NASA TECHNICAL NOTE

NASA TN D-5275

c. 1

NASA TN D-5275

A SIMPLE TECHNIQUE FOR AUTOMATIC COMPUTER EDITING OF BIODATA

by

Ram Swaroop and Kenneth A. West

Computing and Software, Inc.
Field Team at Flight Research Center

and

Charles E. Lewis, Jr.

Flight Research Center
Edwards, Calif.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • JUNE 1969

A SIMPLE TECHNIQUE FOR AUTOMATIC COMPUTER EDITING OF BIODATA

By Ram Swaroop and Kenneth A. West

Computing and Software, Inc.
Field Team at Flight Research Center

and

Charles E. Lewis, Jr.

Flight Research Center
Edwards, Calif.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 – CFSTI price $3.00

A SIMPLE TECHNIQUE FOR AUTOMATIC COMPUTER EDITING OF BIODATA

By Ram Swaroop and Kenneth A. West
Computing and Software, Inc.
Field Team at Flight Research Center

and

Charles E. Lewis, Jr.
Flight Research Center

SUMMARY

Before any data are statistically analyzed, it is always necessary to edit the data to some extent. Furthermore, when large quantities of data are collected, the editing must be performed by automatic means. One common task in the editing process is the identification of observations which deviate markedly from the rest of the sample, commonly known as outliers. A simple statistical technique for identifying the outliers and the necessary computer program is presented in this report. The program requires as input only the data set, sample size, and preselected levels of significance at which outliers are to be identified. It is assumed that the data set is a random sample of size larger than two from a normal population. Two examples are presented to illustrate applications of the described technique.

INTRODUCTION

The NASA Flight Research Center is engaged in an extensive research and development program aimed at advancing the state of the art in medical monitoring of humans in flight (ref. 1). Under this program, more biomedical information is collected in flight than is collected under the sum of all other known flight programs. An effort of this magnitude depends entirely on the Flight Research Center's capacity for collecting, reducing, and analyzing these data by automatic means, including development of new techniques for accomplishing this work. No matter how sophisticated the monitoring, collection, and reduction systems, some editing of the biodata is required before they can be analyzed statistically.

The reduced biodata may contain observations that deviate markedly from the rest of the sample, or from the trend of the data set. Such measurements are termed outlying observations, or outliers. An outlier may be subject to errors other than the usual random fluctuations characterizing the population to which the data belong, or may merely occur too infrequently to be considered in a particular analysis. Most of the statistical tests which are available to detect and decide whether an outlier is too rare to be acceptable are either too complex or too cumbersome for general application. One method suggested originally by Cramér (ref. 2) is derived in this paper and its use demonstrated. This method was chosen because of its simplicity and easy applicability in editing biodata. A computer program for automatic editing was written in FORTRAN IV, based on Cramér's suggestion. The input to the program is sample data, sample size, and a preselected level of significance. The data sample is either a set of observations or the deviations of the observations from a model, whichever is appropriate. Before testing the outliers, the program computes and prints the mean and standard deviation. After the test is performed, the data are printed with each outlier identified by an asterisk. Then the mean, standard deviation, and sample size are recalculated and printed, excluding the outliers. Four repetitions of the test may be made for different levels of significance.

Two examples are presented. In the first example the observations meet the assumption of random sample from a normal population. In the second example the input data are deviations of the observations from a given model.

The program source listing, with instructions, and sample problems are presented in appendixes A to C. For convenience in manual computations, a related probability table of values (table I) is included.

SYMBOLS

$F(x)$ distribution function at x

$f_{\tau}(x)$ density function of τ at point x

$\sum_{i}^{k}$ summation starting from i through k, where i and k are integers between 1 and n, and i is less than k

n sample size

$s = \sqrt{\frac{(n-1)}{n}}$ (standard deviation of sample)

t random variable of Student's distribution

t' random variable related to t distribution

$\bar{x}$ arithmetic average of sample values

$x_1, x_2, \ldots, x_n$ elements of observed sample

$y_1, y_2, \ldots, y_n$	orthogonalized variables obtained from $x_1, x_2, \ldots, x_n$
α	level of significance between 0 and 1
$\Gamma(r)$	gamma function at r. If r is an integer $\Gamma(r) = (r - 1)!$
σ	standard deviation
τ	random variable obtained from Student's t distribution
τ^*	value of τ at α level of significance

BRIEF TEST DESCRIPTION

The outlier test consists of computing a parameter τ^* for the data set and a value τ_i for each member in the data set, comparing each τ_i with the τ^*, and identifying the members with a τ_i greater than τ^*. The set parameter τ^* is a function of the sample and the level of significance at which outliers are to be identified. Each τ_i is a function of the value of the member and the mean and standard deviation of the data set. As previously stated, the members must constitute a random sample of size larger than two from a normal population.

DERIVATION OF TEST

Let $x_1, x_2, \ldots, x_n$ be a random sample of size n from a normal distribution with mean 0 and standard deviation σ. Then the variable

$$t = \frac{x_1}{\sqrt{\frac{1}{n-1} \sum_{2}^{n} x_i^2}}$$

has a Student's central t distribution, but the variable

$$\tau = \frac{x_1}{\sqrt{\frac{1}{n} \sum_{1}^{n} x_i^2}}$$

does not because here the numerator and denominator are not independent. Moreover, $\tau^2 \leqq n$; therefore, the distribution of τ has a zero probability outside the interval $(-\sqrt{n}, \sqrt{n})$. Defining

$$t' = \sqrt{\frac{n-1}{n}} \frac{\tau}{\sqrt{1 - \frac{\tau^2}{n}}} = \frac{\tau\sqrt{n-1}}{\sqrt{n - \tau^2}} = \frac{x_1}{\sqrt{\frac{1}{n-1}\sum_2^n x_i^2}}$$

it is seen that t' is distributed as a central t with $(n - 1)$ degrees of freedom. From this the density function of τ is obtained as

$$f_\tau(x) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})} \left(1 - \frac{x^2}{n}\right)^{\frac{(n-3)}{2}} \qquad \text{for } |x| \leqq \sqrt{n}$$

for $n > 2$. Replacing the x_i variables with new y_i variables by means of an orthogonal transformation such that the first two are

$$y_1 = \sqrt{n}\,\bar{x}$$

and

$$y_2 = \sqrt{\frac{n}{n-1}}(x_1 - \bar{x})$$

it is found that

$$ns^2 = \sum_1^n x_i^2 - n\bar{x}^2 = \sum_2^n y_i^2$$

Consequently, the variable

$$\tau_1 = \frac{x_1 - \bar{x}}{s}$$

which expresses the deviation of the sample value x_1 from the sample mean $\bar{x}$ in terms of measured units of the standard deviation of the sample, becomes

$$\tau_1 = \frac{y_2}{\sqrt{\frac{1}{n-1}\sum_2^n y_i^2}}$$

The variables y_i, $i = 2 \ldots, n$ are independent and normally distributed with 0 mean and standard deviation σ. The variable

$$\frac{\tau_1 \sqrt{n - 2}}{\sqrt{n - 1 - \tau_1^2}}$$

is then distributed as a Student's t with (n - 2) degrees of freedom. These results hold irrespective of the value of the mean, and for any relative deviation

$$\frac{x_i - \overline{x}}{s}$$

PROGRAM APPLICATIONS

Given a sample of data $x_1, \ldots x_n$, compute the sample estimate of the mean

$$\overline{x} = \frac{1}{n}\sum_1^n x_i$$

and

$$s = \sqrt{\frac{\sum_1^n x_i^2 - n\overline{x}^2}{n}}$$

Then compute critical values of τ^* by the relation

$$\tau^* = \frac{t\sqrt{n - 1}}{\sqrt{n - 2 + t^2}}$$

where t is the Student's central t-value at α level with (n - 2) degrees of freedom and n is the sample size. Next compute τ_i values for n sample points

$$\tau_i = \frac{x_i - \overline{x}}{s}$$

If

$$|\tau_i| > \tau^*$$

then x_i is considered an outlier at significance level α.

The program (appendix A) follows this method to detect and identify (by *) the outliers. An option of the program allows the user to repeat the test for different levels of significance. The required input parameters are as follows:

1. Format of the data to be read.

2. The sample size n.

3. The significance level α values.

4. The data format is as specified.

The input procedure is completely described in the comments at the beginning of the source listing (appendix A).

EXAMPLES

Example 1

One-minute heart rates from a 71-minute flight piloted by a student pilot from the Aerospace Research Pilot School at Edwards Air Force Base are used to illustrate the described method of editing for outliers. The heart rates in this example satisfy the assumption of random sampling from a normal population. The sample output, including the outliers marked by asterisks and the standard deviations, is shown in appendix B. The results of the analysis are shown in the following table:

Significance level, percent	10	5	1	0.1	Original data
Reduced sample size	61	67	70	71	71
Standard deviation of reduced sample	4.593	5.711	6.371	6.749	6.749

Example 2

In this example the outlier editing program is used on a flight profile from the Limited Vision Landing Accuracy Study (commonly known as "Cyclops") in progress at the Flight Research Center. The observations are the altitudes of the aircraft at fixed distances from the touchdown target point on the runway. The flight profile model is assumed to be a second-degree polynomial. The deviations of the observations from this flight profile are the inputs to this program. These deviations satisfy the requirements of random sampling from a normal population. The computer output identifying outliers with asterisks is presented in appendix C. The flight profile and outliers are shown in figure 1.

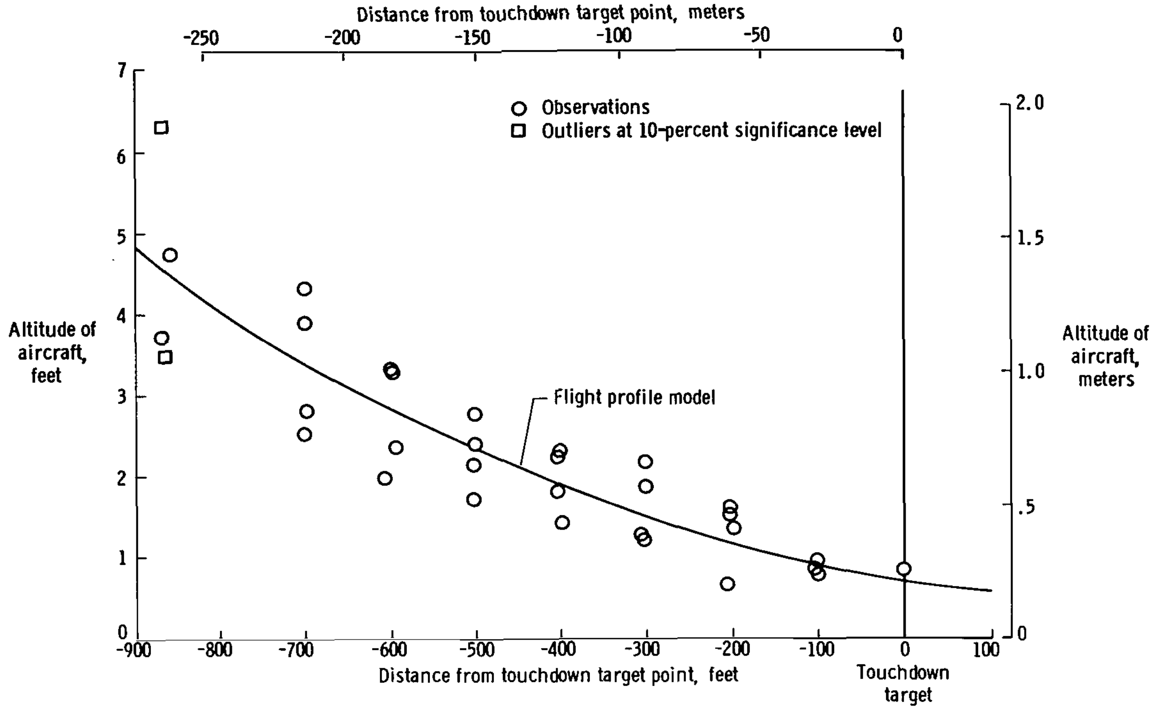


Figure 1.– Observations of aircraft altitude variation with distance from the runway touchdown target point for the Limited Vision Landing Accuracy Study.

The results of the analysis at four levels of significance are presented in the following table:

Significance level, percent	10	5	1	0.1	Original data
Reduced sample size	33	34	34	34	35
Standard deviation of reduced sample	0.453	0.475	0.475	0.475	0.564

REMARKS

Although the method of detecting and identifying the outliers described in this paper is being used for biodata editing at the Flight Research Center, its applicability in other areas is obvious. Results indicate that statistical analyses after removal of outliers have a lower standard deviation than when the outliers are not removed. The sample sizes are usually larger than 30, so reduction in sample size did not affect the validity of the results.

Flight Research Center,
National Aeronautics and Space Administration,
Edwards, Calif., March 25, 1969,
127-51-03-01-24.

APPENDIX A

PROGRAM SOURCE LISTING

```
C      OUTLIE
C
C      PURPOSE
C          TEST EXTREME VALUES OF AN ARRAY
C
C      METHOD
C          LET  X(1) THRU X(NFIN)  BE DISTRIBUTED N(0,5) THEN
C          T = X / SQRT(1/N* SUM(X**2)
C                    WHERE  SUM EXCLUDES VALUE X IN THE NUMERATOR
C          N = SAMPLE SIZE
C
C          TAU = (T*SQRT(N-1))/SQRT(N-2+ T**2)
C          TO TEST OUTLIERS CALCULATE  TAU(OBS) =  (X-U)/S
C              WHERE
C                  S**2 = 1/N*SUM((X-U)**2)
C                  U    =  MEAN
C          IF  ABS(TAU(OBS)) >  TAU ,    X IS AN OUTLIER
C
C      REFERENCE
C          MATHEMATICAL METHODS OF STATISTICS, A. CRAMER 1961
C
C      DESCRIPTION OF PARAMETERS
C          X = ARRAY ASSUMED TO B N(U,S)
C          NLIM = NUMBER OF X VALUES
C          T = STUDENT'S T-VALUE AT ALPHA LEVEL AND N-2 DEGREES
C              OF FREEDOM
C
C      INPUT
C
C      CARD  1    FORMAT OF X ARRAY  READ  IN   (20 A4)
C
C      CARD  2
C
C          COL  1 THRU 5    NLIM = NO. OF X VALUES
C          COL 10           NT = NO. OF ALPHA VALUES
C          COL 11 THRU 15 ALPHA 1
C          COL 16 THRU 20 ALPHA 2
C          COL 21 THRU 25 ALPHA 3
C          COL 26 THRU 30 ALPHA 4
C                          LIMIT OF 4 ALPHA VALUES WHICH MUST BE
C                          EQUAL TO  0.10, 0.05, 0.01, OR 0.001  .
C                          OTHER VALUES WILL NOT BE ACCEPTED BY THE
C                          PROGRAM
C
C          CARD 3 THRU  ...
C              DATA FORMATED AS PRESCRIBED IN CARD 1
C
C
```

```
C       OUTPUT
C           1 MEAN
C           2 STANDARD DEVIATION
C           3 LIST OF DATA WITH THE OUTLIERS IDENTIFIED BY AN ASTERISK
C           4 MEAN WITH OUTLIERS DELETED
C           5 STANDARD DEVIATION WITH OUTLIERS DELETED
C           6 SAMPLE SIZE WITHOUT THE OUTLIERS
C
C           NOTE  STEPS  3 THRU 6 ARE REPEATED FOR EACH LEVEL OF ALPHA
C                 USED.
C
C
       DIMENSION X(1000),ITC(1000), ALPHA(5), FMT(20), TVAL(5)
      REAL*8 SUM1, SUM2, SUMNEW, SUMN2
      DATA IDUM1/4H    /,IDUM2/4H*   /
    1 SUM1 = 0.0
      SUM2 = 0.0
       READ(1,100,END=999) FMT, NLIM, NT,(ALPHA(J),J=1,NT)
  100 FORMAT(20A4/2I5,4F5.3)
      WRITE(3,200) FMT
  200 FORMAT(1H1,30X,'P R O G R A M    T O    T E S T    O U T L I E R S
     1'///' FORMAT OF INPUT X VALUES IS '20A4)
      READ(1,FMT)(X(I),I=1,NLIM )
      DO 20 I=1,NLIM
      SUM1 = SUM1 + X(I)
   20 SUM2 = SUM2 + X(I)*X(I)
      SDEV =  (SUM2 -(SUM1*SUM1)/NLIM)/(NLIM)
      SDEV = SQRT(SDEV)
      XMEAN = SUM1/NLIM
       ALIM = NLIM
       WRITE(3,160) XMEAN, SDEV
  160 FORMAT(1H ,'MEAN                  ='F10.4/' STANDARD DEVIATION =',
     1F10.4)
      DO 30 J=1,NT
       CALL STUDNT(TVAL(J),NLIM, ALPHA(J))
   30 TVAL(J) = (TVAL(J)*SQRT(ALIM-1.))/SQRT(ALIM-2. +TVAL(J)*TVAL(J))
      DO 70 J = 1, NT
      WRITE(3,130)
  130 FORMAT(1H1,'LIST OF INPUT VALUES'//9X,'N',14X,'X VALUE',9X,'N',13X
     A    ,'X VALUE'/)
      SUMNEW = SUM1
       SUMN2 = SUM2
       NCOUNT  = NLIM
      DO 60 I= 1, NLIM
  150 FORMAT(1H1)
       ITC(I)= IDUM1
      TAU = (X(I) - XMEAN)/SDEV
       IF(ABS(TAU).LE.TVAL(J)) GO TO 90
       ITC(I) = IDUM2
      SUMNEW = SUMNEW-X(I)
      SUMN2  = SUMN2 -X(I)*X(I)
      NCOUNT = NCOUNT -1
   90 CONTINUE
   60 CONTINUE
       WRITE(3,110)(I,X(I),ITC(I),I=1,NLIM)
  110 FORMAT(1X,I10,F20.5,A1,I9,F20.5,A1)
      WRITE(3,120) ALPHA(J)
```

```
120 FORMAT(1H0, '* INDICATES OUTLIER VALUE AT 'F5.3,' SIGNIFICANCE LEV
   1EL')
    EXPT = SUMNEW/NCOUNT
    SDEV1= (SUMN2-(SUMNEW*SUMNEW)/NCOUNT)/(NCOUNT)
    SDEV1= SQRT(SDEV1)
    WRITE(3,170) EXPT, SDEV1,NCOUNT
170 FORMAT('0POPULATION PARAMETERS WITH OUTLIERS DELETED'/' MEAN
   1           ='F10.4/' STANDARD DEVIATION ='F10.4/' N
   2 ='I5///)
 70 CONTINUE
    GO TO 1
999 RETURN
    END
    SUBROUTINE STUDNT(T,N,A)
    DIMENSION            TABLE(4, 34),LTAB( 4),ATAB(4)
    DATA  LTAB / 30,40,60,120 /,ATAB/0.10,0.05,0.01,0.001/
    DATA TABLE/6.314,12.706,63.657,636.619,2.920,4.303,9.925,31.598,
   12.353,3.182,5.841,12.924,2.132,2.776,4.604,8.610,2.015,2.571,4.032
   2,6.869,1.943,2.447,3.707,5.959,1.895,2.365,3.499,5.408,1.860,2.306
   3,3.355,5.041,1.833,2.262,3.250,4.781,1.812,2.228,3.169,4.587,1.796
   4,2.201,3.106,4.437,1.782,2.179,3.055,4.318,1.771,2.160,3.012,4.221
   5,1.761,2.145,2.977,4.140,1.753,2.131,2.947,4.073,1.746,2.120,2.921
   6,4.015,1.740,2.110,2.898,3.965,1.734,2.101,2.878,3.922,1.729,2.093
   7,2.861,3.883,1.725,2.086,2.845,3.850,1.721,2.080,2.831,3.819,1.717
   8,2.074,2.819,3.792,1.714,2.069,2.807,3.767,1.711,2.064,2.797,3.745
   9,1.708,2.060,2.787,3.725,1.706,2.056,2.779,3.707,1.703,2.052,2.771
   A,3.690,1.701,2.048,2.763,3.674,1.699,2.045,2.756,3.659,1.697,2.042
   1,2.750,3.646,1.684,2.021,2.704,3.551,1.671,2.000,2.660,3.460,1.658
   2,1.980,2.617,3.373,1.645,1.960,2.576,3.291/
    DO 10 I=1,4
    IF(A    .EQ.ATAB(I)) GO TO 20
 10  CONTINUE
    WRITE(3,1) A
  1 FORMAT(1H0,'ALPHA VALUE = 'F5.4,' IS NOT A TABLED VALUE' )
    WRITE(3,167)
167 FORMAT(1H ,'                         ALPHA SET EQUAL TO 0.10')
     I = 1
 20 I1= I
     IF( N.LE.30) GO TO 40
     DO 30 I  = 2,4
    IF(N-LTAB(I))50,40,30
 30 CONTINUE
C     INTERPOLATE WITH DF 120, INFIN
    T= TABLE(I1,34) +  (1.0/N)/(1.0/LTAB(4))*(TABLE(I1,33)-TABLE(I1,34
   1) )
    RETURN
 40 T= TABLE(I1,N)
    RETURN
 50 T= TABLE(I1,I+29) + (1.0/N -1.0/LTAB(I))/(1.0/LTAB(I-1)-1.0/LTAB(
   1 I))*(TABLE(I1,I+28) - TABLE(I1,I+29))
    RETURN
    END
```

APPENDIX B

OUTPUT FOR EXAMPLE 1

```
    P R O G R A M   T O   T E S T   O U T L I E R S
FORMAT OF INPUT X VALUES IS (14F5.3)
MEAN                =   98.5042
STANDARD DEVIATION  =    6.7494
LIST OF INPUT VALUES
```

N	X VALUE	N	X VALUE
1	101.20000	2	95.20000
3	85.79999*	4	82.20000*
5	96.70000	6	95.29999
7	98.29999	8	101.59999
9	94.89999	10	90.79999
11	92.20000	12	97.00000
13	95.29999	14	94.79999
15	104.20000	16	118.20000*
17	104.29999	18	110.50000*
19	107.50000	20	104.70000
21	104.29999	22	103.39999
23	104.70000	24	101.59999
25	102.20000	26	105.39999
27	103.50000	28	98.00000
29	105.00000	30	103.39999
31	104.00000	32	97.89999
33	100.39999	34	99.29999
35	100.89999	36	98.79999
37	99.89999	38	106.39999
39	96.29999	40	102.70000
41	105.09999	42	110.29999*
43	101.89999	44	96.89999
45	96.50000	46	96.29999
47	95.70000	48	98.00000
49	90.29999	50	88.89999
51	97.29999	52	86.20000*
53	86.50000*	54	96.29999
55	97.09999	56	96.50000
57	89.00000	58	84.70000*
59	97.00000	60	88.59999
61	97.79999	62	95.20000
63	98.39999	64	112.79999*
65	111.09999*	66	97.00000
67	95.50000	68	97.50000
69	93.00000	70	94.59999
71	93.00000		

```
* INDICATES OUTLIER VALUE AT 0.100 SIGNIFICANCE LEVEL

POPULATION PARAMETERS WITH OUTLIERS DELETED
MEAN                =   98.4508
STANDARD DEVIATION  =    4.5932
N                   =   61
```

LIST OF INPUT VALUES

N	X VALUE	N	X VALUE
1	101.20000	2	95.20000
3	85.79999	4	82.20000*
5	96.70000	6	95.29999
7	98.29999	8	101.59999
9	94.89999	10	90.79999
11	92.20000	12	97.00000
13	95.29999	14	94.79999
15	104.20000	16	118.20000*
17	104.29999	18	110.50000
19	107.50000	20	104.70000
21	104.29999	22	103.39999
23	104.70000	24	101.59999
25	102.20000	26	105.39999
27	103.50000	28	98.00000
29	105.00000	30	103.39999
31	104.00000	32	97.89999
33	100.39999	34	99.29999
35	100.89999	36	98.79999
37	99.89999	38	106.39999
39	96.29999	40	102.70000
41	105.09999	42	110.29999
43	101.89999	44	96.89999
45	96.50000	46	96.29999
47	95.70000	48	98.00000
49	90.29999	50	88.89999
51	97.29999	52	86.20000
53	86.50000	54	96.29999
55	97.09999	56	96.50000
57	89.00000	58	84.70000*
59	97.00000	60	88.59999
61	97.79999	62	95.20000
63	98.39999	64	112.79999*
65	111.09999	66	97.00000
67	95.50000	68	97.50000
69	93.00000	70	94.59999
71	93.00000		

* INDICATES OUTLIER VALUE AT 0.050 SIGNIFICANCE LEVEL

POPULATION PARAMETERS WITH OUTLIERS DELETED
MEAN = 98.4463
STANDARD DEVIATION = 5.7113
N = 67

LIST OF INPUT VALUES

N	X VALUE	N	X VALUE
1	101.20000	2	95.20000
3	85.79999	4	82.20000
5	96.70000	6	95.29999
7	98.29999	8	101.59999
9	94.89999	10	90.79999
11	92.20000	12	97.00000
13	95.29999	14	94.79999
15	104.20000	16	118.20000*
17	104.29999	18	110.50000
19	107.50000	20	104.70000
21	104.29999	22	103.39999
23	104.70000	24	101.59999
25	102.20000	26	105.39999
27	103.50000	28	98.00000
29	105.00000	30	103.39999
31	104.00000	32	97.89999
33	100.39999	34	99.29999
35	100.89999	36	98.79999
37	99.89999	38	106.39999
39	96.29999	40	102.70000
41	105.09999	42	110.29999
43	101.89999	44	96.89999
45	96.50000	46	96.29999
47	95.70000	48	98.00000
49	90.29999	50	88.89999
51	97.29999	52	86.20000
53	86.50000	54	96.29999
55	97.09999	56	96.50000
57	89.00000	58	84.70000
59	97.00000	60	88.59999
61	97.79999	62	95.20000
63	98.39999	64	112.79999
65	111.09999	66	97.00000
67	95.50000	68	97.50000
69	93.00000	70	94.59999
71	93.00000		

* INDICATES OUTLIER VALUE AT 0.010 SIGNIFICANCE LEVEL

POPULATION PARAMETERS WITH OUTLIERS DELETED
MEAN = 98.2228
STANDARD DEVIATION = 6.3706
N = 70

LIST OF INPUT VALUES

N	X VALUE	N	X VALUE
1	101.20000	2	95.20000
3	85.79999	4	82.20000
5	96.70000	6	95.29999
7	98.29999	8	101.59999
9	94.89999	10	90.79999
11	92.20000	12	97.00000
13	95.29999	14	94.79999
15	104.20000	16	118.20000
17	104.29999	18	110.50000
19	107.50000	20	104.70000
21	104.29999	22	103.39999
23	104.70000	24	101.59999
25	102.20000	26	105.39999
27	103.50000	28	98.00000
29	105.00000	30	103.39999
31	104.00000	32	97.89999
33	100.39999	34	99.29999
35	100.89999	36	98.79999
37	99.89999	38	106.39999
39	96.29999	40	102.70000
41	105.09999	42	110.29999
43	101.89999	44	96.89999
45	96.50000	46	96.29999
47	95.70000	48	98.00000
49	90.29999	50	88.89999
51	97.29999	52	86.20000
53	86.50000	54	96.29999
55	97.09999	56	96.50000
57	89.00000	58	84.70000
59	97.00000	60	88.59999
61	97.79999	62	95.20000
63	98.39999	64	112.79999
65	111.09999	66	97.00000
67	95.50000	68	97.50000
69	93.00000	70	94.59999
71	93.00000		

* INDICATES OUTLIER VALUE AT 0.001 SIGNIFICANCE LEVEL

POPULATION PARAMETERS WITH OUTLIERS DELETED
MEAN = 98.5042
STANDARD DEVIATION = 6.7494
N = 71

APPENDIX C

OUTPUT FOR EXAMPLE 2

```
    P R O G R A M    T O    T E S T    O U T L I E R S

FORMAT OF INPUT X VALUES IS    (14F5.3)
MEAN                 =    -0.0002
STANDARD DEVIATION =     0.5636
LIST OF INPUT VALUES
```

N	X VALUE	N	X VALUE
1	0.25000	2	0.47400
3	0.47000	4	0.39700
5	0.38100	6	0.34300
7	0.18700	8	-0.13300
9	-0.77000	10	-0.58000
11	-0.46500	12	-0.24300
13	-0.10700	14	-0.29200
15	-0.53800	16	-0.06500
17	0.06800	18	-0.31000
19	1.82600*	20	0.90500
21	0.45000	22	0.01300
23	0.33600	24	0.65400
25	0.41800	26	0.08000
27	0.14000	28	-0.98900*
29	-0.88500	30	-0.88500
31	-0.65300	32	-0.46500
33	-0.24900	34	0.29700
35	-0.06600		

```
* INDICATES OUTLIER VALUE AT 0.100 SIGNIFICANCE LEVEL

POPULATION PARAMETERS WITH OUTLIERS DELETED
MEAN                 =    -0.0255
STANDARD DEVIATION =     0.4533
N                    =    33
```

LIST OF INPUT VALUES

N	X VALUE	N	X VALUE
1	0.25000	2	0.47400
3	0.47000	4	0.39700
5	0.38100	6	0.34300
7	0.18700	8	-0.13300
9	-0.77000	10	-0.58000
11	-0.46500	12	-0.24300
13	-0.10700	14	-0.29200
15	-0.53800	16	-0.06500
17	0.06800	18	-0.31000
19	1.82600*	20	0.90500
21	0.45000	22	0.01300
23	0.33600	24	0.65400
25	0.41800	26	0.08000
27	0.14000	28	-0.98900
29	-0.88500	30	-0.88500
31	-0.65300	32	-0.46500
33	-0.24900	34	0.29700
35	-0.06600		

* INDICATES OUTLIER VALUE AT 0.050 SIGNIFICANCE LEVEL

POPULATION PARAMETERS WITH OUTLIERS DELETED
MEAN = -0.0539
STANDARD DEVIATION = 0.4754
N = 34

LIST OF INPUT VALUES

N	X VALUE	N	X VALUE
1	0.25000	2	0.47400
3	0.47000	4	0.39700
5	0.38100	6	0.34300
7	0.18700	8	-0.13300
9	-0.77000	10	-0.58000
11	-0.46500	12	-0.24300
13	-0.10700	14	-0.29200
15	-0.53800	16	-0.06500
17	0.06800	18	-0.31000
19	1.82600*	20	0.90500
21	0.45000	22	0.01300
23	0.33600	24	0.65400
25	0.41800	26	0.08000
27	0.14000	28	-0.98900
29	-0.88500	30	-0.88500
31	-0.65300	32	-0.46500
33	-0.24900	34	0.29700
35	-0.06600		

* INDICATES OUTLIER VALUE AT 0.010 SIGNIFICANCE LEVEL

POPULATION PARAMETERS WITH OUTLIERS DELETED
MEAN = -0.0539
STANDARD DEVIATION = 0.4754
N = 34

LIST OF INPUT VALUES

N	X VALUE	N	X VALUE
1	0.25000	2	0.47400
3	0.47000	4	0.39700
5	0.38100	6	0.34300
7	0.18700	8	-0.13300
9	-0.77000	10	-0.58000
11	-0.46500	12	-0.24300
13	-0.10700	14	-0.29200
15	-0.53800	16	-0.06500
17	0.06800	18	-0.31000
19	1.82600*	20	0.90500
21	0.45000	22	0.01300
23	0.33600	24	0.65400
25	0.41800	26	0.08000
27	0.14000	28	-0.98900
29	-0.88500	30	-0.88500
31	-0.65300	32	-0.46500
33	-0.24900	34	0.29700
35	-0.06600		

* INDICATES OUTLIER VALUE AT 0.001 SIGNIFICANCE LEVEL

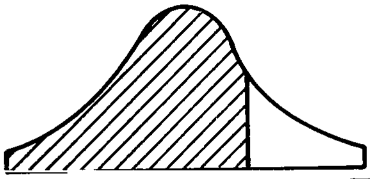
POPULATION PARAMETERS WITH OUTLIERS DELETED
MEAN = -0.0539
STANDARD DEVIATION = 0.4754
N = 34

REFERENCES

1. Roman, James: Long-Range Program to Develop Medical Monitoring in Flight. The Flight Research Program - I. Aerospace Medicine, vol. 36, no. 6, June 1965, pp. 514-518.

2. Cramér, Harald: Mathematical Methods of Statistics. Princeton University Press, 1961, pp. 240, 390.

TABLE I. – UPPER PERCENTAGE POINTS OF THE τ DISTRIBUTION



$$F(\tau^*) = \int_{-\sqrt{n}}^{\tau^*} \frac{1}{\sqrt{n\pi}} \frac{\Gamma\left(\frac{n}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \left(1 - \frac{x^2}{n}\right)^{\frac{n-3}{2}} dx = 1 - \alpha$$

n	τ^*		
	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.001$
3	1.410	1.414	1.414
4	1.645	1.715	1.730
5	1.757	1.918	1.982
6	1.814	2.051	2.178
7	1.848	2.142	2.329
8	1.870	2.207	2.447
9	1.885	2.256	2.540
10	1.896	2.294	2.616
12	1.910	2.348	2.730
14	1.920	2.385	2.812
16	1.926	2.411	2.873
18	1.931	2.421	2.921
20	1.934	2.447	2.959
25	1.940	2.474	3.026
30	1.944	2.492	3.071
40	1.948	2.514	3.127
60	1.952	2.535	3.182
120	1.956	2.555	3.237
∞	1.960	2.576	3.291

NOTE: To calculate τ^* values for sample sizes or significance levels not shown, use the relationship

$$\tau^* = \frac{t\sqrt{n-1}}{\sqrt{n-2+t^2}}$$

where t is Student's t-value at significance level α and $n - 2$ degrees of freedom and n is sample size.

"The aeronautical and space activities of the United States shall be conducted so as to contribute . . . to the expansion of human knowledge of phenomena in the atmosphere and space. The Administration shall provide for the widest practicable and appropriate dissemination of information concerning its activities and the results thereof."

—NATIONAL AERONAUTICS AND SPACE ACT OF 1958

NASA SCIENTIFIC AND TECHNICAL PUBLICATIONS

TECHNICAL REPORTS: Scientific and technical information considered important, complete, and a lasting contribution to existing knowledge.

TECHNICAL NOTES: Information less broad in scope but nevertheless of importance as a contribution to existing knowledge.

TECHNICAL MEMORANDUMS: Information receiving limited distribution because of preliminary data, security classification, or other reasons.

CONTRACTOR REPORTS: Scientific and technical information generated under a NASA contract or grant and considered an important contribution to existing knowledge.

TECHNICAL TRANSLATIONS: Information published in a foreign language considered to merit NASA distribution in English.

SPECIAL PUBLICATIONS: Information derived from or of value to NASA activities. Publications include conference proceedings, monographs, data compilations, handbooks, sourcebooks, and special bibliographies.

TECHNOLOGY UTILIZATION PUBLICATIONS: Information on technology used by NASA that may be of particular interest in commercial and other non-aerospace applications. Publications include Tech Briefs, Technology Utilization Reports and Notes, and Technology Surveys.

Details on the availability of these publications may be obtained from:

SCIENTIFIC AND TECHNICAL INFORMATION DIVISION

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Washington, D.C. 20546