UNITED STATES DEPARTMENT OF THE INTERIOR

GEOLOGICAL SURVEY

COMPUTER PROGRAM FOR THE CALCULATION OF GRAIN-SIZE STATISTICS BY THE METHOD OF MOMENTS

By

Michael B. Sawyer

Open-File Report 77-580

1977

This report is preliminary and has not been edited or reviewed for conformity with U.S. Geological Survey standards and nomenclature.

Contents

Illustrations

Abstract

A computer program is presented for a Hewlett-Packard Model 9830A desk-top calculator (1) which calculates statistics using weight or point count data from a grain-size analysis. The program uses the method of moments in contrast to the more commonly used but less inclusive graphic method of Folk and Ward (1957).

The merits of the program are: (1) it is rapid; (2) it can accept data in either grouped or ungrouped format; (3) it allows direct comparison with grain-size data in the literature that have been calculated by the method of moments; (4) it utilizes all of the original data rather than percentiles from the cumulative curve as in the approximation technique used by the graphic method; (5) it is written in the computer language BASIC, which is easily modified and adapted to a wide variety of computers; and (6) when used in the HP-9830A, it does not require punching of data cards. The method of moments should be used only if the entire sample has been measured and the worker defines the measured grain-size range.

---

(1) Use of brand names in this paper does not imply endorsement of these products by the U.S. Geological Survey.

Introduction

The purpose of this paper is to present a relatively simple computer program that computes moment measures for use on a Hewlett-Packard Model 9830A desk-top calculator. (1) The program calculates the mean, variance, standard deviation, skewness, and kurtosis for a grain-size distribution and keeps track of the number of data values. It also permits an optional printout of the raw data and a non-optional printout of the sample identification number. The moment measure equations are given on the following page (modified from Remington and Schork, 1970).

---

(1) Use of brand names in this paper does not imply endorsement of these products by the U.S. Geological Survey.

$$\text{Mean} = \bar{x} = \frac{1}{n} \sum_{i=1}^{n} f_i x_i$$

$$\text{Variance} = s^2 = \frac{1}{n-1} \left[\sum_{i=1}^{n} f_i x_i^2 - n(\bar{x})^2 \right]$$

$$\text{Standard deviation} = s = \sqrt{s^2}$$

$$\text{Skewness} = a_3 = \frac{1}{s^3} \left[\frac{1}{n} \sum_{i=1}^{n} f_i x_i^3 - \frac{3}{n} \bar{x} \sum_{i=1}^{n} f_i x_i^2 + 2(\bar{x})^3 \right]$$

$$\text{Kurtosis} = a_4 = \frac{1}{s^4} \left[\frac{1}{n} \sum_{i=1}^{n} f_i x_i^4 - \frac{4}{n} \bar{x} \sum_{i=1}^{n} f_i x_i^3 + \frac{6}{n} (\bar{x})^2 \sum_{i=1}^{n} f_i x_i^2 - 3(\bar{x})^4 \right]$$

where f_i is the frequency of the observations, grouped or ungrouped,

x_i is a single observation for ungrouped data or a class midpoint for grouped data,

and n is the total number of observations or the sum of the frequencies.

Data are entered from the keyboard in the format of a frequency followed by a grain-size value. In the case of ungrouped data, a frequency value of "1" would be entered.

The main problem to arise from increased use of the computer has been the lack of users who are trained in computer programming. Many users are able to use existing programs but are unable to write their own programs or to modify existing software. For this reason, publication of computer software has become increasingly common and important to all involved, so that more efficient use may be made of existing computer facilities and personnel. This program arose from a study of the use of image analysis for the determination of grain-size distributions of sediments for the Basin Analysis program element of the U.S. Geological Survey's Uranium and Thorium Exploration Research and Resource Assessment Program. The volume of data produced by an image analyzing computer is such that it can only be handled practically by a computer; therefore, many programs have had to be written. This program is a simplified version of several programs written for use with the image analyzing computer.

Equipment

The Hewlett-Packard equipment used to run this program is as follows:

Model 9830A programmable calculator with 4000 bytes (2000 words) of Read-Write-Memory

Model 9866A thermal printer

String variables ROM (Read-Only-Memory block)

Extended I/O ROM

The only statements in the program that require the Extended I/O ROM are the "PRINT LIN(N)" statements. If this ROM is not available, these statements may be replaced by multiple "PRINT" statements. For example, the statement "60 PRINT LIN(2)", which causes the paper to advance two lines, could be replaced by "60 PRINT" and "61 PRINT". These two statements would also cause the paper to advance two lines. The "PRINT LIN(N)" statement is used in lines 20, 60, 140, 460, and 540 (Figure 1) as an abbreviated method for controlling the format of the printout.

Program operation and format

The explanation of the program operation consists of two parts: an explanation of neccessary program logic and an explanation of the manual operations to be performed by the program user. The term "statement" will be used to refer to the numbered program lines, and the term "operation" will refer to procedures to be performed manually by the program user. Statements performing related operations will be discussed in groups. Figure 1 is a copy of the program, and Figure 2 is a typical printout of the results of a program run.

Figures 1 and 2.--NEAR HERE

(In figure 1, on lines 360, 370, 380, 420, 440, and 450, the carat mark "^" represents an up-arrow.)

```
10 DIM A$[20],B$[3],C$[3],A[3],B[3]
20 PRINT LIN(5)
30 PRINT "WHEN YOU'RE THROUGH ENTERING DATA, ENTER -1,0 AS"
40 PRINT "THE LAST DATA VALUES."
50 PRINT "ALWAYS SEPARATE THE FREQUENCY AND SIZE WITH A COMMA."
60 PRINT LIN(2)
70 DISP "PLEASE READ PRINTOUT."
80 WAIT 3000
90 N=X1=X2=X3=X4=0
100 DISP "DO YOU WANT THE DATA PRINTED";
110 INPUT B$
120 DISP "SAMPLE NO.";
130 INPUT A$
140 PRINT TAB40,"SAMPLE NO.:"A$,LIN(2)
150 F=0
160 F=F+1
170 DISP "ENTER FREQUENCY, SIZE";
180 INPUT A[F],B[F]
190 IF B$="NO" THEN 330
200 IF A[F]<0 THEN 250
210 IF F#3 THEN 340
220 WRITE (15,230)A[1],B[1],A[2],B[2],A[3],B[3]
230 FORMAT F8.3,3X,F8.3,6X,F8.3,3X,F8.3,6X,F8.3,3X,F8.3
240 GOTO 340
250 IF F=1 AND A[F]<0 THEN 330
260 IF F=2 AND A[F]<0 THEN 280
270 IF F=3 AND A[F]<0 THEN 310
280 WRITE (15,290)A[1],B[1]
290 FORMAT F8.3,3X,F8.3
300 GOTO 330
310 WRITE (15,320)A[1],B[1],A[2],B[2]
320 FORMAT F8.3,3X,F8.3,6X,F8.3,3X,F8.3
330 IF A[F]<0 THEN 410
340 N=N+A[F]
350 X1=X1+(A[F]*B[F])
360 X2=X2+(B[F]^2*A[F])
370 X3=X3+(B[F]^3*A[F])
380 X4=X4+(B[F]^4*A[F])
390 IF F#3 THEN 160
400 IF F=3 THEN 150
410 M=X1/N
420 V=(X2-N*M^2)/(N-1)
430 D=SQR(V)
440 S=((X3/N)-((3*M*X2)/N)+(2*M^3))/D^3
450 K=((X4/N)-((4*M*X3)/N)+((6*M^2*X2)/N)-(3*M^4))/D^4
460 PRINT LIN(2)
```

Figure 1.--Computer program for the calculation of grain-size statistics by the method of moments.

```
470 FIXED 3
480 PRINT "MEAN= "M"PHI"
490 PRINT "VARIANCE= "V
500 PRINT "STANDARD DEVIATION= "D"PHI"
510 PRINT "SKEWNESS= "S
520 PRINT "KURTOSIS= "K
530 PRINT "N= "N
540 PRINT LIN(5)
550 STANDARD
560 DISP "MORE DATA TO BE ENTERED";
570 INPUT C$
580 IF C$="YES" THEN 90
590 END
```

Figure 1.--continued

```
WHEN YOU'RE THROUGH ENTERING DATA, ENTER -1,0 AS
THE LAST DATA VALUES.
ALWAYS SEPARATE THE FREQUENCY AND SIZE WITH A COMMA.


                                        SAMPLE NO.:TEST


   1.000      -2.000       3.000     -1.000       7.000      0.000
  10.000       1.000      36.000      2.000      57.000      3.000
  35.000       4.000      22.000      5.000       6.000      6.000
   2.000       7.000       1.000      8.000


MEAN=  3.089   PHI
VARIANCE=  2.484
STANDARD DEVIATION=  1.576   PHI
SKEWNESS= -0.163
KURTOSIS=  3.770
N=  180.000
```

Figure 2.--Typical printout of data and calculated statistics.

Operation I. When the program has been loaded into the calculator memory from a tape cassette or typed in from the keyboard, the "RUN" key and then the "EXECUTE" key are pressed.

Statements 100-110 ask the user if he wishes to have his raw data values printed in order to reduce the possibility of typographical errors in data entry.

Operation II. The answer "YES" or "NO" is typed in response to the question asked in statements 100-110 followed by pressing the "EXECUTE" key.

Statements 120-130 ask the user to enter the sample number.

Operation III. The sample number is typed in response to statements 120-130 followed by pressing the "EXECUTE" key. The sample number has arbitrarily been limited to 20 characters including spaces. Use of larger sample numbers will cause an error message to be displayed.

Statement 140 prints the sample number.

Statements 170-180 ask the user to enter a frequency and a size value.

Operation IV. The user types a frequency value, a comma, and then a size value followed by pressing the "EXECUTE" key.

Statements 190-320 control the format in which the data values are printed. Up to three pairs of data values are printed on each line dependent upon the total number of data values entered. The first value on a line is a frequency value, and the second value is the corresponding size. This format is repeated across the line.

Statements 560-570 ask the user if he wishes to enter a new data set.

Operation V. The user types "YES" or "NO" in response to the question asked in statements 560-570 followed by pressing the "EXECUTE" key.

Statement 580 checks the response to the question in statements 560-570. If the user has more data to be entered, program execution continues at statement 90 and the program, in effect, begins again. If the user is through entering data, the program ends.

Advantages of the program

This program has the following advantages over hand calculation and some of the other computer programs written to calculate statistics for grain-size analyses:

(1) Grain-size data for a large number of samples can be analyzed very rapidly. Approximately thirty typical sets of data can be analyzed per hour using this program.

(2) Most other programs calculate grain-size statistics by the graphic method (Slatt and Press, 1976). Their method was devised to make hand calculation of grain-size statistics easier, but it is only an approximation to the method of moments and disregards the tails of the distribution that many sedimentologists believe to be very important. According to Folk (1968), the method of moments "probably gives a truer picture than the graphic methods, which rely on only a few selected percentage lines." In fact, Folk's graphic mean uses only the 16th, 50th, and 84th percentiles and thus disregards 32 percent of the sample. In addition, his inclusive graphic equations for standard deviation, skewness, and kurtosis only extend outward to the 5th and 95th percentiles, thus disregarding 10 percent of the distribution.

(3) It has been argued by Slatt and Press (1976) that the graphic method should be used because it has been more commonly used in the past and because many of the published grain-size statistics were calculated by this technique. This argument has some strengths, but geologists in recent years have been turning

to the method of moments because it is now practical to use it. The volume of literature containing moments measure statistics is now large enough to warrant use of this technique.

(4) The program can accept data in either grouped or ungrouped format. Ungrouped data, if obtainable, will always produce more accurate results than grouped data because discrete data values are used. Grouping of the data tends to cause a loss of resolution in the data set and thus causes inaccuracy.

(5) The program is written in a modified version of the computer language BASIC. Most major computers and many smaller ones can use this language, and thus the program can be easily adapted to any of these machines.

(6) The same logic used in this program could be used to write a similar program in another computer language such as FORTRAN.

(7) The HP-9830A does not require special knowledge of computer programming in order to be used. Once the program has been typed into memory, its operation is self-explanatory. It also does not require any punching of computer cards.

Disadvantages of the method of moments

Because the method of moments is affected by every data value in the distribution, caution should be used when the distribution is "open-ended" (Folk, 1968). This situation occurs in sieving when the pan fraction is not analyzed by pipette or hydrometer methods to determine its grain-size distribution. There are three possible solutions to this problem. The first is to analyze this fine fraction and then add these data to the data for the coarse fraction. The second solution is to choose an arbitrary mean size for the fine fraction. This will produce some error, but may be acceptable in many cases. The third solution is to essentially ignore this fine fraction and calculate the moments for the coarse fraction only. Where the fine fraction is only a small percentage of the sample, its exclusion may not produce a large error, but the worker should be careful to define the size range of material that was measured.

References cited

Folk, R. L., 1968, Petrology of sedimentary rocks: Hemphill's, Austin, Texas, p. 49.

Folk, R. L. and Ward, W. C., 1957, Brazos River Bar--A study in the significance of grain size parameters: Jour. Sed. Petrology, v. 27, p. 3-26.

Remington, R. D., and Schork, M. A., 1970, Statistics with applications to the biological and health sciences: Prentice-Hall Inc., Englewood Cliffs, N.J., p. 218-224.

Slatt, R. M. and Press, D. E., 1976, Computer program for presentation of grain-size data by the graphic method: Sedimentology, v. 23, p. 121-131.