOR THE POPULATION MEAN AND VARIANCE
 8. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
     THE VARIANCE
 9. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
     THE SAMPLE MEAN
10. CALCULATION OF THE PROBABILITY OF EXCEEDING A STANDARD
11. HYPOTHESIS TESTING
12. POWER SPECTRUM ANALYSIS
13. PROCEED TO NEXT PAGE
14. QUIT
TYPE THE DESIRED OPTION NUMBER AND PRESS ENTER ? 4


          *********************************
          * 4.     STUDENT T               *
          *********************************


IF THE VARIABILITY OF A NORMAL DISTRIBUTION IS ESTIMATED FROM
A SET OF SAMPLES, A STUDENT T DISTRIBUTION INSTEAD OF THE NORMAL
DISTRIBUTION IS USED TO CREATE THE CONFIDENCE INTERVAL FOR THE MEAN.
THE STUDENT T IS TO BE CALCULATED FROM THE PROVIDED CONFIDENCE
LEVEL FOR THE POPULATION MEAN AND THE DEGREES OF FREEDOM WHICH
ARE ONE LESS THEN THE NUMBER OF SAMPLES. THE CONFIDENCE LEVEL
PROVIDED BY USER HERE IS FOR TWO-TAILED TEST.

ANSWER EACH QUESTION AFTER A QUESTION MARK (?) AND THEN PRESS ENTER.
PRESS FUNCTION KEY 1 AND THEN PRESS ENTER TO START OVER AGAIN.

 INPUT CONFIDENCE LEVEL? 95
 INPUT DEGREES OF FREEDOM =? 12


*********************************************************
DEGREES OF FREEDOM                          = 12
CONFIDENCE LEVEL FOR THE POPULATION MEAN= 95  %
THE STUDENT T =                              2.178711
*********************************************************
```

```
DO YOU WISH TO DO ANOTHER CALCULATION (Y/N)? N

        ****************************************
        *     PROGRAM MENU ........PAGE 1      *
        ****************************************

 1. LINEAR REGRESSION
 2. CALCULATION OF NORMAL DEVIATE Z
 3. CALCULATION OF THE PERCENTAGE AREA OF NORMAL DISTRIBU-
     TION (FROM MINUS INFINITY TO NORMAL DEVIATE Z)
 4. CALCULATION OF STUDENT T
 5. CALCULATION OF THE PERCENTAGE AREA OF STUDENT T DIST-
     RIBUTION (FROM MINUS INFINITY TO STUDENT T)
 6. CALCULATION OF CHI SQUARE
 7. CALCULATION OF SAMPLE MEAN,STANDARD DEVIATION, AND CON-
     FIDENCE INTERVALS FOR THE POPULATION MEAN AND VARIANCE
 8. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
     THE VARIANCE
 9. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
     THE SAMPLE MEAN
10. CALCULATION OF THE PROBABILITY OF EXCEEDING A STANDARD
11. HYPOTHESIS TESTING
12. POWER SPECTRUM ANALYSIS
13. PROCEED TO NEXT PAGE
14. QUIT
TYPE THE DESIRED OPTION NUMBER AND PRESS ENTER ? 5


*********************************************************************
* 5. CALCULATION OF THE PERCENTAGE AREA % OF STUDENT T DISTRIBUTION *
*********************************************************************


THIS PROGRAM IS TO CALCULATE THE PERCENTAGE AREA OF STUDENT
T DISTRIBUTION.  THE AREA IS INTEGRATED FROM MINUS INFINITY TO THE
T VALUE WHICH HAS TO BE PROVIDED BY THE USER. THE DEGREES OF
FREEDOM ARE ALSO NEEDED.

ANSWER EACH QUESTION AFTER A QUESTION MARK (?) AND THEN PRESS ENTER.

IF YOU WANT TO START OVER AGAIN, PRESS FUNCTION KEY 1 AND RETURN.
 INPUT DEGREES OF FREEDOM=? 12
 INPUT STUDENT T VALUE =? 2


*****************************************
 DEGREES OF FREEDOM =              12
 THE T VALUE        =              2
 THE PERCENTAGE AREA FOR
 THE STUDENT T DISTRIBUTION =      96.56724  %
*****************************************


DO YOU WISH TO DO ANOTHER CALCULATION (Y/N)? N
```

```
*****************************************
*      PROGRAM MENU ........PAGE 1       *
*****************************************

 1. LINEAR REGRESSION
 2. CALCULATION OF NORMAL DEVIATE Z
 3. CALCULATION OF THE PERCENTAGE AREA OF NORMAL DISTRIBU-
     TION (FROM MINUS INFINITY TO NORMAL DEVIATE Z)
 4. CALCULATION OF STUDENT T
 5. CALCULATION OF THE PERCENTAGE AREA OF STUDENT T DIST-
     RIBUTION (FROM MINUS INFINITY TO STUDENT T)
 6. CALCULATION OF CHI SQUARE
 7. CALCULATION OF SAMPLE MEAN,STANDARD DEVIATION, AND CON-
     FIDENCE INTERVALS FOR THE POPULATION MEAN AND VARIANCE
 8. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
     THE VARIANCE
 9. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
     THE SAMPLE MEAN
10. CALCULATION OF THE PROBABILITY OF EXCEEDING A STANDARD
11. HYPOTHESIS TESTING
12. POWER SPECTRUM ANALYSIS
13. PROCEED TO NEXT PAGE
14. QUIT
TYPE THE DESIRED OPTION NUMBER AND PRESS ENTER ? 6


                 *************************
                 * 6. CHI-SQUARE PROGRAM *
                 *************************


NORMALLY DISTRIBUTED DATA COULD BE TRANSFORMED INTO A UNIT
NORMAL DISTRIBUTION WITH MEAN=0 AND VARIANCE=1.  THE SUM OF
SQUARES OF DEVIATIONS FROM THE SAMPLE MEAN THEN HAS A CHI-SQUARE
DISTRIBUTION WITH (N-1) DEGREES OF FREEDOM WHERE N IS THE NUMBER
OF OBSERVATIONS. ONE OF THE APPLICATIONS FOR THE CHI-SQUARE
DISTRIBUTION IS TO DETERMINE THE CONFIDENCE LIMITS OF THE VARIANCE
ESTIMATION FOR NORMALLY DISTRIBUTED DATA.
TO USE THIS PROGRAM TO DETERMINE THE VALUE OF CHI-SQUARE, THE
USER MUST PROVIDE THE PERCENTAGE AREA AND THE DEGREES OF FREEDOM
FOR THE VARIANCE ESTIMATION.  THE UPPER PERCENTAGE AREA IS THAT
INTEGRATED FROM THE DESIRED VALUE TO INFINITY OF CHI-SQUARE.

ANSWER EACH QUESTION AFTER A QUESTION MARK (?) AND THEN PRESS ENTER.
IF YOU WANT TO START OVER AGAIN, PRESS FUNCTION KEY 1 AND RETURN.

INPUT DEGREES OF FREEDOM=? 12
INPUT UPPER PERCENTAGE AREA %=? 95

*****************************************************
DEGREES OF FREEDOM =                    12
CHI-SQUARE ( 12 , .95 ) = 5.22471
THE PERCENTAGE AREA =                   95  %
*****************************************************


 DO YOU WISH TO DO ANOTHER CALCULATION (Y/N)? N
```

```
*****************************************
*     PROGRAM MENU ........PAGE 1       *
*****************************************

  1. LINEAR REGRESSION
  2. CALCULATION OF NORMAL DEVIATE Z
  3. CALCULATION OF THE PERCENTAGE AREA OF NORMAL DISTRIBU-
     TION (FROM MINUS INFINITY TO NORMAL DEVIATE Z)
  4. CALCULATION OF STUDENT T
  5. CALCULATION OF THE PERCENTAGE AREA OF STUDENT T DIST-
     RIBUTION (FROM MINUS INFINITY TO STUDENT T)
  6. CALCULATION OF CHI SQUARE
  7. CALCULATION OF SAMPLE MEAN,STANDARD DEVIATION, AND CON-
     FIDENCE INTERVALS FOR THE POPULATION MEAN AND VARIANCE
  8. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
     THE VARIANCE
  9. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
     THE SAMPLE MEAN
 10. CALCULATION OF THE PROBABILITY OF EXCEEDING A STANDARD
 11. HYPOTHESIS TESTING
 12. POWER SPECTRUM ANALYSIS
 13. PROCEED TO NEXT PAGE
 14. QUIT
 TYPE THE DESIRED OPTION NUMBER AND PRESS ENTER ? 7

*********************************************************
* 7. CALCULATION OF SAMPLE MEAN, STANDARD DEVIATION, AND *
*    CONFIDENCE INTERVALS FOR THE MEAN AND THE VARIANCE  *
*********************************************************


ANSWER EACH QUESTION AFTER A QUESTION MARK (?) AND THEN PRESS ENTER.
IF YOU WANT TO START OVER AGAIN, PRESS FUNCTION KEY 1 AND RETURN.

DATA MUST BE INPUT TO THE COMPUTER BY THE USER EITHER FROM THE
KEYBOARD OR FROM A FILE ON THE DISK BEFORE CALCULATIONS CAN BE
PERFORMED.


DEFINE Y(VARIABLE SUCH AS TOC, ETC)=? TOC

IS YOUR DATA STORED IN A FILE (Y/N) ? Y
INPUT FILENAME(NO MORE THAN 8 CHARACTERS)
DATA FROM DISK A, TYPE A:  DATA FROM HARD DISK, TYPE C: FIRST
THEN TYPE FILENAME.  EXAMPLE FILENAMES  LIN1 AND LIN2 ARE
AVAILABLE ON DISK A FOR YOUR USE.  IF YOU HAVE ALREADY LOADED
THE PROGRAMS TO THE HARD DISK, THEN THE FILENAMES ARE ON THE
HARD DISK.


DO YOU WISH TO LIST THE FILENAME BEFORE YOU PROCEED(Y/N) ? N
HOW MANY SETS OF DATA TO BE RETRIVED? 2
 TYPE FILENAME FOR SET 1 ? A:LIN1
 TYPE FILENAME FOR SET 2 ? A:LIN2
```

```
SELECT NUMBER OF OPTION:
          1. LIST INPUT DATA
          2. MODIFY OR ADD INPUT DATA
          3. DELETE SOME OF INPUT DATA
          4. STORE DATA
          5. START TO CALCULATE
OPTION ? 1
LISTING OF DATA
FOR SET 1
DATA POINTS      Y
 1              50
 2              30
 3              40
 4              30
 5              35
FOR SET 2
DATA POINTS      Y
 1              40
 2              40
 3              35
 4              45
 5              50
 6              35

SELECT NUMBER OF OPTION:
          1. LIST INPUT DATA
          2. MODIFY OR ADD INPUT DATA
          3. DELETE SOME OF INPUT DATA
          4. STORE DATA
          5. START TO CALCULATE
OPTION ? 5


INPUT CONFIDENCE LEVEL % FOR THE POPULATION MEAN? 95
INPUT CONFIDENCE LEVEL % FOR THE STANDARD DEVIATION? 95

         FOR SAMPLE SET  1
         START TO CALCULATE THE SAMPLE MEAN, STANDARD DEVIATION
         , CONFIDENCE INTERVALS FOR THE MEAN AND STANDARD DEVIATION.
```

```
*****************************************************
NUMBER OF SAMPLES =                          5
MEAN =                                       37
ESTIMATED STANDARD DEVIATION =               8.3666
DEGREES OF FREEDOM =                         4
CONFIDENCE LEVEL FOR THE MEAN =              95  %
THE STUDENT T =                              2.776367

THE CONFIDENCE INTERVAL FOR THE MEAN
 26.61179 <= POPULATION MEAN <= 47.38822

CONFIDENCE LEVEL FOR THE VARIANCE =          95  %
XV2=CHI-SQUARE
XV2( 4 , .975 ) =                            .4834985
XV2( 4 , .025 ) =                            11.15244

THE CONFIDENCE INTERVAL FOR THE STANDARD DEVIATION
 5.010649 <= STANDARD DEVIATION <= 24.06476
*****************************************************

PRESS ENTER TO CONTINUE?
         FOR SAMPLE SET  2
         START TO CALCULATE THE SAMPLE MEAN, STANDARD DEVIATION
         , CONFIDENCE INTERVALS FOR THE MEAN AND STANDARD DEVIATION.



*****************************************************
NUMBER OF SAMPLES =                          6
MEAN =                                       40.83333
ESTIMATED STANDARD DEVIATION =               5.845227
DEGREES OF FREEDOM =                         5
CONFIDENCE LEVEL FOR THE MEAN =              95  %
THE STUDENT T =                              2.570313

THE CONFIDENCE INTERVAL FOR THE MEAN
 34.69979 <= POPULATION MEAN <= 46.96688

CONFIDENCE LEVEL FOR THE VARIANCE =          95  %
XV2=CHI-SQUARE
XV2( 5 , .975 ) =                            .8301781
XV2( 5 , .025 ) =                            12.83213

THE CONFIDENCE INTERVAL FOR THE STANDARD DEVIATION
 3.64869 <= STANDARD DEVIATION <= 14.345
*****************************************************

PRESS ENTER TO CONTINUE?


DO YOU WANT TO DO ANOTHER CALCULATION (Y/N)? N
```

```
*****************************************
*     PROGRAM MENU ........PAGE 1       *
*****************************************

 1. LINEAR REGRESSION
 2. CALCULATION OF NORMAL DEVIATE Z
 3. CALCULATION OF THE PERCENTAGE AREA OF NORMAL DISTRIBU-
    TION (FROM MINUS INFINITY TO NORMAL DEVIATE Z)
 4. CALCULATION OF STUDENT-T
 5. CALCULATION OF THE PERCENTAGE AREA OF STUDENT T DIST-
    RIBUTION (FROM MINUS INFINITY TO STUDENT T)
 6. CALCULATION OF CHI SQUARE
 7. CALCULATION OF SAMPLE MEAN,STANDARD DEVIATION, AND CON-
    FIDENCE INTERVALS FOR THE POPULATION MEAN AND VARIANCE
 8. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
    THE VARIANCE
 9. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
    THE SAMPLE MEAN
10. CALCULATION OF THE PROBABILITY OF EXCEEDING A STANDARD
11. HYPOTHESIS TESTING
12. POWER SPECTRUM ANALYSIS
13. PROCEED TO NEXT PAGE
14. QUIT
TYPE THE DESIRED OPTION NUMBER AND PRESS ENTER ? 8


*************************************************************************
* 8. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF THE VARIANCE *
*************************************************************************


THE USER HAS TO PROVIDE THE ALLOWABLE ERROR RATIO ( WIDTH OF CONFIDENCE
INTERVAL OF STANDARD DEVIATION / STANDARD DEVIATION ) AND THE CONFIDENCE
LEVEL FOR THE MEAN.

ANSWER EACH QUESTION AFTER A QUESTION MARK (?) AND THEN PRESS ENTER.
IF YOU WANT TO START OVER AGAIN, PRESS FUNCTION KEY 1 AND RETURN.

ALLOWABLE ERROR RATIO = ? .5
CONFIDENCE LEVEL % =    ? 95
     ***********************************************
      NUMBER OF SAMPLE REQUIRED=                34
      THE CONFIDENCE LEVEL FOR THE MEAN=        95
      ALLOWABLE ERROR OF STANDARD DEVIATION=    .5
     ***********************************************


        DO YOU WISH TO DO ANOTHER CALCULATION (Y/N)? N
```

```
*****************************************
*      PROGRAM MENU ........PAGE 1       *
*****************************************

 1. LINEAR REGRESSION
 2. CALCULATION OF NORMAL DEVIATE Z
 3. CALCULATION OF THE PERCENTAGE AREA OF NORMAL DISTRIBU-
     TION (FROM MINUS INFINITY TO NORMAL DEVIATE Z)
 4. CALCULATION OF STUDENT T
 5. CALCULATION OF THE PERCENTAGE AREA OF STUDENT T DIST-
     RIBUTION (FROM MINUS INFINITY TO STUDENT T)
 6. CALCULATION OF CHI SQUARE
 7. CALCULATION OF SAMPLE MEAN,STANDARD DEVIATION, AND CON-
     FIDENCE INTERVALS FOR THE POPULATION MEAN AND VARIANCE
 8. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
     THE VARIANCE
 9. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
     THE SAMPLE MEAN
10. CALCULATION OF THE PROBABILITY OF EXCEEDING A STANDARD
11. HYPOTHESIS TESTING
12. POWER SPECTRUM ANALYSIS
13. PROCEED TO NEXT PAGE
14. QUIT
TYPE THE DESIRED OPTION NUMBER AND PRESS ENTER ? 9


*********************************************************************
* 9. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF THE MEAN *
*********************************************************************


THIS PROGRAM IS TO DETERMINE THE SAMPLE NUMBER BASED ON THE ACCURACY
OF THE MEAN.  IF THE CALCULATED SAMPLE NUMBER IS LESS THEN 3, THEN
3 IS SELECTED. THE USER HAS TO PROVIDE THE FOLLOWING:

CONFIDENCE LEVEL % FOR THE MEAN
COEFFICIENT OF VARIATION (STANDARD DEVIATION/SAMPLE MEAN) IN %
ERROR % OF THE MEAN REQUIRED ((SAMPLE MEAN-POPULATION
 MEAN )/POPULATION MEAN)

ANSWER EACH QUESTION AFTER A QUESTION MARK (?) AND THEN PRESS ENTER.
PRESS FUNCTION KEY 1 AND THEN PRESS ENTER TO START OVER AGAIN.

      CONFIDENCE LEVEL % FOR THE MEAN=? 95
      COEFFICIENT OF VARIATION IN %=  ? 50
      ERROR % OF THE MEAN=            ? 10
*************************************************************
CONFIDENCE LEVEL % FOR THE MEAN=                          95
COEFFICIENT OF VARIATION (STANDARD DEVIATION/SAMPLE MEAN= 50  %
ERROR OF THE MEAN=                                        10  %
NUMBER OF SAMPLES REQUIRED=                                99
*************************************************************
```

```
DO YOU WISH TO DO ANOTHER CALCULATION (Y/N)? N

      ****************************************
      *     PROGRAM MENU ........PAGE 1      *
      ****************************************

 1. LINEAR REGRESSION
 2. CALCULATION OF NORMAL DEVIATE Z
 3. CALCULATION OF THE PERCENTAGE AREA OF NORMAL DISTRIBU-
    TION (FROM MINUS INFINITY TO NORMAL DEVIATE Z)
 4. CALCULATION OF STUDENT T
 5. CALCULATION OF THE PERCENTAGE AREA OF STUDENT T DIST-
    RIBUTION (FROM MINUS INFINITY TO STUDENT T)
 6. CALCULATION OF CHI SQUARE
 7. CALCULATION OF SAMPLE MEAN,STANDARD DEVIATION, AND CON-
    FIDENCE INTERVALS FOR THE POPULATION MEAN AND VARIANCE
 8. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
    THE VARIANCE
 9. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
    THE SAMPLE MEAN
10. CALCULATION OF THE PROBABILITY OF EXCEEDING A STANDARD
11. HYPOTHESIS TESTING
12. POWER SPECTRUM ANALYSIS
13. PROCEED TO NEXT PAGE
14. QUIT
TYPE THE DESIRED OPTION NUMBER AND PRESS ENTER ? 10


*******************************************************
* 10. PROBABILITY OF AN EFFLUENT EXCEEDING A STANDARD *
*******************************************************


THIS PROGRAM IS TO INVESTIGATE THE PROBABILITY OF EXCEEDING
A STANDARD. THIS REQUIRES THE KNOWLEDGE OF :

1. POPULATION MEAN
2. STANDARD DEVIATION
3. THE STANDARD NOT TO BE EXCEEDED

ANSWER EACH QUESTION AFTER A QUESTION MARK (?) AND PRESS ENTER.
PRESS FUNCTION KEY 1 AND THEN PRESS ENTER TO START OVER AGAIN.

 INPUT STANDARD=                    ? 100
 INPUT POPULATION MEAN=             ? 120
 INPUT STANDARD DEVIATION=          ? 20


*******************************************************
STANDARD=                                   100
POPULATION MEAN=                            120
STANDARD DEVIATION=                         20
PROBABILITY OF EXCEEDING STANDARD =  84.13449  %
*******************************************************
```

```
DO YOU WISH TO DO ANOTHER CALCULATION (Y/N)? N

        ****************************************
        *     PROGRAM MENU ........PAGE 1      *
        ****************************************

 1. LINEAR REGRESSION
 2. CALCULATION OF NORMAL DEVIATE Z
 3. CALCULATION OF THE PERCENTAGE AREA OF NORMAL DISTRIBU-
    TION (FROM MINUS INFINITY TO NORMAL DEVIATE Z)
 4. CALCULATION OF STUDENT T
 5. CALCULATION OF THE PERCENTAGE AREA OF STUDENT T DIST-
    RIBUTION (FROM MINUS INFINITY TO STUDENT T)
 6. CALCULATION OF CHI SQUARE
 7. CALCULATION OF SAMPLE MEAN,STANDARD DEVIATION, AND CON-
    FIDENCE INTERVALS FOR THE POPULATION MEAN AND VARIANCE
 8. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
    THE VARIANCE
 9. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
    THE SAMPLE MEAN
10. CALCULATION OF THE PROBABILITY OF EXCEEDING A STANDARD
11. HYPOTHESIS TESTING
12. POWER SPECTRUM ANALYSIS
13. PROCEED TO NEXT PAGE
14. QUIT
TYPE THE DESIRED OPTION NUMBER AND PRESS ENTER ? 11

          *************************
          * 11. HYPOTHESIS TESTING *
          *************************

HYPOTHESIS TESTING IS TO TEST WHETHER A SAMPLE COMES FROM
A PARTICULAR DISTRIBUTION. IN ORDER TO USE THIS HYPOTHESIS
TESTING, THE USER CAN SELECT ANY OPTIONS IF NO DATA POINTS
HAVE BEEN ENTERED.  IF DATA POINTS ARE TO BE ENTERED FROM THE
KEY BOARD OR FROM A FILE ON THE DISK, THEN OPTION 3 MUST BE SELECTED.

GROUP 1                              GROUP 2
     1. POPULATION MEAN                   1. POPULATION MEAN
     2. SAMPLE MEAN                       2. ONE SAMPLE VALUE
     3. NUMBER OF SAMPLE                  3. POPULATION STANDARD DEVIATION
     4. POPULATION STANDARD DEVIATION     4. CONFIDENCE LEVEL % FOR THE MEAN
     5. CONFIDENCE LEVEL % FOR THE MEAN
GROUP 3
     1. POPULATION MEAN
     2. SAMPLE MEAN
     3. NUMBER OF SAMPLE
     4. SAMPLE STANDARD DEVIATION
     5. CONFIDENCE LEVEL % FOR THE MEAN

ANSWER EACH QUESTION AFTER A QUESTION MARK (?) AND THEN PRESS ENTER.
IF YOU WANT TO START OVER AGAIN, PRESS FUNCTION KEY 1 AND RETURN.
INPUT GROUP NUMBER (1-3) =            ? 3
DEFINE Y(VARIABLE SUCH AS TOC, ETC)=? TOC
```

```
DO YOU HAVE TO INPUT DATA SET (Y/N)? Y
IS YOUR DATA STORED IN A FILE (Y/N) ? Y
INPUT FILENAME(NO MORE THAN 8 CHARACTERS)
DATA FROM DISK A, TYPE A:  DATA FROM HARD DISK, TYPE C: FIRST
THEN TYPE FILENAME.  EXAMPLE FILENAMES  LIN1 AND LIN2 ARE
AVAILABLE ON DISK A FOR YOUR USE.  IF THE PROGRAMS HAVE BEEN
LOADED TO THE HARD DISK, THEN THE FILENAMES ARE ON THE HARD DISK.


DO YOU WISH TO LIST THE FILENAME BEFORE YOU PROCEED(Y/N) ? N
HOW MANY SETS OF DATA TO BE RETRIVED? 2
 TYPE FILENAME FOR SET 1 ? A:LIN1
 TYPE FILENAME FOR SET 2 ? A:LIN2

SELECT NUMBER OF OPTION:
          1. LIST INPUT DATA
          2. MODIFY OR ADD INPUT DATA
          3. DELETE SOME OF INPUT DATA
          4. STORE DATA
          5. START TO CALCULATE
OPTION ? 1
LISTING OF DATA
FOR SET 1
DATA POINTS       Y
 1              50
 2              30
 3              40
 4              30
 5              35
FOR SET 2
DATA POINTS       Y
 1              40
 2              40
 3              35
 4              45
 5              50
 6              35

SELECT NUMBER OF OPTION:
          1. LIST INPUT DATA
          2. MODIFY OR ADD INPUT DATA
          3. DELETE SOME OF INPUT DATA
          4. STORE DATA
          5. START TO CALCULATE
OPTION ? 5
```

```
 FOR SET                                    1
 INPUT POPULATION MEAN=                  ? 40
 INPUT CONFIDENCE LEVEL % FOR THE POPULATION MEAN=? 95


*************************************************************
FOR SET                                     1
POPULATION MEAN=                            40
SAMPLE MEAN=                                37
NUMBER OF SAMPLES=                          5
SAMPLE STANDARD DEVIATION=               8.3666
THE STUDENT T FOR 95  % CONFIDENCE LEVEL=          2.776367
THE CALCULATED T VALUE=                     -.8017838
THE SAMPLE HAS A MEAN EQUAL TO THE POPULATION MEAN
*************************************************************

 FOR SET                                    2
 INPUT POPULATION MEAN=                  ? 40
 INPUT CONFIDENCE LEVEL % FOR THE POPULATION MEAN=? 95


*************************************************************
FOR SET                                     2
POPULATION MEAN=                            40
SAMPLE MEAN=                                40.83333
NUMBER OF SAMPLES=                          6
SAMPLE STANDARD DEVIATION=               5.845227
THE STUDENT T FOR 95  % CONFIDENCE LEVEL=          2.570313
THE CALCULATED T VALUE=                      .3492146
THE SAMPLE HAS A MEAN EQUAL TO THE POPULATION MEAN
*************************************************************
DO YOU WANT ANOTHER CALCULATION (Y/N)? N
```

```
*****************************************
*     PROGRAM MENU ........PAGE 1       *
*****************************************

 1. LINEAR REGRESSION
 2. CALCULATION OF NORMAL DEVIATE Z
 3. CALCULATION OF THE PERCENTAGE AREA OF NORMAL DISTRIBU-
    TION (FROM MINUS INFINITY TO NORMAL DEVIATE Z)
 4. CALCULATION OF STUDENT T
 5. CALCULATION OF THE PERCENTAGE AREA OF STUDENT T DIST-
    RIBUTION (FROM MINUS INFINITY TO STUDENT T)
 6. CALCULATION OF CHI SQUARE
 7. CALCULATION OF SAMPLE MEAN,STANDARD DEVIATION, AND CON-
    FIDENCE INTERVALS FOR THE POPULATION MEAN AND VARIANCE
 8. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
    THE VARIANCE
 9. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
    THE SAMPLE MEAN
10. CALCULATION OF THE PROBABILITY OF EXCEEDING A STANDARD
11. HYPOTHESIS TESTING
12. POWER SPECTRUM ANALYSIS
13. PROCEED TO NEXT PAGE
14. QUIT
TYPE THE DESIRED OPTION NUMBER AND PRESS ENTER ? 12


                 ******************************
                 * 12. POWER SPECTRUM ANALYSES *
                 ******************************


THIS PROGRAM CONDUCTS A POWER SPECTRUM ANALYSIS IN WHICH
THE MINIMUM SAMPLING FREQUENCY REQUIRED TO CAPTURE A SPECIFIC
WATER QUALITY EVENT CAN BE DETERMINED.  THE PROGRAM DETERMINES
THE MAGNITUDE OF COMPONENTS OF THE TOTAL VARIANCE OF A RECORD
THAT RECUR AT CONSTANT TIME INTERVALS.

ANSWER EACH QUESTION AFTER A QUESTION MARK (?) AND THEN PRESS ENTER.
PRESS FUNCTION KEY 1 AND THEN PRESS ENTER TO START OVER AGAIN.


DEFINE X (TIME UNIT SUCH AS DAYS, HOURS)= ? HOURS
DEFINE Y(DEPENDENT VARIABLE SUCH AS TOC)= ? TOC
INPUT SAMPLING TIME INTERVAL (IN TIME UNIT)=? 5
 YOUR DATA STORED IN A FILE MUST BE IN Y(DEPENDENT VAR)
 FORMAT SUCH AS 10,20,25,...
 AN EXAMPLE FILE TEST3.DAT IS ON THIS DISK FOR YOUR USE.
IS YOUR DATA STORED IN A FILE (Y/N) ? Y
DO YOU WISH TO LIST THE FILENAME BEFORE YOU CONTINUE TO PROCEED
(Y/N)? N
INPUT FILENAME(NO MORE THAN 8 CHARACTERS)
DATA FROM DISK A, TYPE A:  DATA FROM HARD DISK, TYPE C: FIRST
THEN TYPE FILENAME? A:TEST3.DAT
```

```
SELECT NUMBER OF OPTION:
      1. LIST INPUT DATA
      2. MODIFY OR ADD INPUT DATA
      3. DELETE SOME OF THE DATA
      4. PERFORM CALCULATIONS
      5. STORE DATA
      6. GO TO MAIN MENU
      7. DO ANOTHER CALCULATION
 OPTION ? 1

RECORD OF DATA OF DEPENDENT VARIABLE AT SAMPLING TIME INTERVALS
        Y

849.96        964.51        916.73        879.38        983.39
1168.8        1235.43       1045.1        672.61        351.6
264.6         381.75        504.82        470.01        306.53
194.99        271.45        477.85        612.64        532.92
305.88        155.24        251.97        553.75        849.81
964.5         916.78        879.36        983.27        1168.68
1235.47       1045.33       672.88        351.75        264.58
381.64        504.78        470.1         306.65        195.01
271.33        477.7         612.62        533.05        306.05
155.27        251.81        553.51        849.65        964.49
916.84        879.34        983.15        1168.57       1235.51
1045.55       673.16        351.91        264.55        381.53
504.74        470.19        306.78        195.03        271.21
477.55        612.59        533.19        306.21        155.3
251.65        553.27        849.49
                  THE TOTAL NUMBER OF SAMPLES=     73


SELECT NUMBER OF OPTION:
      1. LIST INPUT DATA
      2. MODIFY OR ADD INPUT DATA
      3. DELETE SOME OF THE DATA
      4. PERFORM CALCULATIONS
      5. STORE DATA
      6. GO TO MAIN MENU
      7. DO ANOTHER CALCULATION
 OPTION ? 4

     INPUT NUMBER OF LAGS REQUIRED FOR THE CALCULATION OF
     AUTOCORRELATION COEFFICIENTS. THE SUGGESTED NUMBER IS ABOUT
     15 % OF THE TOTAL NUMBER OF SAMPLES. SELECT A NUMBER THAT
     CAN DIVIDE 360 AND GENERATE AN INTEGER

     THE DEFAULT LAG NUMBER IS= 12

     DO YOU WANT TO CHANGE THE LAG NUMBER (Y/N)? N
     SAMPLE MEAN=                                  601.4425
     ESTIMATED STANDARD DEVIATION=                 323.0803

      DO YOU WANT TO SEE THE DETAILS (Y/N)? N
```

```
LAG        VARIANCE CONTRIBUTION/        SAMPLING INTERVAL
           TOTAL VARIANCE % ( >.5 %)     TO CAPTURE A SPECIFIC
           (EVENTS OF SIGNIFICANCE       WATER QUALITY EVENT
           LARGE NUMBER, GREATER
           SIGNIFICANCE)
 0               26.22311                INFINITE
 1               32.52712                 40
 2               19.12219                 20
 3               7.247379                 13.33333
 4               10.41866                 10
 5               4.551088                 8

LAG  NOMINAL PERIOD
 0     INFINITE
 1         120
 2         60
 3         40
 4         30
 5         24

TO CAPTURE ALL SPECIFIC WATER QUALITY EVENTS, THE MINIMUM SAMPLING
INTERVAL SHOULD BE 8  TIME UNITS
 PRESS ENTER TO CONTINUE?
THE NORMAL PERIOD MEANS THAT THE EVENT WILL REPEAT ITSELF AFTER
A PERIOD OF TIME. FOR EXAMPLE, THERE IS AN EVENT THAT REPEATS
ITSELF EVERY 24  TIME UNITS. IN ORDER TO CAPTURE THE EVENT,
THE MINIMUM SAMPLING INTERVAL MUST BE 8  TIME UNITS OR LESS.
OF COURSE, THIS ALSO CAPTURES ALL EVENTS WITH A PERIOD GREATER
THAN 24  TIME UNITS.

SELECT NUMBER OF OPTION:
      1. LIST INPUT DATA
      2. MODIFY OR ADD INPUT DATA
      3. DELETE SOME OF THE DATA
      4. PERFORM CALCULATIONS
      5. STORE DATA
      6. GO TO MAIN MENU
      7. DO ANOTHER CALCULATION
 OPTION ? 6
```

```
*****************************************
*     PROGRAM MENU ........PAGE 1       *
*****************************************

 1. LINEAR REGRESSION
 2. CALCULATION OF NORMAL DEVIATE Z
 3. CALCULATION OF THE PERCENTAGE AREA OF NORMAL DISTRIBU-
     TION (FROM MINUS INFINITY TO NORMAL DEVIATE Z)
 4. CALCULATION OF STUDENT T
 5. CALCULATION OF THE PERCENTAGE AREA OF STUDENT T DIST-
     RIBUTION (FROM MINUS INFINITY TO STUDENT T)
 6. CALCULATION OF CHI SQUARE
 7. CALCULATION OF SAMPLE MEAN,STANDARD DEVIATION, AND CON-
     FIDENCE INTERVALS FOR THE POPULATION MEAN AND VARIANCE
 8. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
     THE VARIANCE
 9. CALCULATION OF SAMPLE NUMBER BASED ON THE ACCURACY OF
     THE SAMPLE MEAN
10. CALCULATION OF THE PROBABILITY OF EXCEEDING A STANDARD
11. HYPOTHESIS TESTING
12. POWER SPECTRUM ANALYSIS
13. PROCEED TO NEXT PAGE
14. QUIT
TYPE THE DESIRED OPTION NUMBER AND PRESS ENTER ? 13


*******************************************
* PROGRAM MENU....................PAGE 2  *
*******************************************


15  COMPARING TWO MEANS
16. CALCULATION OF THE PERCENTAGE AREA % IN F-DISTRIBUTION
17. CALCULATION OF THE F VALUE IN F-DISTRIBUTION
18. TEST FOR SIGNIFICANT DIFFERENCE BETWEEN
    VARIABILITIES OF TWO SAMPLES
19. TEST FOR SIGNIFICANT DIFFERENCE BETWEEN THE
    POPULATION VARIANCE AND THE SAMPLE VARIANCE
20. RETURN TO PREVIOUS PAGE
21. QUIT

TYPE THE DESIRED OPTION NUMBER AND PRESS ENTER ? 15
```

```
**************************
* 15. COMPARING TWO MEANS
**************************

THIS PROGRAM IS TO COMPARE TWO MEANS IN ORDER TO DETERMINE IF BOTH
MEANS ORIGINATE FROM THE SAME POPULATION.  FOR EXAMPLE, TWO DIFFERENT
PROCESSES ARE COMPARED TO DETERMINE ANY STATISTICAL DIFFERENCE EXISTS.
THE COMPARISION IS TWO-TAILED TEST. THE NULL HYPOTHESIS TO BE TESTED
IS Ho: MEANS ARE EQUAL. AGAINST Ha: MEANS ARE NOT EQUAL.
IF YOU WANT TO DETERMINE WHETHER PROCESS 1 IS BETTER THAN PROCESS 2,
THEN THE COMPARISION IS ONE-TAILED. THE NULL HYPOTHESIS TO BE TESTED
IS Ho: MEANS ARE EQUAL. AGAINST ALTERNATIVE HYPOTHESIS Ha: MEAN 1
IS GREATER THAN MEAN 2 OR VICE VERSUS.
TWO GROUPS ARE CONSIDERED HERE.  IN GROUP 2, DATA SET DON'T HAVE
TO BE ENTERED.  IN GROUP 1, DATA SET MAY BE ENTERED FROM THE KEY
BOARD OR FROM A FILE ON THE DISK. IF DATA SET ARE NOT AVAILABLE,
THE INFORMATION REQUESTED MAY BE ENTERED FROM THE KEY BOARD.
BEFORE THE TWO MEANS ARE COMPARED, THE TWO SAMPLE STANDARD DEVIATIONS
MUST BE COMPARED BY USING F-TEST TO DETERMINE WHETHER THEY ARE
SIGNIFICANTLY DIFFERENT OR NOT.  THE EQUATION TO POOL THE SAMPLE
STANDARD DEVIATIONS DEPENDS ON IT. THEREFORE, IF THEY ARE NOT COMPARED,
GO BACK TO PROGRAM MENU AND SELECT VARIABILITY TEST FOR SIGNIFICANT
DIFFERENCE OF THE TWO SAMPLES AND THEN USE THIS PROGRAM.
 GO BACK  TO PROGRAM MENU (Y/N)? N

GROUP 1                                  GROUP 2
1. TWO SAMPLE MEANS                      1. TWO SAMPLE MEANS
2. NUMBER OF SAMPLES FROM BOTH           2. NUMBER OF SAMPLES FROM BOTH
   SETS OF SAMPLES                          SETS OF SAMPLES
3. SAMPLE STANDARD DEVIATIONS FROM       3. STANDARD DEVIATION
   BOTH SETS OF SAMPLES
4. CONFIDENCE LEVEL % REQUIRED           4. CONFIDENCE LEVEL % REQUIRED
   FOR THE COMPARISON                       FOR THE COMPARISON

   ANSWER EACH QUESTION AFTER A QUESTION MARK (?) AND THEN PRESS ENTER.
     IF YOU WANT TO START OVER AGAIN, PRESS FUNCTION KEY 1 AND RETURN.


     INPUT GROUP NUMBER (1-2)              =? 1

     ARE THE TWO STANDARD DEVIATIONS SIGNIFICANTLY DIFFERENT (Y/N)? N
     WILL DATA SET BE PROVIDED ?  PLEASE ENTER (Y/N)? Y

     DEFINE Y(VARIABLE SUCH AS TOC, ETC)=? TOC

     IS YOUR DATA STORED IN A FILE (Y/N) ? Y
   INPUT FILENAME(NO MORE THAN 8 CHARACTERS)
   DATA FROM DISK A, TYPE A:  DATA FROM HARD DISK , TYPE C: FIRST
   THEN TYPE FILENAME.  EXAMPLE FILENAMES  LIN1 AND LIN2 ARE
   AVAILABLE ON DISK A FOR YOUR USE.  IF THE PROGRAMS HAVE BEEN
   LOADED INTO THE HARD DISK, THEN THE FILENAMES ARE ON THE HARD DISK.
```

```
        DO YOU WISH TO LIST THE FILENAME BEFORE YOU PROCEED(Y/N) ? N

      TYPE FILENAME FOR SET 1 ? A:LIN1
      TYPE FILENAME FOR SET 2 ? A:LIN2

SELECT NUMBER OF OPTION:
          1. LIST INPUT DATA
          2. MODIFY OR ADD INPUT DATA
          3. DELETE SOME OF INPUT DATA
          4. STORE DATA
          5. START TO CALCULATE
OPTION ? 1
LISTING OF DATA
FOR SET 1
DATA POINTS       Y
 1              50
 2              30
 3              40
 4              30
 5              35
FOR SET 2
DATA POINTS       Y
 1              40
 2              40
 3              35
 4              45
 5              50
 6              35

SELECT NUMBER OF OPTION:
          1. LIST INPUT DATA
          2. MODIFY OR ADD INPUT DATA
          3. DELETE SOME OF INPUT DATA
          4. STORE DATA
          5. START TO CALCULATE
OPTION ? 5

      IF NULL HYPOTHESIS Ho: MEANS ARE EQUAL. AGAINST ALTERNATIVE HYPOTHESIS
      Ha: MEANS ARE NOT EQUAL. THEN IT IS TWO-TAILED TEST.
      ENTER Y TO THE FOLLOWING QUESTION.
      IF YOU ARE GOING TO DETERMINE ONE MEAN IS SIGNIFICANTLY GRATER THAN
      THE OTHER OR VICE VERSUS, THEN IT IS ONE-TAILED TEST. ENTER N TO THE
      FOLLOWING QUESTION.
      TO DETERMINE ANY SIGNIFICANT DIFFERENCE BETWEEN THESE TWO MEANS (Y/N)? Y
      THE COMPARISION IS TWO-SIDED.
        INPUT CONFIDENCE LEVEL FOR THE COMPARISON      =? 95
```

```
*************************************************************
THE SAMPLES ARE TOC
MEAN FOR SAMPLE 1                        = 37
MEAN FOR SAMPLE 2                        = 40.83333
NUMBER OF SAMPLES FOR SET 1              = 5
NUMBER OF SAMPLES FOR SET 2              = 6
ESTIMATED STANDARD DEVIATION FOR SAMPLE 1  = 8.3666
ESTIMATED STANDARD DEVIATION FOR SAMPLE 2  = 5.845227
POOLED STANDARD DEVIATION                  = 7.077612
THE STUDENT T FOR 95  % CONFIDENCE LEVEL= 2.262207
THE CALCULATED  STUDENT T                     =-.8944457
A 95 % INTERVAL STATEMENT FOR THE DIFFERENCE IS
      5.861824  >= DIFFERENCE >=-13.52849
THE TWO MEANS HAVE THE SAME POPULATION
THIS IS TWO-SIDED TEST.
*************************************************************

    DO YOU WANT TO DO ANOTHER CALCULATION (Y/N)? N


        ******************************************
        * PROGRAM MENU....................PAGE 2  *
        ******************************************


 15  COMPARING TWO MEANS
 16. CALCULATION OF THE PERCENTAGE AREA % IN F-DISTRIBUTION
 17. CALCULATION OF THE F VALUE IN F-DISTRIBUTION
 18. TEST FOR SIGNIFICANT DIFFERENCE BETWEEN
     VARIABILITIES OF TWO SAMPLES
 19. TEST FOR SIGNIFICANT DIFFERENCE BETWEEN THE
     POPULATION VARIANCE AND THE SAMPLE VARIANCE
 20. RETURN TO PREVIOUS PAGE
 21. QUIT

 TYPE THE DESIRED OPTION NUMBER AND PRESS ENTER ? 16


*************************************************************
* 16. CALCULATION OF THE PERCENTAGE AREA % OF F-DISTRIBUTION *
*************************************************************


 THIS PROGRAM IS TO CALCULATE THE PERCENTAGE AREA OF F-DISTRIBUTION.
 THE AREA IS INTEGRATED FROM 0 TO THE DEFINED VALUE OF F WHICH IS
 PROVIDED BY THE USER.  THE DEGREES OF FREEDOM FOR THE TWO VARIANCES
 ARE ALSO NEEDED.

 ANSWER EACH QUESTION AFTER A QUESTION MARK (?) AND THEN PRESS ENTER.
 IF YOU WANT TO START OVER AGAIN, PRESS FUNCTION KEY 1 AND RETURN.

  INPUT DEGREES OF FREEDOM FOR VARIANCE 1=    ? 12
  INPUT DEGREES OF FREEDOM FOR VARIANCE 2=    ? 12
  INPUT THE DESIRED VALUE OF F= ? 3
```

```
*************************************************************
  THE PERCENTAGE AREA IN THE F-DISTRIBUTION =  96.56724  %
  DEGREES OF FREEDOM FOR VARIANCE 1 =          12
  DEGREES OF FREEDOM FOR VARIANCE 2 =          12
  THE F VALUE =                                3
*************************************************************

DO YOU WISH TO DO ANOTHER CALCULATION (Y/N)? N


        ******************************************
        * PROGRAM MENU....................PAGE 2  *
        ******************************************


 15  COMPARING TWO MEANS
 16. CALCULATION OF THE PERCENTAGE AREA % IN F-DISTRIBUTION
 17. CALCULATION OF THE F VALUE IN F-DISTRIBUTION
 18. TEST FOR SIGNIFICANT DIFFERENCE BETWEEN
     VARIABILITIES OF TWO SAMPLES
 19. TEST FOR SIGNIFICANT DIFFERENCE BETWEEN THE
     POPULATION VARIANCE AND THE SAMPLE VARIANCE
 20. RETURN TO PREVIOUS PAGE
 21. QUIT

 TYPE THE DESIRED OPTION NUMBER AND PRESS ENTER ? 17


 ************************************************************
 *     17. CALCULATION OF THE F VALUE IN F-DISTRIBUTION      *
 ************************************************************


 THIS PROGRAM IS TO CALCULATE THE F VALUE FOR A KNOWN AREA OF
 F-DISTRIBUTION. THE AREA IS THAT INTEGRATED FROM ZERO F TO THE
 DESIRED F IN F-DISTRIBUTION. THE DEGREES OF FREEDOM FOR THE TWO
 VARIANCES ARE ALSO NEEDED.

 ANSWER EACH QUESTION AFTER A QUESTION MARK (?) AND THEN PRESS ENTER.
 IF YOU WANT TO START OVER AGAIN, PRESS FUNCTION KEY 1 AND RETURN.

  INPUT DEGREES OF FREEDOM FOR VARIANCE 1=    ? 12
  INPUT DEGREES OF FREEDOM FOR VARIANCE 2=    ? 12
  INPUT THE DESIRED PERCENTAGE AREA % =       ? 95


        *******************************************************
           THE PERCENTAGE AREA IN THE F-DISTRIBUTION =      95  %
           DEGREES OF FREEDOM FOR VARIANCE 1 =        12
           DEGREES OF FREEDOM FOR VARIANCE 2 =        12
           THE F VALUE =                2.686624
        *******************************************************
```

DO YOU WISH TO DO ANOTHER CALCULATION (Y/N)? N

```
*****************************************
* PROGRAM MENU...................PAGE 2  *
*****************************************
```

15 COMPARING TWO MEANS
16. CALCULATION OF THE PERCENTAGE AREA % IN F-DISTRIBUTION
17. CALCULATION OF THE F VALUE IN F-DISTRIBUTION
18. TEST FOR SIGNIFICANT DIFFERENCE BETWEEN
    VARIABILITIES OF TWO SAMPLES
19. TEST FOR SIGNIFICANT DIFFERENCE BETWEEN THE
    POPULATION VARIANCE AND THE SAMPLE VARIANCE
20. RETURN TO PREVIOUS PAGE
21. QUIT

TYPE THE DESIRED OPTION NUMBER AND PRESS ENTER ? 18

```
*****************************************************
*  18. TESTS FOR SIGNIFICANT DIFFERENCE BETWEEN     *
*            VARIABILITIES OF TWO SAMPLES           *
*****************************************************
```

THE OBJECTIVE IS TO TEST THE DIFFERENCE IN VARIABILITY BETWEEN TWO SAMPLES WHEN THE STANDARD DEVIATION IS NOT KNOWN. FOR EXAMPLE, A NEW EQUIPMENT IS USED TO MEASURE A COMPOUND AND IT IS EXPECTED THAT THE MEASUREMENT UNIFORMITY WOULD IMPROVE. THE QUESTION TO ASK NOW IS THAT WHETHER THE IMPROVEMENT (MORE UNIFORM) REALLY EXISTS OR JUST BY CHANCE. THE NULL HYPOTHESIS Ho: VARIABILITIES ARE EQUAL. AGAINST Ha: VARIABILITY BEFORE IS GREATER THAN VARIABILITY AFTER. THIS IS ONE-TAILED TEST. HOWEVER, IF THE NULL HYPOTHESIS Ho: VARIABILITIES ARE EQUAL AGAINST Ha: VARIABILITIES ARE NOT EQUAL, THEN THIS IS TWO-TAILED TEST.
THE FOLLOWING DATA, FOR EXAMPLE, ARE OBTAINED BEFORE AND AFTER.

	BEFORE	AFTER
NUMBER OF SAMPLES	20	25
ESTIMATED STANDARD DEVIATION	0.88	0.63

TO BE SURE OF A SIGNIFICANT IMPROVEMENT IN VARIABILITY, A RATIO (F RATIO) OF TWO VARIABILITIES BEFORE AND AFTER MUST BE CALCULATED. AND COMPARED WITH THE F VALUE OBTAINED WITH DEGREES OF FREEDOM 19 AND 24 AND AT 95 % CONFIDENCE LEVEL.

PRESS ENTER TO CONTINUE?

THE F RATIO= SQUARE OF (.88/.63)=1.951121
THE F (19,24,.95) =2.039895

```
IF A REAL IMPROVEMENT EXISTS, IT WOULD BE NECESSARY FOR THE CALCULATED
VALUE TO EXCEED 2.03 TO REPORT THAT AN IMPROVEMENT IN VARIABILITY
EXISTS WITH A 95 % CHANCE OF BEING CORRECT. THIS IS ONE-TAILED TEST.
IN THIS CASE THERE IS NO IMPROVEMENT IN VARIABILITY.

TO USE THIS PROGRAM, THE USER MUST PROVIDE THE ESTIMATED STANDARD
DEVIATIONS AND NUMBER OF SAMPLES BEFORE AND AFTER.  THE CONFIDENCE
LEVEL AND ONE- OR TWO-TAILED ARE ALSO NEEDED. ARRANGE THE VARIANCE
OF SAMPLE 1 TO BE GREATER THEN THAT OF SAMPLE 2  IN TWO-TAILED TEST.
INPUT ONE- OR TWO-TAILED (1 OR 2) ? 1
WILL DATA SET BE ENTERED BY THE USER (Y/N)? Y


ANSWER EACH QUESTION AFTER A QUESTION MARK (?) AND THEN PRESS ENTER.
IF YOU WANT TO START OVER AGAIN, PRESS FUNCTION KEY 1 AND RETURN.

IN GENERAL, DATA MUST BE PROVIDED BY THE USER EITHER FROM THE KEYBOARD
OR FROM A FILE ON THE DISK.


DEFINE Y(VARIABLE SUCH AS TOC, ETC)=? TOC

IS YOUR DATA STORED IN A FILE (Y/N) ? Y
INPUT FILENAME(NO MORE THAN 8 CHARACTERS)
DATA FROM DISK A, TYPE A:  DATA FROM HARD DISK, TYPE C: FIRST
THEN TYPE FILENAME.  EXAMPLE FILENAMES  LIN1 AND LIN2 ARE
AVAILABLE ON DISK A FOR YOUR USE.  IF YOU HAVE LOADED THE PROGRAMS
INTO THE HARD DISK, THEN THE FILENAMES ARE ON THE HARD DISK.


DO YOU WISH TO LIST THE FILENAME BEFORE YOU PROCEED(Y/N) ? N

 TYPE FILENAME FOR SET 1 ? A:LIN1
 TYPE FILENAME FOR SET 2 ? A:LIN2

SELECT NUMBER OF OPTION:
          1. LIST INPUT DATA
          2. MODIFY OR ADD INPUT DATA
          3. DELETE SOME OF INPUT DATA
          4. STORE DATA
          5. PROCEED TO THE SELECTED PROGRAM
OPTION ? 1
LISTING OF DATA
FOR SET 1
DATA POINTS      Y
 1              50
 2              30
 3              40
 4              30
 5              35
FOR SET 2
DATA POINTS      Y
 1              40
 2              40
 3              35
 4              45
 5              50
 6              35
```

```
SELECT NUMBER OF OPTION:
          1. LIST INPUT DATA
          2. MODIFY OR ADD INPUT DATA
          3. DELETE SOME OF INPUT DATA
          4. STORE DATA
          5. PROCEED TO THE SELECTED PROGRAM
OPTION ? 5
         INPUT CONFIDENCE LEVEL REQUIRED=      ? 95

                ******************************************************
                   THIS IS ONE-TAILED TEST.
                   CONFIDENCE LEVEL                    =       95  %
                   NUMBER OF SAMPLES BEFORE            =       5
                   NUMBER OF SAMPLES AFTER             =       6
                   SAMPLE MEAN BEFORE                  =       37
                   SAMPLE MEAN AFTER                   =       40.83333
                   SAMPLE STANDARD DEVIATION BEFORE    =       8.3666
                   SAMPLE STANDARD DEVIATION AFTER     =       5.845227
                   F VALUE AT  95  % CONFIDENCE =              5.192483
                   CALCULATED F RATIO                  =       2.04878
                   AN IMPROVEMENT IN VARIABILITY DOES NOT EXIST
                ******************************************************


       DO YOU WANT TO DO ANOTHER CALCULATION (Y/N)? N


                ******************************************
                * PROGRAM MENU....................PAGE 2  *
                ******************************************


         15  COMPARING TWO MEANS
         16. CALCULATION OF THE PERCENTAGE AREA % IN F-DISTRIBUTION
         17. CALCULATION OF THE F VALUE IN F-DISTRIBUTION
         18. TEST FOR SIGNIFICANT DIFFERENCE BETWEEN
             VARIABILITIES OF TWO SAMPLES
         19. TEST FOR SIGNIFICANT DIFFERENCE BETWEEN THE
             POPULATION VARIANCE AND THE SAMPLE VARIANCE
         20. RETURN TO PREVIOUS PAGE
         21. QUIT

         TYPE THE DESIRED OPTION NUMBER AND PRESS ENTER ? 19
```

```
*************************************************************
*  19. TEST FOR SIGNIFICANT DIFFERENCE FOR THE POPULATION   *
*          VARIABILITY AND THE SAMPLE VARIABILITY            *
*************************************************************


THE OBJECTIVE IS TO TEST DIFFERENCES BETWEEN THE SAMPLE VARIABILITY
AND THE POPULATION VARIABILITY.  THE TEST WILL INDICATE THAT
WHETHER THE SAMPLE VARIABILITY IS AN IMPROVEMENT OVER THE
POPULATION VARIABILITY.  FOR EXAMPLE, DOES A LOWER VALUE OF SAMPLE
VARIABILITY MEAN THAT THE NEW MEASUREMENTS ARE SIGNIFICANTLY
MORE UNIFORM?
TO TEST FOR A DIFFERENCE IN VARIABILITY, THE CHI SQUARE TEST
USING THE FOLLOWING FORMULA IS UTILIZED.

CHI SQUARE= DEGREES OF FREEDOM  X SAMPLE VARIANCE / POPULATION
            VARIANCE
PRESS ENTER TO CONTINUE?
IF THE CALCULATED CHI SQUARE IS LARGER THAN THE CHI SQUARE
AT A UPPER 5 % CHI SQUARE DISTRIBUTION AND (N-1) DEGREES OF
FREEDOM, THEN THE SAMPLE VARIABILITY IS SIGNIFICANTLY GREATER
THAN THE POPULATION VARIABILITY.  IF THE CALCULATED CHI SQUARE
IS IN BETWEEN THE CHI SQUARE VALUES AT UPPER 5 AND 95 % CHI SQUARE
DISTRIBUTION, THEN THE SAMPLE VARIABILITY IS NOT SIGNIFICANTLY
LARGER OR SMALLER THAN THE POPULATION VARIABILITY. IF THE CALCULATED
CHI SQUARE VALUE IS SMALLER THAN THE CHI SQUARE VALUE AT A UPPER
95 % LEVEL, THEN THE SAMPLE VARIABILITY IS SIGNIFICANTLY SMALLER
THAN THE POPULATION VARIABILITY.
THE UPPER X % HERE ARE THE TOTAL AREA INTEGRATED FROM INFINITE TO
THE DESIRED VALUE OF CHI-SQUARE.

ANSWER EACH QUESTION AFTER A QUESTION MARK (?) AND THEN PRESS ENTER.
IF YOU WANT TO START OVER AGAIN, PRESS FUNCTION KEY 1 AND RETURN.

INPUT SAMPLE STANDARD DEVIATION=? 20
INPUT POPULATION STANDARD DEVIATION=? 30
INPUT NUMBER OF SAMPLES  =? 20


*****************************************************
SAMPLE STANDARD DEVIATION =          20
POPULATION STANDARD DEVIATION =      30
NUMBER OF SAMPLES =                 20
CHI-SQUARE ( 19 , .95 ) = 10.11973
CALCULATED CHI SQUARE        = 8.444444
THE SAMPLE VARIANCE IS SIGNIFICANTLY SMALLER THAN THE POPULATION
VARIANCE.
*****************************************************


 DO YOU WISH TO DO ANOTHER CALCULATION (Y/N)? N
```

```
*****************************************
* PROGRAM MENU....................PAGE 2  *
*****************************************
```

15 COMPARING TWO MEANS
16. CALCULATION OF THE PERCENTAGE AREA % IN F-DISTRIBUTION
17. CALCULATION OF THE F VALUE IN F-DISTRIBUTION
18. TEST FOR SIGNIFICANT DIFFERENCE BETWEEN
    VARIABILITIES OF TWO SAMPLES
19. TEST FOR SIGNIFICANT DIFFERENCE BETWEEN THE
    POPULATION VARIANCE AND THE SAMPLE VARIANCE
20. RETURN TO PREVIOUS PAGE
21. QUIT

TYPE THE DESIRED OPTION NUMBER AND PRESS ENTER ? 21

.....END.....

NOTE: IF YOU WANT TO SEE THIS PROGRAM MENU AGAIN,
JUST TYPE EMSLSTAT AND THEN PRESS ENTER.

APPENDIX A

DEFINITIONS OF BASIC STATISTICS

NORMAL DISTRIBUTION

The probabilistic model for the frequency distribution of a continuous random variable is called the probability distribution. While these distributions may assume a variety of shapes, a very large number of random variables observed in nature possess a frequency distribution which is bell-shaped and can be approximated by using a normal curve. The density function of a normal distribution of mean μ and variance σ^2 is given by the equation:

$$f(y) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right), \quad -\infty < y < \infty \qquad \text{Eq.(A-1)}$$

This equation for the normal distribution is constructed such that the area under the curve will represent probability. Hence, the total area is equal to one.

The normal distribution is completely determined by two parameters: μ, the population mean of the distribution, and σ, the population standard deviation. The parameter μ is the center of the distribution while σ is a measure of the dispersion of the data from the population mean. Thus, a change in μ merely slides the curve right or left without changing its profile, while a change in σ widens or narrows the curve without changing the location of its center.

In practice, variables seldom are in a range of values from "minus infinity" to "plus infinity". Nevertheless, the relative frequency distribution for many types of measurements will generate a bell-shaped figure which may be approximated by the function shown in Figure A-1. One property of the normal distribution as illustrated in the figure is that randomly selected observations will have approximately a 68.3% probability of falling within the interval $\mu \pm \sigma$, 95.5% within the interval $\mu \pm 2\sigma$ and 99.7% within the interval $\mu \pm 3\sigma$.

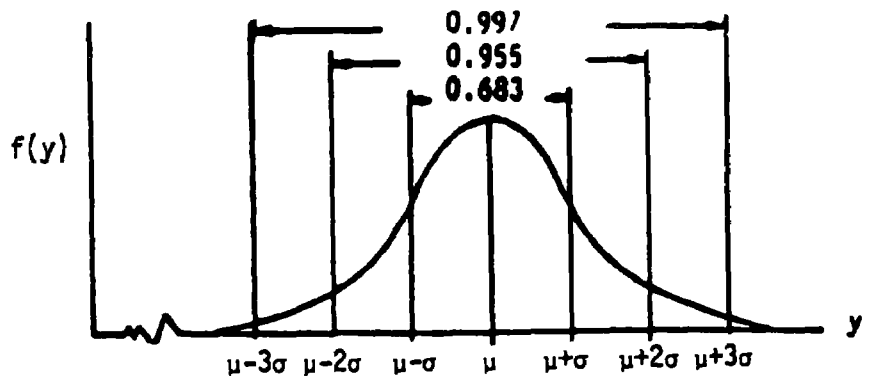

Figure A-1. Normal Distribution

SAMPLE MEAN, $\bar{y}$

A specified number of items (a sample) randomly drawn from a large body of data (a population) is presumed to represent the population. One of the most common and useful measures of the center of the distribution for the sample is the arithmetic mean of a set of measurements. This is often referred to as the sample mean. The arithmetic mean of a set of n measurements $y_1, y_2, y_3, \ldots, y_n$ is equal to the sum of the measurements divided by n. It is used to estimate the population mean μ and can be calculated from the following formula:

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad \text{Eq. (A-2)}$$

SAMPLE VARIANCE, s^2

The population variance σ^2 is an indicator of the spread of a probability distribution about its mean. It is estimated by the sample variance s^2. The sample variance of a set of n measurements $y_1, y_2, y_3, \ldots, y_n$ is equal to the sum of the square of the deviations of the measurements from the mean divided by the degrees of freedom (n-1). The sample variance s^2 can be calculated from the following formula:

$$s^2 = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1} \qquad \text{Eq. (A-3)}$$

SAMPLE STANDARD DEVIATION, s

The sample standard deviation s is the positive square root of the sample variance s^2. It is the estimate of the standard deviation σ which is defined as the distance along the abscissa from the mean to the point of inflection on the normal curve. The standard deviation is, as was the variance, a measure of dispersion and has the same unit of measure as the mean μ. The sample standard deviation can be calculated from the following formula:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n - 1}} \qquad \text{Eq.(A-4)}$$

Both the variance and the standard deviation play an important role in statistics. Since $\bar{y}$ approximates μ and s approximates σ, those percentages as described in the normal distribution will hold approximately for $\bar{y} \pm s$, $\bar{y} \pm 2s$, and $\bar{y} \pm 3s$.

SAMPLE COEFFICIENT OF VARIATION, CV

Another measure of dispersion from the mean is the coefficient of variation CV. The CV provides a measure of the dispersion relative to the location of the data set, so that the spread of the data in sets with different means can be compared. It can be calculated from the following formula:

$$CV = \frac{s}{\bar{y}} \qquad \text{Eq.(A-5)}$$

CENTRAL LIMIT THEOREM

The central limit theorem states that, if random samples of n observations are repeatedly drawn from a population with a finite mean μ and a standard deviation σ, then, when n is large the frequency distribution of the sample means will be approximately bell-shaped. Thus, when n is large, the sample mean, $\bar{y}$, will be approximately normally distributed with mean equal to μ and standard deviation $\sigma/\sqrt{n}$. The approximation will become more accurate as n becomes larger.

RANDOM SAMPLING

A set of observations may be regarded as a random sample from the population if each member of the sample is a random drawing from the whole population. Mathematically speaking, suppose that a sample of n measurements is drawn from a population consisting of N total measurements. There are the following different combinations of n measurements which can be selected from the population:

$$C_n^N = \frac{N!}{n!(N - n)!} \qquad \text{Eq.(A-6)}$$

If the sampling is conducted in such a way that each of the C_n^N samples has an equal probability of being selected, the sampling is said to be random and the result is said to be a random sample.

STUDENT t DISTRIBUTION

A random variable y having a student t distribution with $\beta=(n-1)$ degrees of freedom has a probability density function (pdf) of the form:

$$F(y) = \frac{\Gamma\left(\frac{\beta+1}{2}\right)}{\sqrt{\beta\pi}\ \Gamma(\beta/2)} \quad \frac{1}{(1+y^2/\beta)^{(\beta+1)/2}} \qquad \text{Eq.(A-7)}$$

It is noteworthy that a student t distribution with infinite degrees of freedom is a standard normal distribution. The student t distribution instead of the normal distribution is utilized when the variance of the normal distribution must be estimated from a set of observations. The distribution of

$$t_\beta = \frac{\bar{y} - \mu}{s/\sqrt{n}} \qquad \text{Eq.(A-8)}$$

is a student t distribution with $\beta=n-1$ degrees of freedom. A student t distribution with $\beta = 4$ degrees of freedom is shown in Figure A-2.

CHI SQUARE DISTRIBUTION, X_β^2

If s^2 is the variance of a random sample of size n from a normal distribution, the quantity $\beta s^2/\sigma^2$ has a chi square distribution as shown in Figure A-3. This quantity is represented by X_β^2. The chi square distribution is characterized by one parameter β, the degrees of freedom (n-1). It has a mean value of β and a variance of 2β.

A test based on the chi square test determines whether there is any significant difference between the sample variability and the population variability. For example, does a lower value of sample variance mean that the new measurements are significantly more uniform?

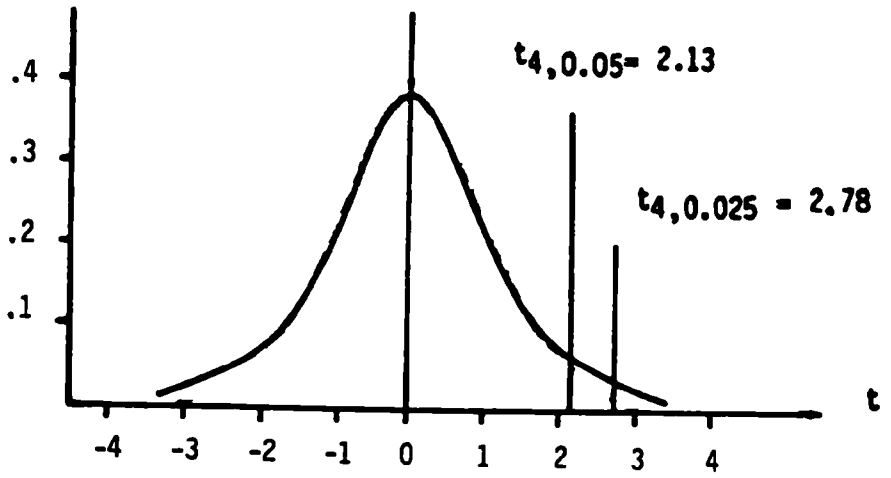

Figure A-2. Distribution of student t with β = 4 degrees of freedom.

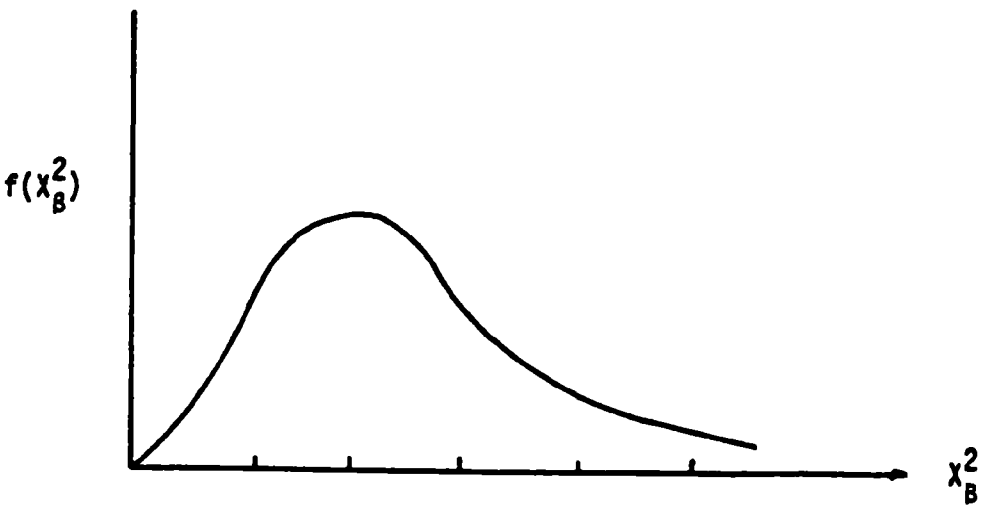

Figure A-3. Chi square distribution.

The mathematical form for this distribution is:

$$f(X_\beta^2) = \frac{1}{2^{\beta/2}\,\Gamma(\beta/2)} \quad (X_\beta^2)^{(\beta-2)/2} \exp\,(-X_\beta^2/2) \quad \text{for } X_\beta^2 > 0 \qquad \text{Eq.(A-9)}$$

$$= 0 \qquad\qquad \text{for } X_\beta^2 \leq 0$$

F DISTRIBUTION

Suppose that a sample of n_1 observations is randomly drawn from a normal distribution having variance σ_1^2, a second sample of n_2 observations is also randomly drawn from a second normal distribution having variance σ_2^2, and estimates s_1^2 and s_2^2 of the two population variances are calculated, having β_1 and β_2 degrees of freedom. The ratio $(X_{\beta_1}^2/\beta_1)/(X_{\beta_2}^2/\beta_2)$ has an F distribution having β_1 and β_2 degrees of freedom. The F distribution may be used to test whether the variances are equal $(\sigma_1^2 = \sigma_2^2)$ by comparing the ratio of the sample variances (s_1^2/s_2^2) with the F distribution having β_1 and β_2 degrees of freedom.

The mathematical form for this distribution is:

$$f(F) = \frac{\Gamma\left(\frac{\beta_1+\beta_2}{2}\right)\beta_1^{\beta_1/2}\,\beta_2^{\beta_2/2}}{\Gamma(\beta_1/2)\,\Gamma(\beta_2/2)}\, F^{\beta_1/2-1}\,(\beta_2+\beta_1 F)^{-(\beta_1+\beta_2)/2} \quad \text{for } F > 0 \qquad \text{Eq.(A-10)}$$

$$= 0 \qquad\qquad \text{for } F \leq 0$$

APPENDIX B

DESCRIPTIONS OF STATISTICAL SAMPLING PROGRAMS ON THE DISK

B.1 CURVE FITTING WITH A LINEAR REGRESSION

A linear regression is a curve fitting technique. It is based on fitting a straight line through a series of observed data points so that the sum of squares of the deviations of these points from the line are minimized. A straight line is expressed by the mathematical form:

$$y = a+bx \qquad \text{Eq.(B-1)}$$

where:

$$a = \overline{y} - b\overline{x} \qquad \text{Eq.(B-1a)}$$

$$b = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \qquad \text{Eq.(B-1b)}$$

$$\sum = \sum_{i=1}^{n} \qquad \text{Eq.(B-1c)}$$

y_i = dependent variable or observed data point

x_i = independent variable

After fitting a line through a set of data, it is necessary to determine if the fit is good. One indicator is the coefficient of correlation (cc), which can have a value between -1 and +1. It may be expressed as:

$$cc = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{(n\sum x_i^2 - \sum x_i \sum x_i)\quad(n\sum y_i^2 - \sum y_i \sum y_i)}} \qquad \text{Eq.(B-2)}$$

A positive coefficient of correlation indicates that increases in the independent variable will result in increases in the dependent variable. The variables are directly related. If the coefficient is negative, then they are indirectly related. A coefficient of correlation of o means that the variables are not related.

B.2 NORMAL DEVIATE Z

Figure A-1 in Appendix A shows the percentage of elements of the population contained in various intervals of a normal distribution. About 68.3% of the area under the normal curve falls within $\mu \pm 1\sigma$. About 95.5% of the area under the curve falls within $\mu \pm 2\sigma$. And about 99.7% of the area under the curve falls within $\mu \pm 3\sigma$. If we introduce Z, which is defined as the distance from the population mean in units of the standard deviation, we can produce a standard normal profile as shown in Figure B-1. The Z value is calculated by the following formula:

$$Z = \frac{y - \mu}{\sigma} \qquad \text{Eq.(B-3)}$$

By redefining Z as the distance from the population mean in units of the standard deviation of the average, the same standard normal profile as shown in Figure B-1 is still produced. Now, however, it represents the normal distribution for the sample average. The Z is calculated by the following formula:

$$Z = \frac{\bar{y} - \mu}{\frac{\sigma}{\sqrt{n}}} \qquad \text{Eq.(B-4)}$$

where:

n is the total number of observations.

$\bar{y}$ is the sample average.

$\frac{\sigma}{\sqrt{n}}$ is the standard deviation of the average.

The percentage area in Figure B-1 only depends upon the value of Z. For example, the percentage area between Z = ±1 is 68.26%. To use this program to obtain the Z value requires the user to provide the confidence level or the percentage area. For example, the Z value for a 95% confidence level in a two-sided test is 1.959961.

B.3 PERCENTAGE AREA UNDER THE NORMAL CURVE

This program is the opposite of the program in (B.2). In this program the user needs to provide the Z value in order to obtain the corresponding percentage area. This program is very useful. For example, it can answer the question, "What is the probability that a single observation y, drawn from a normal distribution with population mean μ=6 and variance σ^2=4, lies between 6 and 9?" By using equation (B-3), two values of Z, namely, 0 and 1.5 are obtained. The corresponding areas for Z=0 and 1.5 are 0.5 and 0.9332, respectively. Therefore, the probability that a single observation y, drawn at random from a normal population with μ=6 and σ^2=4, will have a value (y_i) between 6 and 9 is (.9322-0.5) = 0.4322 or 43.22%. The user must input a value of Z in order to calculate the area of normal distribution (integrated from minus infinity to the desired Z value).

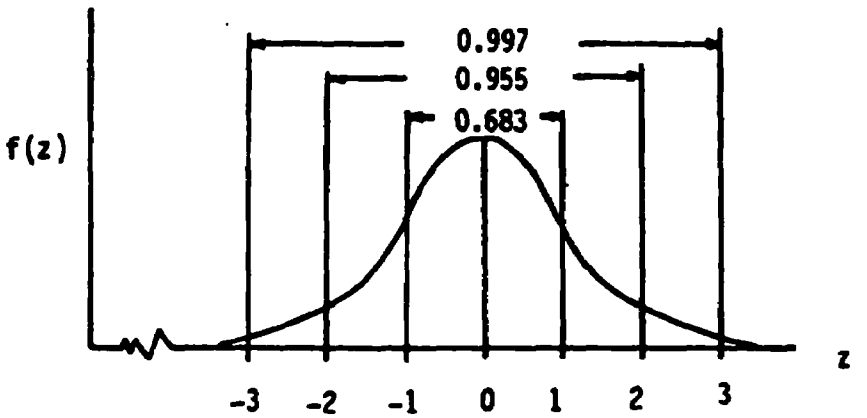

Figure B-1. Standard normal distribution.

B.4 STUDENT t

If the variability σ^2 of a normal distribution is estimated from a set of samples, the student t, instead of the normal deviate Z, should be used when the confidence interval for the mean is to be calculated. The normal curve is replaced by a student t distribution which varies according to the degrees of freedom.

This program requires the user to provide the desired confidence level and the degrees of freedom, $\beta=n-1$, in order to obtain the value of t in a two-sided test.

B.5 PERCENTAGE AREA UNDER THE STUDENT t

This program is the opposite of the program in (B.4). This program requires the input of the value of t in order to obtain the percentage area of the student t distribution. The area is integrated from minus infinity to the provided t value. For example, the percentage area for $\beta = 15$ and $t = 2.7$ is 99.18%.

B.6 CHI SQUARE

One of the applications for the chi square distribution, other than those described in the previous section, is to determine the confidence limits of the variance estimation for normally distributed data.

To use this program to determine the value of chi square, the user must provide the degrees of freedom and desired percentage area (integrated from 0 to the desired chi square value). For example, the value of chi square for β=15 and percentage area = 95% is 24.99.

B.7 SAMPLE MEAN, STANDARD DEVIATION, AND CONFIDENCE INTERVALS FOR THE MEAN AND VARIANCE

If a sample is taken from a population, the sample average will seldom be exactly the same as the population mean. An estimate of an interval that will bracket the population mean is then made. If such interval estimates were made a large number of times, and actually did contain the true mean in 95% of the cases, it might be said that we are operating at a 95% confidence level (C.L.). The interval estimates are called 95% confidence intervals (C.I.). The expressions for confidence intervals for the mean and the variance are:

Confidence Interval (C.I.) for the Population Mean μ if σ is unknown

$$\bar{y} - t_{\beta;\alpha/2} \frac{s}{\sqrt{n}} \leq \mu \leq \bar{y} + t_{\beta;\alpha/2} \frac{s}{\sqrt{n}} \qquad \text{Eq.(B-5)}$$

where:

β = degrees of freedom, (n-1)
α = a significant level, % = C.L. = 100% - α%
$\underline{n}$ = number of samples
$\bar{y}$ = sample mean
$t_{\beta;\alpha/2}$ = student t at degrees of freedom, β, and a significant level α.
s = sample standard deviation.

Confidence Interval (C.I.) for the Population Mean μ if σ is known

$$\bar{y} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{y} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \qquad \text{Eq.(B-6)}$$

where:

$Z_{\alpha/2}$= normal deviate Z at a significant level α.

Confidence Interval for the Variance, σ^2

$$\frac{\beta s^2}{X^2_{\beta;\alpha/2}} \leq \sigma^2 \leq \frac{\beta s^2}{X^2_{\beta;(1-\alpha/2)}} \qquad \text{Eq.(B-7)}$$

The user must provide sample data (individual observations), number of samples, the desired confidence levels for the mean, and the standard deviation in order to use this program.

B.8 DETERMINATION OF THE NUMBER OF SAMPLES

The number of samples necessary to reasonably characterize a water or wastewater can be determined if background data on the concentration and variance of the concentration of the parameter are available. Two techniques can be used to determine the required number of samples; one is based on the allowable confidence interval for the standard deviation, the other on the accuracy of the mean.

Determining the Number of Samples Based on the Accuracy of the Mean

The relationship among the number of samples n, the coefficient of variation $s/\bar{y}$, the accuracy of sample mean μ, the student t with degrees of freedom β, and confidence level $(1-\alpha)$ can be expressed as:

$$n = \left(\frac{t_{\beta;\alpha/2} \dfrac{s}{\bar{y}}}{\dfrac{(\mu - \bar{y})}{\bar{y}}} \right)^2 \qquad \text{Eq.(B-8)}$$

This program requires the user to input the following information:

- Confidence level for the mean
- Coefficient of variation $s/\bar{y}$ (CV)
- Error of the mean $(\mu - \bar{y})/\bar{y}$

For example, given α = 5%, CV=0.5, $(\mu-\bar{y})/\bar{y}$=0.25, the number of samples needed would be 18.

If the coefficient of variation is not already available, it can be estimated by collecting three or four samples to determine the sample mean and the standard deviation.

Determining the Number of Samples Based on the Accuracy of the Sample Variance

The relationship among the number of samples n, the chi square X^2_β and confidence level for the variance s^2 can be expressed as:

$$\frac{\Delta}{s} = \sqrt{\beta}\left[\frac{1}{\sqrt{X^2_{\beta;\ 1-\alpha/2}}} - \frac{1}{\sqrt{X^2_{\beta;\ \alpha/2}}}\right] \qquad \text{Eq.(B-9)}$$

where Δ is the allowable width of the confidence interval for the standard deviation with a confidence level (100% - α).

To apply this method to determine the sample number required, the user must provide the following information:

- Confidence level (100% - α)
- Relative error of the standard deviation Δ/s.

For example, given $\Delta/s = 0.5$ and (100% - α) = 95%, 36 samples would be required.

B.9 PROBABILITY OF EXCEEDING A STANDARD

The probability of exceeding a standard is one of the statistical methods to determine the percentage violation of a parameter being monitored. The user must provide the following information in order to use this program:

- Population mean μ
- Standard deviation σ
- The standard that should not be exceeded, Y.

The probability P of an effluent exceeding a standard can be determined by:

- calculating the Z value using the following formula:

$$Z = \frac{Y - \mu}{\sigma} \qquad \text{Eq.(B-10)}$$

- determining the area from Z to ∞ from the standard normal distribution.

For example, given Y = 100, μ = 75, σ = 18, the probability (P) would be 8.23%.

B.10 HYPOTHESIS TESTING

Hypothesis testing determines whether a sample comes from a particular distribution with a specific parameter. The information required for this program is as follows:

Group 1
1. population mean, μ
2. sample mean, $\bar{y}$
3. number of samples, n
4. standard deviation, σ
5. confidence level for the mean

Group 2
1. population mean, μ
2. one sample value, y
3. standard deviation, σ
4. confidence level for the mean

Group 3
1. population mean, μ
2. sample mean, $\bar{y}$
3. number of sample, n
4. sample standard deviation, s
5. confidence level for the mean

For example, given μ = 100, $\bar{y}$ = 120, n = 10, σ = 50, confidence level = 95%, determine whether the sample comes from a particular distribution with a population mean = 100.

Solution:

Substituting the given conditions into equation (B-4),

$$Z = \frac{\bar{y}-\mu}{\frac{\sigma}{\sqrt{n}}} = 1.265$$

which is less than the normal deviate Z = 1.96 at a 95% confidence level. We conclude that the sample, therefore, has a mean equal to the population mean.

B.11 POWER SPECTRUM ANALYSIS

The use of statistics discussed so far depends on the assumption that the data record remains in equilibrium about a constant mean. For a long-term monitoring program the data may consist of a harmonic and a trend that the variance will not be a random dispersion about a constant mean. If trend and harmonics are not identified or removed, distortions can occur both in data processing and in conclusions on the probability distribution of the measured parameter. Two techniques used to evaluate and identify these components are trend removal and power spectrum analysis.

A trend may be defined as a harmonic component whose period is longer than the record length. The technique to remove a trend requires the method of the least squares or regression analysis. Linear regression can be expressed by a straight line which has the same equation as (B-1):

$$y = a+bx$$

The coefficients a and b are also calculated by the same equations (B-1a) and (B-1b). The new time series, y_i^*, after a linear trend is removed is:

$$y_i^* = y_i - (a+bx_i) \qquad \text{Eq.(B-11)}$$

The other technique is called the power spectrum analysis which is a statistical method for analyzing time-dependent records. It is used to analyze a long, continuous record with high frequency of data acquisition. It should not be used for short surveys or low frequency monitoring when limited amounts of data are available, or if part of the record is missing. The general rule is that if we want to monitor one year's data, we should have accumulated a data record for ten years.

A time-dependent record can be resolved into a spectrum. Any dominant periodicities in the record will appear as peaks in the power spectrum. Therefore, the spectral analysis is used to extract any regular variations of the respective parameter with respect to time. It leads to uncover some of the phenomena governing the variation of the parameter being studied by connecting the frequencies corresponding to the peaks of the power spectrum to physical, chemical and other factors which may be present in the record. In short, power spectrum analysis computes the following:

- Those parts of the total variance of a record which recur at constant time intervals.
- Those parts of the total variance of a record which are not recurring in character either a trend or random fluctuations.
- The frequencies at which different factors cause the record to vary.
- The determination of optional sampling frequency.

The Computation Procedures of a Power Spectrum

An important requirement for the computation of a power spectrum is that there is no missing data in the record. If some of the data are missing, their values must be interpolated before the spectral analysis is attempted. No more than 5% of the total data should be interpolated. The following are the computational procedures:

. Calculates the sample mean of the record.

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad \text{Eq.(B-12)}$$

where n is the number of measurements

. Calculate autocorrelation coefficients C_r.

$$C_r = \frac{1}{n-r} \sum_{i=1}^{n-r} (y_r y_{i+r} - \bar{y}^2) \qquad \text{Eq.(B-13)}$$

where:

$r = 0, 1, 2, \ldots, m$

m = the total number of lags to which the computation is carried out.

n = the number of samples in the record.

. Fourier cosine transform V_r for each autocorrelation coefficient.

$$V_r = \frac{K}{m} \left[C_o + C_m \cos(\pi r) + 2 \sum_{q=1}^{m-1} C_q \cos\left(\frac{q\pi r}{m}\right) \right] \qquad \text{Ea.(B-14)}$$

where:

$K = \frac{1}{2}$ for $r = 0$ and $r = m$

$K = 1$ for $r = 1, 2, . , m - 1$.

- Smooth some distortion of the spectrum for the small sample size.

$$U_0 = 0.54\ V_0 + 0.46\ V_1;$$
$$U_r = 0.23\ V_{r-1} + 0.54\ V_r + 0.23\ V_{r+1};$$
$$r = 1, 2, 3, \ldots, m-1 \qquad \text{Ea.(B-15)}$$
$$U_m = 0.46\ V_{m-1} + 0.54\ V_m;$$

where:

$U_0, U_1, \ldots, U_m$ are power spectrum estimates corresponding to lags 0, 1, 2, ..., m, respectively.

- Calculate percentage contribution of each lag to the total variance of the record.

$$P_r = \frac{U_r}{\sum_{i=0}^{m} U_i} \times 100\% \qquad \text{Eq.(B-16)}$$

Each of these estimates represents the part of the total record variance that is estimated to occur within a certain period of time.

- Calculate period corresponding to each lag t_r.

$$t_r = \frac{2\ m\ \Delta t}{r} \qquad \text{Eq.(B-17)}$$

where:

t_r is the period corresponding to lag r.

Δt is the sampling interval.

The spectral values, U_r, represent estimates over a range of periods from having a band with limits of

$$\left(\frac{r}{2\ m\Delta t} - \frac{1}{2m\Delta t}\right) \text{ to } \left(\frac{r}{2m\Delta t} + \frac{1}{2m\Delta t}\right) \qquad \text{Eq.(B-18)}$$

In other words, U_r are average values for all frequencies within a band with a lag r.

At lag 0, the period becomes infinite. Thus, the spectral estimate includes all the record variance that does not recur during the length of the record used in the analysis. Therefore, it includes any random fluctuations and linear trends in the record.

The longest period other than zero frequency period is determined by the number of lags used in the computation. It is generally recommended that the number of lags be no greater than 15% of the total number of points in the record.

Determination of sampling frequency

The shortest period that is theoretically possible to resolve with a given sampling interval is one which is twice as large as the sampling interval. In practice, the sampling interval should be equal or less than one-third the length of the shortest period which we want to resolve; for example, if we want to resolve a 24-hour period, sampling intervals of 8 hours are required. In mathematical form, the highest frequency which can be resolved from a discrete record with sampling interval Δt is

$$f_{max} = \frac{1}{2\Delta t} \qquad \text{Eq.(B-19)}$$

Example:

The wastewater influent for the city of Racine, Wisconsin, was sampled hourly in the summer of 1974 and analyzed for TOC. The record is shown in Figure 5. The mean and variance were calculated to be 70.56 mg/L and 1262.07 mg^2/L^2, respectively.

The power spectrum corresponding to the record of Figure 5 is obtained as depicted in Figure 6. This power spectrum exhibits a significant peak at the 1/24 hour frequency and a less significant peak at 1/8 hour. Since the last significant peak in the spectrum occurs at the 1/8 hour frequency, the sampling frequency should be at least two times the frequency of the last significant peak, i.e., 1/4 hour frequency. Therefore, a sampling interval less than 4 hours should be selected.

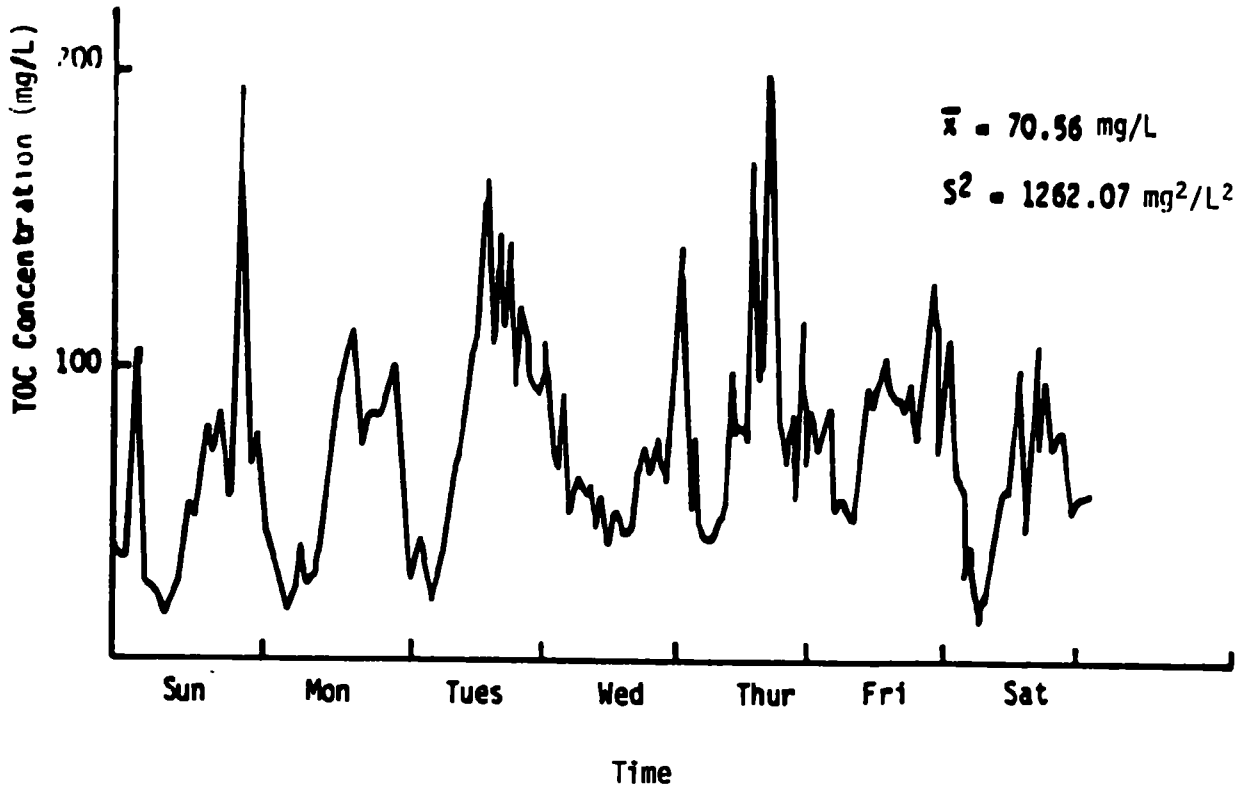

Figure B-2. Time record of TOC of municipal wastewater at Racine, Wisconsin.

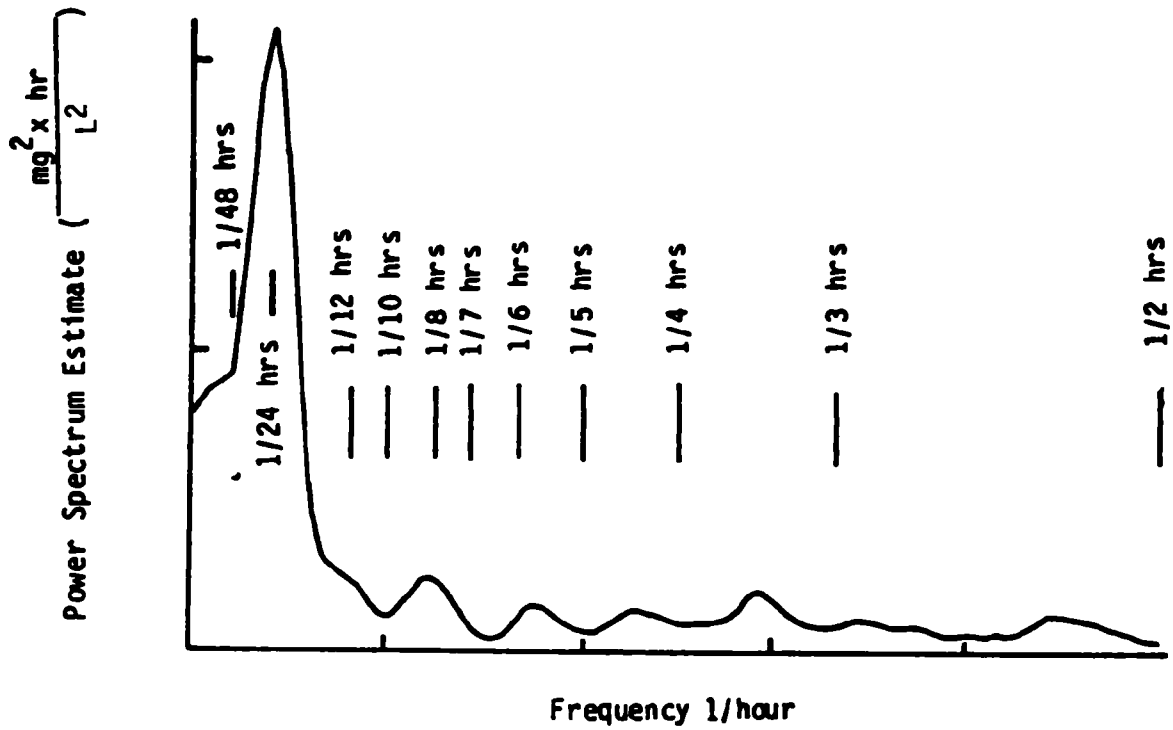

Figure B-3. Power spectrum of TOC concentration of municipal wastewater at Racine, Wisconsin.

B.12 COMPARING TWO MEANS

This program is to compare two means in order to determine if both means originate from the same population. For example, two different processes are compared to determine if any statistical difference exists. The comparison is a two-tailed test. The null hypothesis Ho: "means are equal" is against Ha: "means are not equal." If you want to determine whether process 1 is better than process 2, then the comparison is one-tailed. The null hypothesis Ho: "means are equal" is against alternative hypothesis Ha: "mean 1 is greater than mean 2."

Before the two means are compared, the two sample standard deviations must be compared by using F-test to determine whether they are significantly different or not. The equation to pool the sample standard deviations depends on it.

To use this program, the information required is as follows:

Group 1.

1. Two sample means
2. Number of samples from both sets of data
3. Sample standard deviations from both sets of data
4. Confidence level required for the comparison

In Group 2, the calculating processes are similar to the above example. However, the user must provide the population standard deviation and the normal deviate Z. The Z test statistic is calculated from the formula:

$$Z = \frac{\bar{y}_1 - \bar{y}_2}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}\ \sigma} \qquad \text{Eq.(B-21)}$$

B.13 PERCENTAGE AREA UNDER THE F DISTRIBUTION

The F statistic is the ratio of two estimates of variance. It is a two-parameter distribution, the degrees of freedom in two estimates of variance. It is used to test hypothesis concerning the treatment effects or significant difference between variabilities of two samples.

The calculation of the percentage area in this program is an integration from zero to the desired value of F. The necessary conditions to obtain the area are the degrees of freedoms for the two variances and the desired F value. For example, if the degrees of freedom for both variances are 12, respectively and the desired value of F is 3, then the percentage area in the F distribution is 96.567%.

B.14 F DISTRIBUTION

This program is to calculate the F value for any desired percentage area of the F distribution. It is the inverse of the program in (B.13).

B.15 SIGNIFICANT TEST BETWEEN VARIABILITIES OF TWO SAMPLES

The objective is to test the difference in variability between two

samples. For example, new equipment is used to measure a compound and it is expected that the measurement uniformity would improve (less variance or more precise). The question to ask is whether the improvement really exists or has that occurred by chance. To be sure of a significant improvement in variability, a ratio (F-ratio) of two variabilities before and after must be calculated, and compared with the F value at a 95% confidence level and degrees of freedom for both sets of data. If a real improvement does exist, it would be necessary for the calculated F value to exceed the F value at a 95% confidence level. Then it can be reported that an improvement in variability exists with a 95% chance of being correct. The null hypothesis Ho: "variabilities are equal" is against Ha: "variability before is greater than variability after." This is a one-tailed test. However, if the null hypothesis Ho: "variabilities are equal" is against Ha: "variabilities are not equal," then this is a two-tailed test.

To use this program, the user must provide the sample standard deviations and number of samples before and after. The confidence level and one- or two-tailed are also needed.

For example, given $\overline{y}_1 = 79.1$, $\overline{y}_2 = 76.2$, $n_1 = 7$, $n_2 = 5$, $s_1 = 5.10$, $s_2 = 3.33$, determine whether a significant difference between the variabilities does exist.

Solution:

- First, calculate the ratio of the two variances s_1^2/s_2^2

$$\frac{s_1^2}{s_2^2} = 2.35 \qquad \text{Eq.(B-22)}$$

- Second, calculate the F value at a 95% confidence level of 6 amd 4 degrees of freedom.

$$F\ (6,\ 4,\ 0.95) = 6.16 \qquad \text{Eq.(B-23)}$$

- Third, compare s_1^2/s_1^2 and F (6, 4, 0.95).

  Since the ratio of the two variances is less than the value of F (6,4,0.95) it is correct to assume that the variances are not significantly different.

B.16 SIGNIFICANT TEST BETWEEN THE POPULATION VARIABILITY AND THE SAMPLE VARIABILITY

The objective is to test the difference between the sample variability and the population variability. For example, does a lower value of sample variability from a new measurement mean that it is now more uniform than the past population variance? To answer this question, a chi square test using the following formula must be utilized:

$$\text{chi square} = \frac{\beta\ s^2}{\sigma^2} \qquad \text{Eq.(B-24)}$$

where:

β = degrees of freedom, (n-1)

s^2 = sample variance

σ^2 = population variance

If the calculated chi square value from equation (B-24) is larger than that at an upper 5% chi square with (n-1) degrees of freedom, then the sample variability is significantly greater than the population variability. If the calculated chi square value is in between the upper 5 and 95% chi square values, then the sample variability is not significantly larger or smaller than the population variability. On the other hand, if the calculated chi square value is smaller than that at an upper 95% level, then the sample variability is significantly smaller than the population variability.

APPENDIX C

NOMENCLATURE

μ	Population Mean
σ	Standard Deviation
β	Degrees of Freedom
χ^2_β	Chi-Square with Degrees of Freedom β
α	A Significance Level
y_i	Observations at i=1, 2, 3, ...
$\bar{y}$	Sample Mean
s	Sample Standard Deviation
n	Number of Samples
cc	Coefficient of Correlation
Z	Normal Deviate
t	Student t
Exp	Exponential Function
Γ	Gamma Function