# BUREAU OF MINERAL RESOURCES, GEOLOGY AND GEOPHYSICS

RECORD 1983/31

# RECORD

## TRIAV:

## A COMPUTER PROGRAM FOR AVERAGING SEVERAL TRIANGULAR DISTRIBUTIONS

by

A.L. HINDE

RECORD 1983/31

# TRIAV: A COMPUTER PROGRAM FOR AVERAGING SEVERAL TRIANGULAR DISTRIBUTIONS

by

A.L. HINDE

# CONTENTS

## SUMMARY

This record describes a FORTRAN program entitled TRIAV which can be used to average several triangular distributions. The program is available on the BMR's Hewlett-Packard computer and was developed to produce probabilistic estimates of undiscovered resources.

Input required for the program is as follows:

(i) the number of triangular distributions to be averaged;

(ii) a title; and

(iii) the individual triangular distributions specified by the minimum, most likely, and maximum values.

Output consists of two tabulations of values in terms of cumulative probability. One tabulation consists of 15 equally spaced points on the cumulative probability axis, the other consists of 15 equally spaced values. In addition, an approximate representation of the frequency histogram is plotted and provision has been made to allow the range to be reduced so that greater detail can be obtained in a smaller region of the average cumulative probability curve.

## INTRODUCTION

In the assessment of the undiscovered resources of a region, experts may provide a range of estimates of the potential. The consolidation of these opinions, into one estimate, is the aim of the subjective probability technique of resource assessment. One method of tackling this is to obtain from each expert an estimate of the undiscovered resources in the form of a triangular distribution and then to combine all these triangular distributions by averaging them. This record describes the program TRIAV, which was developed in BMR for this purpose.

A triangular distribution is a probability distribution characterised by a minimum value, a maximum value, and a most probable value. The average of several triangular distributions is not itself a triangular distribution, but a more complex distribution. The program TRIAV calculates and prints the cumulative probabilities for this average distribution at 15 equally spaced values as well as the values at 15 equally spaced cumulative probabilities.

If more than one region is involved in an assessment, the average distributions for each region may be added statistically. In this way the total resources of, for example all of Australia, may be estimated. This statistical addition is handled by program DELPHI, which has been documented elsewhere (E.L. Smith, 1981). For convenience, however, instructions for running program DELPHI are included in Appendix 1.

## THEORY

An example of a triangular distribution, for the random variable, x, is depicted in figure 1.

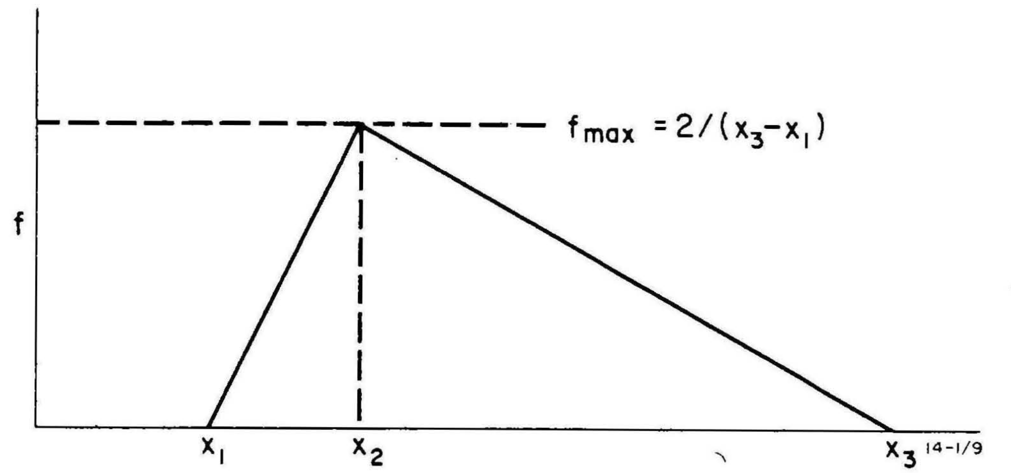



Figure 1. - An example triangular distribution.

The cumulative probability, at a point X, is given by:

$$
\begin{aligned}
p(x < X) &= 0,\ X < x_1 \\
&= \frac{(X-x_1)^2}{(x_2-x_1)}\frac{f_{max}}{2},\quad x_1 < X < x_2 \\
&= \left\{\frac{(X-x_3)}{(x_2-x_3)} + 1\right\}\frac{f_{max}}{2}(X-x_2) + \frac{(x_2-x_1)}{2} f_{max}, \\
&\qquad\qquad x_2 < X < x_3 \\
&= 1,\ x_3 < X
\end{aligned}
\tag{1}
$$

Now consider two overlapping triangular distributions as in figure 2. The cumulative probability of the combined (averaged) distribution at a value of X is the sum of the cumulative probabilities divided by two.

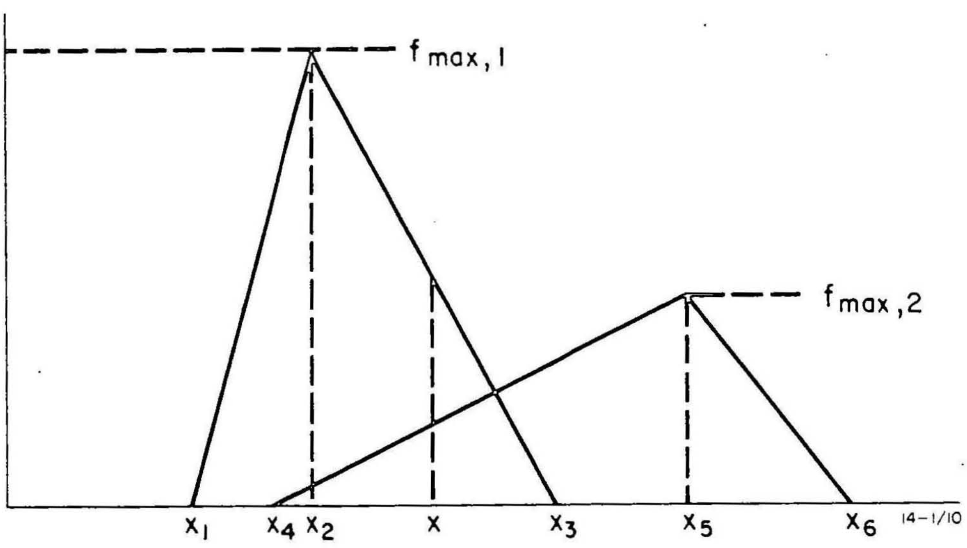



Figure 2. - Two overlapping triangular distributions.

For example, at point X, the cumulative probability is given by

$$
\begin{aligned}
p(x < X) = \tfrac{1}{2}\Big\{&\tfrac{1}{2}\left[(X-x_3)/(x_2-x_3) + 1\right] f_{max,1}(X-x_2) \\
&+ \tfrac{1}{2}(x_2-x_1)f_{max,1} + \tfrac{1}{2}(X-x_4)^2 f_{max,2}/(x_5-x_4)\Big\}
\end{aligned}
\tag{2}
$$

The entire cumulative curve will consist of a number of parabolic segments, the first between $x_1$ and $x_4$, the second between $x_4$ and $x_2$, and so on.

This theory is readily extended to the averaging of more than two triangular distributions. Note also that the triangular distributions can overlap, or not overlap, in any way.

## PROGRAM DESCRIPTION

The program reads in the values of $x_1$ (minimum value), $x_2$ (most likely value), and $x_3$ (maximum value) for each triangular distribution, and determines the range using the smallest value of $x_1$ and the largest value of $x_3$. The combined distribution is calculated in two ways. Firstly, the range is divided by 15 equally spaced points with one at each end point. At each point, the cumulative frequency for the combined distribution is calculated. The probability density function is also calculated: it is approximated by the slopes between the points. Hence, whereas the cumulative frequency is calculated at each of the 15 points, the probability density is approximated at the midpoints between the 15 points.

Secondly, the cumulative frequency range, (0 to 1), is divided by 15 equally spaced points and the values corresponding to these points calculated. A numerical method is used to determine these values.

The means and standard deviations of each input triangular distribution are calculated from the equations,

$$\text{mean} = (x_1 + x_2 + x_3)/3 \qquad (3)$$

$$\text{standard deviation} = \left[(x_1^2 + x_2^2 + x_3^2 - x_1x_2 - x_1x_3 - x_2x_3)/18\right]^{\frac{1}{2}} \qquad (4)$$

The mean of the combined distribution is calculated as the average of the individual means. In addition, the mean and standard deviation of the combined distribution are calculated numerically, using the 15-point cumulative curve. A comparison of the means obtained in these two ways gives an indication of how well the final distribution is represented by the 15-point cumulative curve.

The numerical method of calculating the mean of a distribution, represented by its cumulative curve, is given by:

$$E(x) \quad = \quad \max(x) \text{ - area under the cumulative curve} \qquad (5)$$

where $E(x)$ denotes the average or expected value of x,

$\max(x)$ is the upper end point of the range of x,

and area under the curve $= \frac{1}{2}\sum_{i=2}^{n}(x_i - x_{i-1})(F_i + F_{i-1})$

where $x_i$ = the x-value at the i'th point,

$F_i$ = the cumulative probability at the i'th point

and $n$ = the number of points (15 in the program)

The numerical method for calculating the variance (and hence the standard deviation) of a distribution in terms of its cumulative probability is given by,

$$var(x) = E(x^2) - E(x)^2,$$

where $E(x^2)$ is given by:

$$E(x^2) = \max(x)^2 - \sum_{i=2}^{n}\Big[F_{i-1}(x_i(x_{i-1}+x_i)-2x_{i-1}^2) + F_i(2x_i^2 - x_{i-1}(x_{i-1}+x_i))\Big] /3 \qquad (6)$$

These results can be derived by integrating by parts, the equations for $E(x)$ and $E(x^2)$, to get them in terms of the cumulative function, F; then integrating over each straight-line segment of the cumulative curve.

An allowance has been made in the program for reducing the range of the averaged distribution in order to get greater detail. This is especially useful when, for example, one of the triangular distributions is very spread out, while the others are clustered into a very small part of the range. This smaller part can be expanded by specifying a smaller range. If the range is reduced, however, it is not possible to calculate the mean and standard deviation in the numerical manner, as the range is not individual triangular distributions, is printed.

## INPUT DATA

In order to run the program, a file called 'DATTR', with the data described below, needs to be prepared.

| Line | Cols | Variable | Description | Format for entire line |
|---|---|---|---|---|
| 1 | 1-5 | N | Number of triangular distributions to be averaged (maximum 25) | (I5,2F10.0) |
| | 6-15 | XMINI* | Optional value specifying the lower end of the range (see note 1) | |
| | 16-25 | XMAXI* | Optional value specifying the upper end of the range (see note 1) | |
| 2 | 1-80 | ITITLE | Descriptive title (e.g. deposit). | (40A2) |
| 3 | 1-10 | X1* | Minimum value of triangular variable | (3F10.0) |
| | 11-20 | X2* | Most probable value of triangular variable | |
| | 21-30 | X3* | Maximum value of triangular variable | |

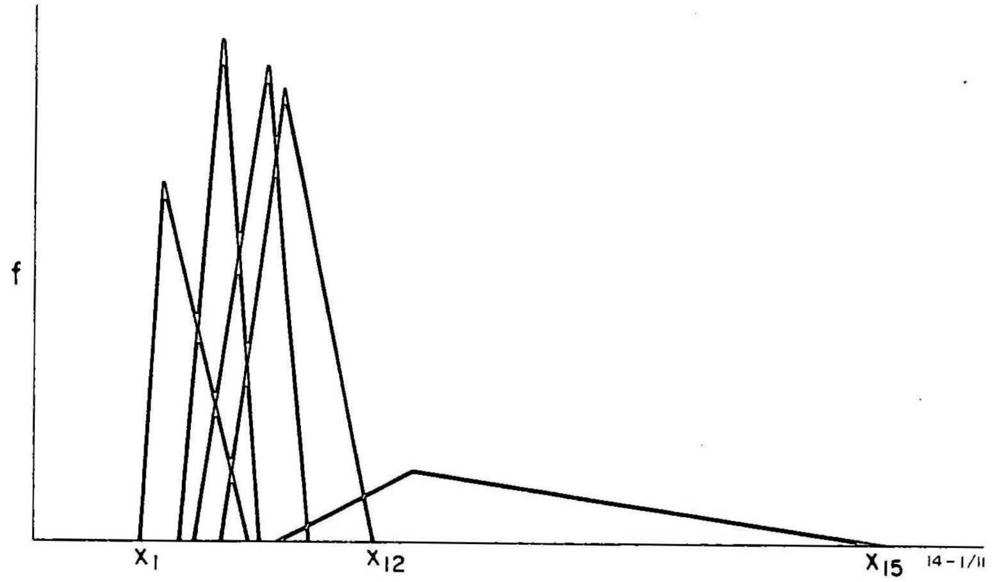
Card 3 is repeated N times.

The entire sequence, lines 1 to 3, may be repeated any number of times as the program does not stop until the end-of-file has been reached.

Note 1. If both XMINI and XMAXI are not specified, the range of the averaged distribution is determined using the smallest minimum value and the largest maximum value of the triangular distributions. This range can be made narrower by providing values for XMINI and XMAXI. This is useful for obtaining greater detail over selected portions of the range. An example where this is advisable is shown in Figure 3 below. Several triangular distributions are clustered between $x_1$ to $x_{15}$. To minimise the loss of information, TRIAV should be run both for the full range and for the range $x_1$ to $x_{12}$.

---

* These variables should be entered with a decimal point.




## RUNNING THE PROGRAM

The program resides in the BMR's Hewlett-Packard computer, and is called TRIAV. To run the program, after preparing the data file 'DATAF', use the commands:

| | |
|---|---|
| RP,TRIAV | (to restore program TRIAV) |
| RU,TRIAV,6 | (to run program TRIAV) |
| LP,-N | (to send the output to the printer) |

### Example terminal session

This example of a complete terminal session, from log-on to log-off, is given so that persons with little or no experience of the computer system may run the program. Data or information that would be displayed on the terminal by the computer is underlined; data or information that is to be typed in by the user is not underlined. The symbol "^" indicates that a blank space is to be typed. The return key is to be pressed after each line is entered.

For a greater understanding of the commands involved here, the reader is referred to 'B.M.R. HP-1000 computer user's reference manual' (1983).

Press any key

```
PLEASE LOG-ON:  ALAN.RESAS
SESSION etc...
PREVIOUS TOTAL SESSION TIME etc...
SECURITY CODE?  915
!ED,-                              (Run the editor in input mode)
:^^^^4^^^^^^^^1.0^^^^^^^^4.0       (4 triangular distributions to be averaged;
                                    limit the range to 1.0 to 4.0)
:EXAMPLE 1                         (title)
:3.5^^^^^^^^4.0^^^^^^^^6.2         (first triangular distribution)
:1.0^^^^^^^^1.5^^^^^^^^4.0         (second    "         "     )
:1.1^^^^^^^^1.2^^^^^^^^2.1         (third     "         "     )
:1.0^^^^^^^^8.0^^^^^^^18.5         (fourth    "         "     )
:^^^^2
:EXAMPLE 2
:100.0^^^^^^150.0^^^^^^200.0
:100.0^^^^^^180.0^^^^^^210.0
:*EOD
:DU,DATTR                          (Duplicate the data into file 'DATTR',
                                    if an error occurs try SF,DATTR)
DATTR CLOSED 10 LINES
!RP,TRIAV                          (Restore program TRIAV)
!RU,TRIAV,6                        (Run program TRIAV)
TRIAV LUN 06 SPOOL FILE -0555      (-0555 is the spool file produced by
                                    the program; the actual number varies
                                    from run to run)
!LP,-0555 (or LP,-N)               (Send the spool file to the printer)
!EX                                (Log-off)
```

The output corresponding to this example is given in Appendix II.

ACKNOWLEDGEMENTS

The author wishes to thank the First Assistant Director, Lee Ranford, for initially requesting this work and assistance with the text; his supervisor, David Forman, for assistance with the text; and Brian Elliott for helpful suggestions and additions.

## REFERENCES

B.M.R. HP-1000 computer user's reference manual, 6/10/83.

Newendorp, P.D., 1975 - Decision analysis for petroleum exploration. The Petroleum Publishing Company, Tulsa, Oklahoma.

Smith, E.L., 1981 - DELPHI: A computer program for the summation of hydrocarbon resources using Monte Carlo simulation. BMR record 1981/65, BMR MF176, unpublished.

# APPENDIX I - RUNNING PROGRAM DELPHI

# APPENDIX I - RUNNING PROGRAM DELPHI

## Introduction

Program DELPHI is a FORTRAN program residing on the CSIRO, CYBER 76 computer. It is stored on the BMR's disc, DMR3922, under the project identification (PID) of DMRXVA, and user identification (UID) of DMRVRA. It was written by Eugene Smith, who provided complete documentation (Smith, 1981). This Appendix is a brief user's guide, taken largely from Smith (1981), and generalised for the assessment of any resource.

DELPHI is designed to summate probability distributions of hydrocarbon resources. It can be used for other resources, such as uranium resources, but the output will be documented in terms of oil and gas, and units relevant to oil and gas. Input is called for both oil and gas; the input for gas should be left blank when the program is used for single resources.

Addition is carried out by sampling each input distribution at random, and then adding the random values. This process is repeated many times and the results are given as a cumulative probability curve of resources. This is known as a Monte Carlo method (Newendorp, 1975).

## Running the program

To obtain the program, the user may LOG-ON under the above PID and UID, then copy the program into the workspace via the command,

```
*/ATTACH,DELPHI,ID=DMRXVA,SN=DMR3922./
```

The data, at the end, will need to be changed for each new run.

The file should contain the following:

1. DELPHI,PO100,T30,MS140000.
2. COMMENT. Name and extension number of person running the job.
3. FTN.
4. LGO.
5. *EOS
6. The program

7. *EOS
8. The data

To submit the program for execution, use the command

\K/A/

and pick up the output from the collection point at the node.

Input Data

| card | cols | variable | description | format for entire card |
|---|---|---|---|---|
| 1 | 1-80 | TITLE | Title of the study (80 characters) | 8A10 |
| 2 | 1-80 | CB | Person(s) who compiled the data (80 characters) | 8A10 |
| 3 | 1-80 | CD | Compilation date (80 characters) | 8A10 |
| 4 | 1-2 | NF | Number of data sets (e.g. basins or areas)(max 50) | I2 |

Cards 5 through N are to be repeated NF times:

| card | cols | variable | description | format for entire card |
|---|---|---|---|---|
| 5 | 1-80 | FIELD | Name of data set (basin or area) (80 characters) | 8A10 |
| 6 | 1 | TYPO | Type of input data distribution (1 character): K = constant, U = uniform, T = triangular, C = cumulative (see Note 1.) | A1,2I2,A1,F4.2,7F10.3/ 8F10.3/8F10.3/7F10.3 |
| | 2-3 | NUM | Number of pairs of values used to define the distribution (max 15) (see note 1.) | |
| | 4-5 | MAGO | Magnitude (power of ten) of resource values in distribution (max 99) | |
| | 6 | UNITO | Volumetric unit of measure for resource (1 character) B = barrels, M = cubic metres (see Note 2.) | |
| | 7-10 | ORISK | Existence risk (use 1.0 if no risk) | |
| | 11-20<br>21-30<br>31-40<br>41-50 | OILVAL(1)<br>OFREQ(1)<br>OILVAL(2)<br>OFREQ(2)<br>⋮ | Up to 15 pairs of values may be entered to describe the distribution. The first value of each pair (OILVAL) is the volume of oil (or the amount of resource); the second value (OFREQ) is the associated frequency or probability. If there are 4 or more pairs of values, additional cards are required. | |

.
.
.
.

N (The card following the last data card for oil for this basin or area)

This card should be left blank for a single resource - it is intended for gas in petroleum assessments.

Note 1. The type of distribution outputted by program TRIAV is cumulative (C), with 15 pairs of values.

Note 2. Use cubic metres and ignore output values in barrels.

# APPENDIX II - EXAMPLE PRINTOUT FROM TRIAV

EXAMPLE 1

TRIANGULAR DISTRIBUTIONS TO BE AVERAGED

| TRIANGULAR DISTRIBUTION NUMBER | INPUT VALUES: X1 (MINIMUM VALUE) | X2 (MOST LIKELY VALUE) | X3 (MAXIMUM VALUE) | CALCULATED STATISTICS OF THE TRIANGULAR DISTRIBUTIONS: MEAN | STANDARD DEVIATION |
|---|---|---|---|---|---|
| 1 | 3.5000 | 4.0000 | 6.2000 | 4.5667 | .5864 |
| 2 | 1.0000 | 1.5000 | 4.0000 | 2.1667 | .6562 |
| 3 | 1.1000 | 1.2000 | 2.1000 | 1.4667 | .2248 |
| 4 | 1.0000 | 8.0000 | 8.5000 | 5.8333 | 1.7119 |

AVERAGE DISTRIBUTION FOR THE N TRIANGULAR DISTRIBUTIONS

(1) AT EQUALLY SPACED VALUES OF X

| X | CUM.FREQ. | FREQUENCY | FREQUENCY HISTOGRAM |
|---|---|---|---|
| 1.000 | 0.0000 | .0000E+00 | * |
| 1.214 | .0400 | .1865E+00 | ****************** |
| 1.429 | .1563 | .5427E+00 | **************************************************** |
| 1.643 | .2587 | .4781E+00 | ********************************************** |
| 1.857 | .3341 | .3516E+00 | ********************************** |
| 2.071 | .3813 | .2203E+00 | ********************* |
| 2.286 | .4099 | .1337E+00 | ************* |
| 2.500 | .4357 | .1204E+00 | ************ |
| 2.714 | .4589 | .1082E+00 | ********** |
| 2.929 | .4794 | .9592E-01 | ********* |
| 3.143 | .4974 | .8367E-01 | ******** |
| 3.357 | .5127 | .7143E-01 | ******* |
| 3.571 | .5263 | .6359E-01 | ****** |
| 3.786 | .5505 | .1131E+00 | *********** |
| 4.000 | .5892 | .1802E+00 | ***************** |

TRUNCATED RANGE

(2) AT EQUALLY SPACED FREQUENCIES

| X | CUM.FREQ. |
|---|---|
| 1.00 | 0.0000 |
| 1.27 | .0714 |
| 1.40 | .1429 |
| 1.54 | .2143 |
| 1.71 | .2857 |
| 1.95 | .3571 |
| 2.44 | .4286 |
| 3.18 | .5000 |
| 3.91 | .5714 |
| 4.26 | .6429 |
| 4.66 | .7143 |
| 5.15 | .7857 |
| 5.85 | .8571 |
| 7.12 | .9286 |
| 4.00 | 1.0000 |

MEAN VALUE OF X (CALCULATED BY AVERAGING THE MEANS OF THE TRIANGULAR DISTRIBUTIONS) = 3.508

PERCENTILES OF THE AVERAGE DISTRIBUTION

PROBABILITY OF VALUE BEING <OR= TO THE VALUE X SHOWN

| PROBABILITY(%) | X |
|---|---|
| 10 | 1.325 |
| 20 | 1.520 |
| 30 | 1.760 |
| 40 | 2.212 |
| 50 | 3.180 |

EXAMPLE 2

TRIANGULAR DISTRIBUTIONS TO BE AVERAGED

| TRIANGULAR DISTRIBUTION NUMBER | INPUT VALUES X1 (MINIMUM VALUE) | X2(MOST (LIKELY VALUE) | X3 (MAXIMUM VALUE) | CALCULATED STATISTICS OF THE TRIANGULAR DISTRIBUTIONS MEAN | STANDARD DEVIATION |
|---|---|---|---|---|---|
| 1 | 100.0000 | 150.0000 | 200.0000 | 150.0000 | 20.4124 |
| 2 | 100.0000 | 180.0000 | 210.0000 | 163.3333 | 23.2140 |

AVERAGE DISTRIBUTION FOR THE N TRIANGULAR DISTRIBUTIONS

(1) AT EQUALLY SPACED VALUES OF X

| X | CUM.FREQ. | FREQUENCY | FREQUENCY HISTOGRAM |
|---|---|---|---|
| 100.000 | 0.0000 | .0000E+00 | * |
| 107.857 | .0097 | .1232E-02 | ***** |
| 115.714 | .0387 | .3696E-02 | ************* |
| 123.571 | .0871 | .6161E-02 | ********************* |
| 131.429 | .1549 | .8625E-02 | ***************************** |
| 139.286 | .2420 | .1109E-01 | ************************************* |
| 147.143 | .3485 | .1355E-01 | ********************************************* |
| 155.000 | .4694 | .1538E-01 | *************************************************** |
| 162.857 | .5865 | .1491E-01 | ************************************************* |
| 170.714 | .6984 | .1423E-01 | *********************************************** |
| 178.571 | .8048 | .1355E-01 | ********************************************* |
| 186.429 | .8974 | .1178E-01 | *************************************** |
| 194.286 | .9593 | .7881E-02 | ************************** |
| 202.143 | .9906 | .3987E-02 | ************* |
| 210.000 | 1.0000 | .1190E-02 | **** |

TRUNCATED RANGE

(2) AT EQUALLY SPACED FREQUENCIES

| X | CUM.FREQ. |
|---|---|
| 100.00 | 0.0000 |
| 121.35 | .0714 |
| 130.18 | .1429 |
| 137.01 | .2143 |
| 142.69 | .2857 |
| 147.71 | .3571 |
| 152.35 | .4286 |
| 157.02 | .5000 |
| 161.83 | .5714 |
| 166.77 | .6429 |
| 171.87 | .7143 |
| 177.13 | .7857 |
| 182.70 | .8571 |
| 189.89 | .9286 |
| 210.00 | 1.0000 |

MEAN VALUE OF X (CALCULATED BY AVERAGING THE MEANS OF THE TRIANGULAR DISTRIBUTIONS) = 156.667

PERCENTILES OF THE AVERAGE DISTRIBUTION

PROBABILITY OF VALUE BEING (OR= TO THE VALUE X SHOWN

| PROBABILITY(%) | X |
|---|---|
| 10 | 125.064 |
| 20 | 135.496 |
| 30 | 143.563 |
| 40 | 150.490 |
| 50 | 157.054 |
| 60 | 163.804 |
| 70 | 170.836 |
| 80 | 178.214 |
| 90 | 186.759 |