DANIEL F. MERRIAM, Editor

FORTRAN IV COMPUTER PROGRAM FOR FITTING OBSERVED COUNT DATA TO DISCRETE DISTRIBUTION MODELS OF BINOMIAL, POISSON AND NEGATIVE BINOMIAL

By

CHARLES W. ONDRICK
Kansas Geological Survey

and

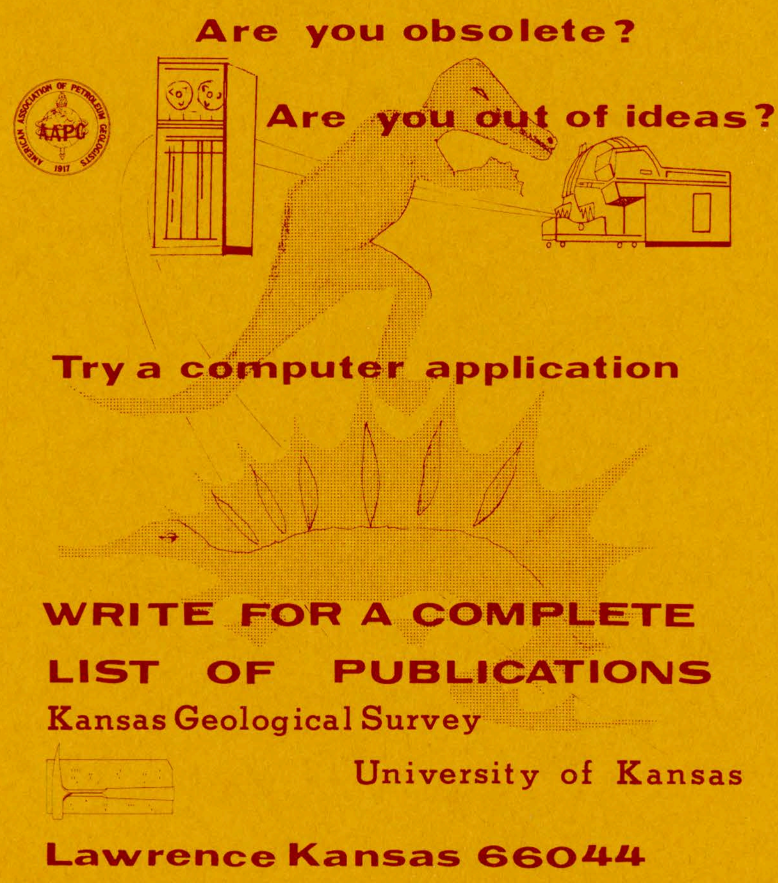
JOHN C. GRIFFITHS
Pennsylvania State University

in cooperation with the
American Association of Petroleum Geologists
Tulsa, Oklahoma

COMPUTER CONTRIBUTION 35
State Geological Survey
The University of Kansas, Lawrence
1969

Editor's Remarks

For a limited time the Geological Survey will make available on magnetic tape the program (Binom) as described here for $15.00 (US). If punched cards are required, please add another $10.00 for handling charges.



*Active Member, °Associate Member, +Junior Member, American Association of Petroleum Geologists.

FORTRAN IV COMPUTER PROGRAM FOR FITTING OBSERVED COUNT DATA TO DISCRETE DISTRIBUTION MODELS OF BINOMIAL, POISSON AND NEGATIVE BINOMIAL

by

Charles W. Ondrick and John C. Griffiths

INTRODUCTION

Data may be divided into two distinct types, measurement and count. Measurement data exhibits a continuous range in variation whereas count data varies in discrete steps. The present discussion and computer program is concerned primarily with count data and the positive binomial, Poisson, and negative binomial discrete distribution models.

The types of frequency distributions generated are dependent upon sampling arrangement, measurement technique and distribution of the constituents to be analyzed. Consider a conceptual experiment where the outcome can be thought of as success or failure, for example quartz grain (success) and not quartz grain (failure). Constant probability models such as the positive binomial and Poisson are expected if the following conditions are fulfilled (Student 1919, p. 211):

1. The chance of a point falling on a particular item is the same for each point.
2. The chance of a point falling on a particular item is the same for each particular item.
3. The fact that a point has fallen on a particular item does not affect the chance of subsequent points falling on that item.

According to Student (1919), condition three is often unfulfilled. If a point falling on a particular item, for example a quartz grain, decreases the chance of points falling on quartz grains, the positive binomial discrete-distribution model is expected. The negative binomial distribution model, a nonconstant probability model, is expected, however, if a point falling on a quartz grain increases the chance of points falling on quartz grains. The negative binomial has been called a contagious or epidemic distribution because the occurrence of cases of a contagious disease is clustered, that is if one case is found the chances of finding additional cases may be increased. The distribution free statistic chi square (χ^2) may be used to test the similarity of the observed to the expected frequency distribution model.

Acknowledgments. - We would like to express our thanks to Mrs. Janet Singer, College of Earth and Mineral Sciences, The Pennsylvania State University for her help during development of this program.

Positive Binomial and Poisson Distribution Models

If the above conditions one through three are fulfilled and the probability of occurrence of a particular item ranges between approximately .05 and .95 the positive binomial frequency distribution may be an appropriate model.

The positive binomial is defined as

$$N(q+p)^n$$

where (N) is the number of trials (traverses), (n) number of points within a trial, p the probability of an occurrence and q the probability of nonoccurrence (1-p). The general term for the expansion of the positive binomial is

$$P_X = \frac{n!}{X!\,(n-X)!}\, p^X q^{n-X}$$

where n, p, q have been defined above and X is the number of successes or events. P_X is the expected probability of the event (success).

The Poisson model may be considered a special case of the positive binomial model. If conditions one through three above are fulfilled and the probability of occurrence of a particular item is less than .05 or greater than .95 the Poisson frequency distribution is the expected model. The general term for the expansion of the Poisson is

$$P_X = \frac{(np)^X}{X!}\, e^{-np}$$

Negative Binomial Distribution Model

The negative binomial may be used as an alternative model. This is a nonconstant probability model and condition three above is not fulfilled. If a point falls on a particular item the probability that the next point will fall again on that item tends to be increased. A detailed theoretical discussion of the negative binomial is presented in Bliss and Fisher (1953).

The negative binomial is defined as

$$N(q-p)^{-k}$$

where N, remains as previously defined, p the probability of occurrence of the event, q the probability of nonoccurrence; and $P=\bar{x}/k$ and $q=1+P$. The exponent k is first approximated by the equation

$$k_1 = \frac{\overline{X}^2}{S^2 - \overline{X}}$$

and then successively substituting new trial values of k in the equation

$$Z_i = S\left(\frac{A_x}{k'_i + X}\right) - N\ln\left(1 + \frac{\overline{X}}{k'_i}\right)$$

until the Z value equals zero. In order to make the Z values converge on zero (0.00005 in the present program) the Newton-Raphson technique is used and a derivative of Z is then calculated (DZ value).

The values R, P and Q in the program output are the estimates of the parameters of the negative binomial and L is the number of iterations required for Z to converge on zero. P and Q are defined above (p, q) and R is equal to P/Q.

The general term for the expansion of the negative binomial is

$$P_x = \frac{(k+X-1)!}{X!\,(k-1)!}\ \frac{R^X}{q^k}$$

GEOLOGICAL EXAMPLES

The geological literature is markedly deficient in the use of positive binomial, Poisson and negative binomial models when in fact much of the data collected are in the form of counts. Examples include petrographic modal analysis, estimations of grain packing, and frequency of natural resource targets. Griffiths (1960a, 1960b, 1962, 1967) and Ondrick (1965, 1968) provide examples of the fitting of observed petrographic modal analysis data to expected distribution models and their geological interpretation. Griffiths and Drew (1966) and Griffiths (1966) extend the applications to the search for natural resources.

The fit of the observed distribution of sphene in the Rensselaer Graywacke, Rensselaer County, New York to both the Poisson and negative binomial models is given in Figure 1. The sampling consisted of 40 sets of 5 points each per thin section or 4800 sets for the 120 thin sections counted. Note the extended tail of the observed distribution toward the higher number of successes and the smaller value of χ^2 (4.79) for the negative binomial fit in comparison to the Poisson fit (213.49). In the observed distribution there are too few 1 successes and too many 2, 3, 4 successes to fit the expected Poisson distribution. It is concluded therefore that sphene is a rare mineral in the Rensselaer Graywacke and that it occurs throughout the rock in blobs and blotches. That is, during point counting, if a point falls on sphene the probability that the next point will land on sphene is increased. The probability of occurrence is in fact .03 ($P = \overline{x}/n = .03$).

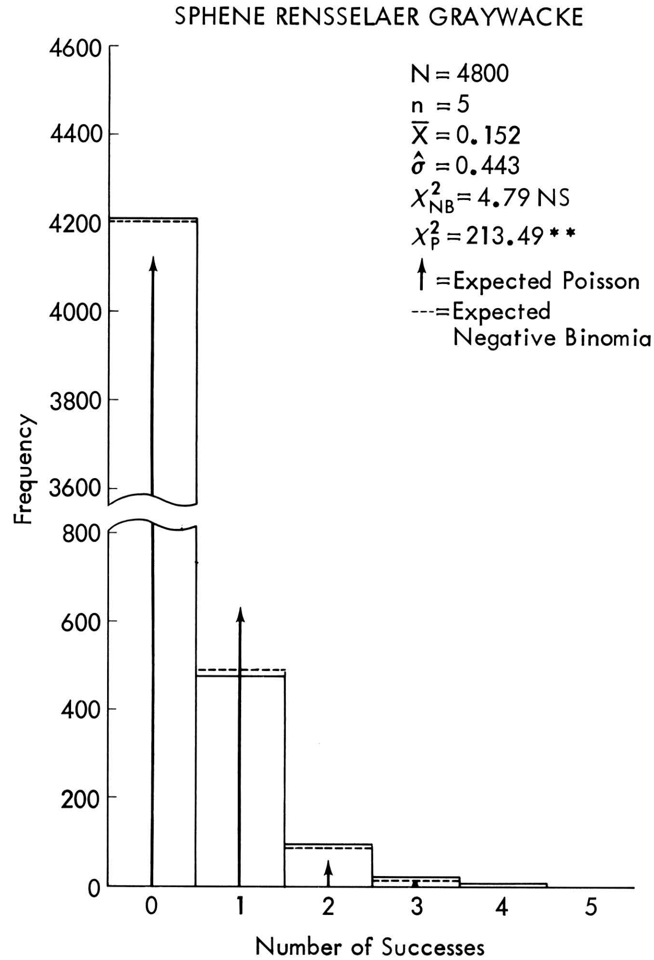


Figure 1.- Observed distribution of sphene in Rensselaer Graywacke, Troy, New York compared with expected Poisson and negative binomial distributions.

An example of the fit of the observed distribution of quartz in the Cow Run Sandstone to the positive binomial distribution is presented in Figure 2. The sampling consisted of 6 traverses of 200 points each per thin section or 156 traverses for 26 thin sections. It is concluded that quartz is a common mineral in the Cow Run Sandstone and is distributed randomly with constant probability throughout the rock.

Another example of a negative binomial distribution is illustrated by the frequency of occurrence of epidote in thin sections of the Rensselaer Graywacke; the sampling arrangement is similar to that of sphene given above and the conclusions concerning the distribution of epidote are that it occurs also as blobs and blotches and is not distributed randomly through the rock.

In a somewhat different context the frequency distribution of successes in a simulated grid sampling of Kansas for oil and gas occurrences again yields a negative binomial distribution. The original data are

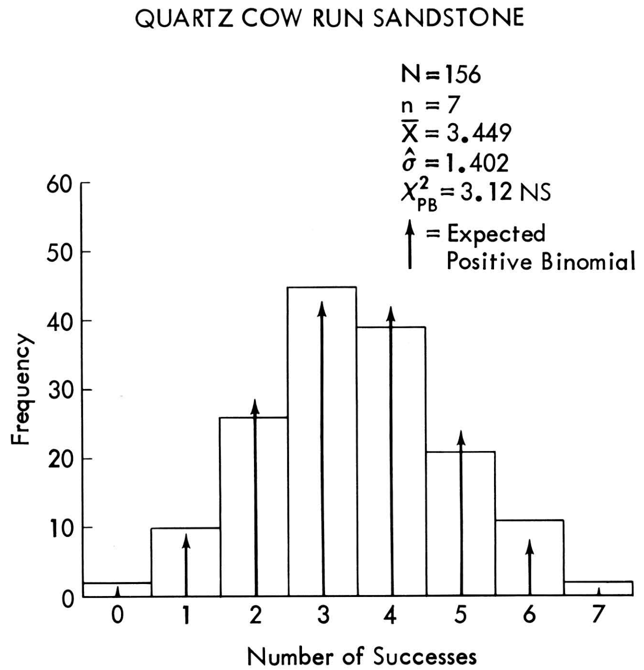


Figure 2.- Observed distribution of quartz in Cow Run Sandstone St. Marys, West Virginia compared with expected positive binomial distribution.

published in Griffiths (1966, Table I, p. 192) and the sampling arrangement consisted of 401 sets of 5 wells with 5 mile spacing between wells. This frequency distribution is rejected by chi-square if compared with a positive binomial (P<0.001) and fits a negative binomial (90>P>80). The implications are that the oil and gas fields of Kansas are clustered and not randomly distributed.

PROGRAM DESCRIPTION

A flow diagram depicting the systematics of the discrete distribution computer program is presented in Figure 3. Observed discrete frequency distributions may be compared against expected positive binomial, Poisson and negative binomial models. Options are provided to calculate either the expected constant probability models (Poisson or positive binomial) or the contagious distribution model (negative binomial) or both. If the observed frequency is compared against a constant probability model, it will automatically be compared against the Poisson if the probability of occurrence is less than .05 and to the positive binomial if the probability of occurrence is equal to or greater than .05.

Limitations

The calculation of the expected Poisson and positive binomial distributions require the computation of the factorial of certain parameters. If the expected frequency of these constant probability models

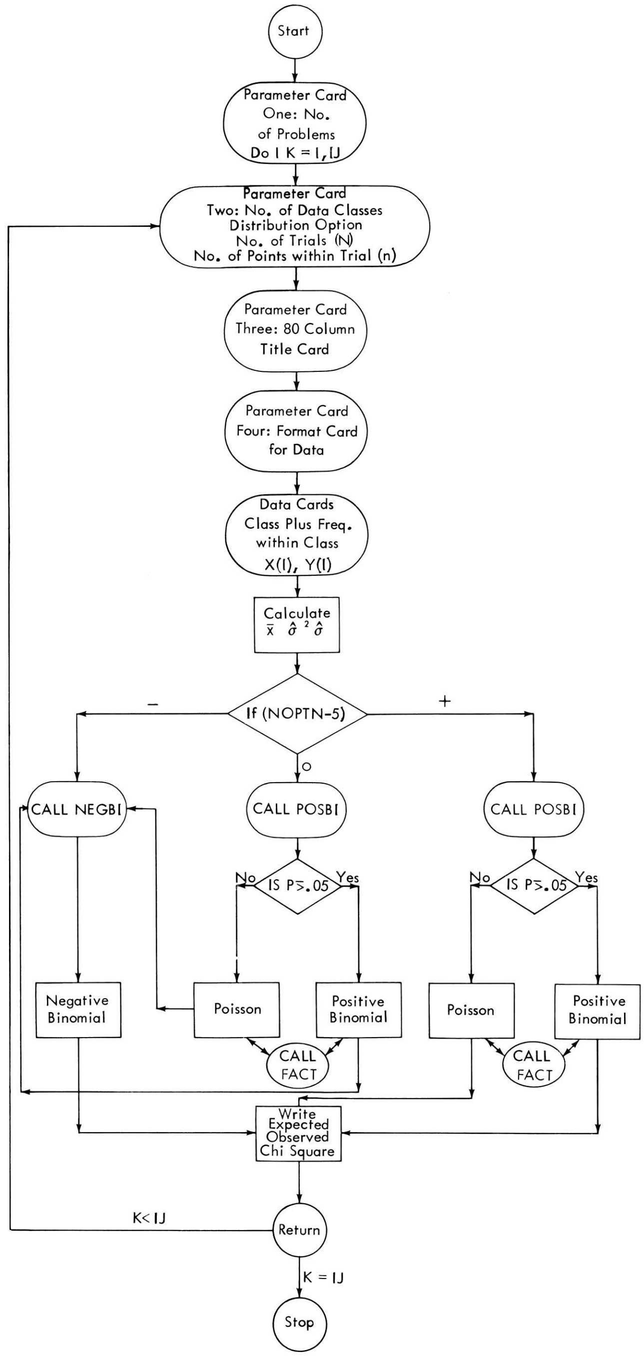


Figure 3.- Flow diagram for Discrete Distributions computer program.

is to be calculated, the number of frequency classes and the size of (n) should be limited to no greater than 10.

Approximations of the probability values of the positive binomial for n = 11 to 49 and 50 to 100 may be obtained from the National Bureau of Standards, "Tables of the Binomial Probability Distribution" (1950) and from Romig (1953) respectively.

Data may be classified into a maximum of 60 classes for the negative binomial distribution.

Input to Program

Card 1	Contains the number of problems to be analyzed. FORMAT I5
Card 2	Contains four parameters (right justify all parameters). FORMAT (2I5, 2F10.0)

Columns	
1-5	Number of classes in frequency distribution
6-10	Distribution option (see Distribution Option below)
11-20	N - number of trials (traverses or sets)
21-30	n - number of points within a trial

Card 3	Contains an 80 column alphanumeric title
Card 4	Contains the format for the data of the form (FX. X, FX. X)
Card 5 + N	Contains the data, which includes the class (0., 1., 2., 3. ..60.) followed by the frequency within the class.

Cards 2 through 5 + N are repeated for each problem to be analyzed.

Distribution Option

The observed data are fitted to the expected Poisson or positive binomial and a X^2 comparison is executed by punching a 6 in column 10 of input Card 2. The observed frequency distribution may be compared with the expected negative binomial model by punching a 4 in column 10 of input card 2. Both the expected Poisson or positive binomial and negative binomial distributions may be compared against the observed discrete distribution by punching a 5 in column 10 of the same parameter card.

Program Compatibility and Adaptation

The program described here has been compiled and executed successfully on the GE 635, IBM 360/Models 50 and 67 and the IBM 1130 computers. The card reader and printer peripherals have been assigned variable names which may be set equal to the particular numerical designation at a given computer center by changing the statements MAIN 82 and 83.

Minor program modifications are necessary for a successful compilation and execution on the IBM 1130 computer. The DIMENSION statements must precede the COMMON statements in all subroutines. Statements MAIN 120, 125, and 135 must be altered to delete the variable format option provided in larger computers. Logical IF statements are not compilable on the 1130 therefore statements FACT 70 and 75 have to be altered to:

```
20 IF (KJ-N) 21, 45, 21
21 IP=IP*(KJ-N)
```

Remove DIMENSION statements from MAIN program. Because of the limited core size of the IBM 1130 computer the subroutines POSBI and NEGBI must be placed in overlay.

REFERENCES

Bliss, C.I., and Fisher, R.A., 1953, Fitting the negative binomial distribution to biological data and note on the efficient fitting of the negative binomial: Biometrics, v. 9, p. 176-200.

Griffiths, J.C., 1960a, Modal analysis of sediments: Revue de Geog. Phys, et de Geol. Dynam., Paris, France, v. 3, Fasc. 1, p. 29-48.

Griffiths, J.C., 1960b, Frequency distributions in accessory mineral analysis: Jour. Geol., v. 68, p. 353-365.

Griffiths, J.C., 1962, Frequency distributions of some natural resource materials: Pennsylvania State Univ. Mineral Ind. Expt. Sta. Circ. No. 63, p. 174-198.

Griffiths, J.C., 1966, Exploration for Natural resources: Jour. Operations Res. Soc. Amer., v. 14, p. 189-209.

Griffiths, J.C., 1967, Scientific method in the analysis of sediments: McGraw-Hill Book Co., New York 508 p.

Griffiths, J.C., and Drew, L.J., 1966, Grid spacing and success ratios in exploration for natural resources, in Proc. of Symposium and Short Course on Computers and Operations Research in Mineral Industries: Pennsylvania State Univ. Mineral Ind. Expt. Sta. Spec. Publ. 2-65, v. 1, p. Q1-24.

National Bureau of Standards, 1950, Tables of the binomial probability distribution: Natl. Bur, Std. U.S. Appl. Math. Ser. 6, 387 p.

Ondrick, C.W., 1965, Statistical comparison of the Keener and Big Injun sands, Pleasants County, West Virginia: unpublished masters thesis, Dept. of Geochemistry and Mineralogy, Pennsylvania State Univ., 185 p.

Ondrick, C.W., 1968, Petrography and geochemistry of the Rensselaer Graywacke, Troy, New York: unpublished doctorate dissertation, Dept. of Geochemistry and Mineralogy, Pennsylvania State Univ., 218 p.

Romig, H., 1953, 50-100 binomial tables: John Wiley & Sons, Inc., New York, 172 p.

Student, 1919, An explanation of deviation from Poisson's law in practice: Biometrika, v. 12, p. 211-215.

Listing of program

```
C     ***********************************************************************
C     **********************DISCRETE DISTRIBUTIONS**********************
C     ***********************************************************************
C     PARAMETER CARD ONE  NUMBER OF SETS OF DATA IN FORMAT I5,
C     PARAMETER CARD TWO  CONTAINS FOUR PARAMETERS IN THE FOLLOWING ORDE
C     NUMBER OF CLASSES OF DATA, OPTION NEGATIVE BINOMIAL FIT ONLY
C     (PUNCH A 4), OPTION POISSON/POSITIVE BINOMIAL FIT ONLY(PUNCH A 6),
C     OPTION BOTH NEGATIVE BINOMIAL AND POISSON/POSITIVE BINOMIAL FIT
C     (PUNCH A 5),NUMBER OF SETS OF N POINTS COMPOSING THE DISTRIBUTION,
C     NUMBER OF POINTS WITHIN A SET. FORMAT(2I5,2F10.0)
C     PARAMETER CARD THREE  AN 80 COLUMN ALPHANUMERIC TITLE CARD.
C     PARAMETER CARD FOUR  FORMAT OF DATA CARDS (FX,X, FX.X)
C     ***********************************************************************
      COMMON X(60),Y(60),TITLE(20),FMT(20),NCLAS,BIGN,SMLLN,AVEX,SF,VAR
      COMMON IRDO,IWR
      DIMENSION PHI(60),CHI(60),ACC(60),PROB(60),EK(60),Z(60),DER(60)
      DIMENSION DZ(60),AEXP(60),ZEXP(60),PX(60),CHISQ(60),AOBS(60)
      IRDO=5
      IWR=6
      READ (IRDO,5) ISETS
    5 FORMAT (I5)
      DO 80 I=1,ISETS
      READ (IRDO,10) NCLAS,NOPTN,BIGN,SMLLN
   10 FORMAT (2I5,2F10.0)
      READ (IRDO,15) TITLE
   15 FORMAT (20A4)
      READ (IRDO,20) FMT
   20 FORMAT (20A4)
      DO 25 J=1,NCLAS
      READ (IRDO,FMT) X(J),Y(J)
   25 CONTINUE
      SF=0.0
      SFD=0.0
      SFD2=0.0
      DO 30 J=1,NCLAS
      SF=SF+Y(J)
      SFD=SFD+X(J)*Y(J)
      SFD2=SFD2+(X(J)**2.)*Y(J)
   30 CONTINUE
      AVEX=SFD/SF
      VAR=((SFD2)/SF)-(AVEX**2)
      STDV=SQRT(VAR)
      WRITE (IWR,35)
   35 FORMAT (1H1)
      WRITE (IWR,40)
   40 FORMAT (//1H ,40X,50H**********DISCRETE DISTRIBUTIONS PROGRAM*****
     1*****)
      WRITE (IWR,45) TITLE
   45 FORMAT (//1H ,25X,20A4//)
      WRITE (IWR,50)
   50 FORMAT (1H ,40X,7HAVERAGE,8X,8HVARIANCE,8X,18HSTANDARD DEVIATION/)
      WRITE (IWR,55) AVEX,VAR,STDV
   55 FORMAT (1H ,39X,F8.3,8X,F8.3,13X,F9.4//)
      IF (NOPTN-5) 60,70,65
   60 CALL NEGBI
      GO TO 75
```

```
65 CALL POSBI
   GO TO 75
70 CALL POSBI
   CALL NEGBI
75 CONTINUE
80 CONTINUE
   STOP
   END
```

```
      SUBROUTINE FACTR(KJ,IP)
C     ****************************************************************
C     SUBROUTINE TO CALCULATE THE FACTORIAL OF A NUMBER
C     ****************************************************************
      COMMON X(60),Y(60),TITLE(20),FMT(20),NCLAS,BIGN,SMLLN,AVEX,SF,VAR
      COMMON IRDO,IWR
      IF (KJ-1) 5,5,10
    5 IP=1
      GO TO 45
   10 N=1
      DO 30 I=1,KJ
      IF (I-1) 35,15,20
   15 IP=KJ*(KJ-N)
      GO TO 25
   20 IF (KJ.EQ.N) GO TO 45
      IP=IP*(KJ-N)
   25 N=N+1
   30 CONTINUE
   35 WRITE (IWR,40)
   40 FORMAT (1H ,53X,24HLOOP INDEX LESS THAN ONE)
   45 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE POSBI
C     ****************************************************************
C     SUBROUTINE TO CALCULATE THE POISSON/POSITIVE BINOMIAL
C     ****************************************************************
      COMMON X(60),Y(60),TITLE(20),FMT(20),NCLAS,BIGN,SMLLN,AVEX,SF,VAR
      COMMON IRDO,IWR
      DIMENSION AEXP(60),ZEXP(60),PX(60),CHISQ(60),AOBS(60)
      P=AVEX/SMLLN
      IF (P-0.05) 5,35,35
C     ****************************************************************
C     GENERATION OF THE POISSON DISTRIBUTION
C     ****************************************************************
    5 WRITE (IWR,10)
   10 FORMAT (///1H0,48X,35HCALCULATION OF THE EXPECTED POISSON)
      WRITE (IWR,15)
   15 FORMAT (/1H0,27X,5HCLASS,5X,20HEXPECTED PROBABILITY,5X,18HEXPECTED
```

```
   1 FREQUENCY,5X,18HOBSERVED FREQUENCY)
     ZLN=0.0
     ZLM=0.0
     ZLP=0.0
     DO 25 J=1,NCLAS
     JA=J-1
     CALL FACTR(JA,IP)
     PX(J)=AVEX**JA/(FLOAT(IP)*2.718**AVEX)
     ZEXP(J)=BIGN*PX(J)
     ZLN=ZLN+PX(J)
     ZLM=ZLM+ZEXP(J)
     ZLP=ZLP+Y(J)
     WRITE (IWR,20) JA,PX(J),ZEXP(J),Y(J)
  20 FORMAT (1H0,29X,I3,10X,F12.8,7X,F15.4,8X,F15.4)
  25 CONTINUE
     WRITE (IWR,30)ZLN,ZLM,ZLP
  30 FORMAT (1H0,28X,6HTOTALS,8X,F12.8,7X,F15.4,8X,F15.4)
     GO TO 50
C    *****************************************************************
C    GENERATION OF THE POSITIVE BINOMIAL DISTRIBUTION
C    *****************************************************************
  35 MA=SMLLN
     WRITE (IWR,40)
  40 FORMAT (///1H0,43X,45HCALCULATION OF THE EXPECTED POSITIVE BINOMIA
    1L)
     WRITE (IWR,15)
     ZLN=0.0
     ZLM=0.0
     ZLP=0.0
     DO 45 J=1,NCLAS
     I=J-1
     KOK=MA-I
     CALL FACTR(MA,JJ)
     CALL FACTR(I,JC)
     CALL FACTR(KOK,JNV)
     PX(J)=FLOAT(JJ)*(P**I)*(1.-P)**(KOK)/(FLOAT(JNV)*FLOAT(JC))
     ZEXP(J)=BIGN*PX(J)
     ZLN=ZLN+PX(J)
     ZLM=ZLM+ZEXP(J)
     ZLP=ZLP+Y(J)
     WRITE (IWR,20)I,PX(J),ZEXP(J),Y(J)
  45 CONTINUE
     WRITE (IWR,30)ZLN,ZLM,ZLP
C    *****************************************************************
C    FOLDING EXPECTED-OBSERVED CLASSES (IF )1.0 IN A TAIL CLASS FOLD)
C    *****************************************************************
  50 KX=1
     J=1
     AOBS(1)=Y(1)
     AEXP(1)=ZEXP(1)
     IF (ZEXP(KX)-1.0) 55,80,80
  55 KX=KX+1
     AOBS(J)=AOBS(J)+Y(KX)
     AEXP(J)=AEXP(J)+ZEXP(KX)
  60 IF (AEXP(J)-1.0) 65,80,80
  65 KX=KX+1
     AEXP(J)=AEXP(J)+ZEXP(KX)
     AOBS(J)=AOBS(J)+Y(KX)
     IF (KX-(NCLAS+1)) 60,60,70
  70 WRITE (IWR,75)
```

```
   75 FORMAT (//1H0,41X,49H*****ALL CLASS FREQUENCIES ARE LESS THAN ONE*
     1****)
      GO TO 190
   80 DO 85 J=2,NCLAS
      AEXP(J)=ZEXP(J)
      AOBS(J)=Y(J)
      IF (ZEXP(J)-1.0) 90,85,85
   85 CONTINUE
   90 JZ=J
      SUM1=0.0
      SUM2=0.0
      DO 95 JA=JZ,NCLAS
      SUM1=SUM1+ZEXP(JA)
      SUM2=SUM2+Y(JA)
   95 CONTINUE
      AEXP(JZ)=SUM1
      AOBS(JZ)=SUM2
      KK=0.0
      IF (AEXP(JZ)-1.0) 100,105,105
  100 KK=1.0
      AEXP(JZ-1)=AEXP(JZ-1)+AEXP(JZ)
      AOBS(JZ-1)=AOBS(JZ-1)+AOBS(JZ)
C     *****************************************************************
C     COMPUTATION OF CHI SQUARE
C     *****************************************************************
  105 IF (KK-1) 115,110,110
  110 KA=JZ-1
      GO TO 120
  115 KA=JZ
  120 IF (KA-2) 130,130,125
  125 S=0.0
      GO TO 140
  130 WRITE (IWR,135)
  135 FORMAT (//1H0,29X,74H**********NO CHI SQUARE WAS COMPUTED DEGREES
     1OF FREEDOM ARE ZERO**********)
      GO TO 190
  140 DO 145 I=1,KA
      CHISQ(I)=(AOBS(I)-AEXP(I))**2/AEXP(I)
  145 S=S+(AOBS(I)-AEXP(I))**2/AEXP(I)
      WRITE (IWR,150)
  150 FORMAT (1H1,53X,25HCALCULATION OF CHI SQUARE)
      WRITE (IWR,155)
  155 FORMAT (//1H ,42X,46HCLASS      OBSERVED      EXPECTED      CHI SQUAR
     1E)
      DO 160 J=1,KA
      LOVE=J-1
  160 WRITE (IWR,165) LOVE,AOBS(J),AEXP(J),CHISQ(J)
  165 FORMAT (1H0,44X,I3,5X,F8.0,5X,F8.3,5X,F9.3)
      KB=KA-2
      IF (P-.05) 180,180,170
  170 WRITE (IWR,175) S,KB
  175 FORMAT (//1H0,36X,31HCHI SQUARE TO A BINOMIAL FIT = ,F9.3,5H WITH,
     11X,I3,3X,18HDEGREES OF FREEDOM)
      GO TO 190
  180 WRITE (IWR,185) S,KB
  185 FORMAT (//1H0,36X,30HCHI SQUARE TO A POISSON FIT = ,F9.3,5H WITH,1
     1X,I3,3X,18HDEGREES OF FREEDOM)
  190 CONTINUE
      RETURN
      END
```

```
      SUBROUTINE NEGBI
C     ********************************************************************
C     SUBROUTINE TO CALCULATE THE NEGATIVE BINOMIAL
C     ********************************************************************
C     ********************************************************************
C     FIRST K = (AVERAGE)**2/(VARIANCE - AVERAGE)
C     Z VALUE = SUM(CUMULATIVE FREQUENCY/FIRST K + CLASS) - (SUM OF
C     FREQUENCIES * (LN(1.0 + (AVERAGE/ESTIMATE OF K(L)))))
C     WHERE Z(L) = S(AX(J)/EK(L) + X(J)) - NLOG(1.0 + AVEXOEK(L))
C     DERIVATIVE DZ = S(AX(J)/((EK(L) + X(J)**2) - N*(AVEX)/EK(L)*(EK(L)
C      + AVEX))
C     THEN  EK(L+1) = EK(L) + Z(L)/DZ(L)
C     L=NUMBER OF Z AND DZ VALUES CALCULATED
C     R = AVERAGE/(EK + AVERAGE)
C     P = AVERAGE/EK
C     Q = 1.0 + P
C     ********************************************************************
      COMMON X(60),Y(60),TITLE(20),FMT(20),NCLAS,BIGN,SMLLN,AVEX,SF,VAR
      COMMON IRDO,IWR
      DIMENSION PHI(60),CHI(60),ACC(60),PROB(60),EK(60),Z(60),DER(60)
      DIMENSION DZ(60)
      WRITE (IWR,5) TITLE
    5 FORMAT (1H1,25X,20A4//)
      WRITE (IWR,10)
   10 FORMAT (1H0,38X,55H*****CALCULATION OF THE EXPECTED NEGATIVE BINOM
     1IAL*****)
      L=1
      EK(L)=((AVEX**2)/(VAR-AVEX))
      WRITE (IWR,15)
   15 FORMAT (//1H ,62X,7HFIRST K)
      WRITE (IWR,20) EK(L)
   20 FORMAT (1H ,58X,F11.6//)
      KEY=0
      WRITE (IWR,25)
   25 FORMAT (1H ,55X,7HZ VALUE,7X,8HDZ VALUE)
      DO 125 L=1,60
      SY=0.0
      SACC=0.0
      SDER=0.0
      DO 30 J=1,NCLAS
      SY=SY+Y(J)
      ACC(J)=((SF-SY)/(EK(L)+X(J)))
      SACC=SACC+ACC(J)
      DER(J)=((SF-SY)/((EK(L)+X(J))**2))
      SDER=SDER+DER(J)
   30 CONTINUE
      PT2=(SF*(ALOG(1.0+(AVEX/EK(L)))))
      Z(L)=SACC-PT2
      PT3=((SF*AVEX)/(EK(L)*(EK(L)+AVEX)))
      DZ(L)=SDER-PT3
      WRITE (IWR,35) Z(L),DZ(L)
   35 FORMAT (1H ,51X,F12.6,3X,F14.8)
      KEY=KEY+1
      GO TO (40,50,90,65,105),KEY
   40 IF (Z(L)) 45,130,85
   45 EK(L+1)=EK(L)-0.1
      GO TO 125
   50 IF (Z(L)) 55,130,60
   55 KEY=KEY-1
      GO TO 45
   60 EK(L+1)=EK(L)+(Z(L)/DZ(L))
```

```
      KEY=KEY+1
      GO TO 125
   65 IF (Z(L)) 70,130,75
   70 KEY=KEY-2
      GO TO 100
   75 IF (Z(L)-0.00005) 130,130,80
   80 KEY=KEY-2
      GO TO 60
   85 EK(L+1)=EK(L)+0.1
      KEY=KEY+1
      GO TO 125
   90 IF (Z(L)) 1)0,130,95
   95 KEY=KEY-2
      GO TO 85
  100 EK(L+1)=EK(L)+(Z(L)/DZ(L))
      KEY=KEY+1
      GO TO 125
  105 IF (Z(L)) 115,130,110
  110 KEY=KEY-2
      GO TO 60
  115 IF (Z(L)+0.00005) 120,130,130
  120 KEY=KEY-2
      GO TO 100
  125 CONTINUE
  130 WRITE (IWR,135)
  135 FORMAT(1H0,42X,20HMAXIMUM LIKELIHOOD K,8X,7HFINAL Z,8X,1HL)
      WRITE (IWR,140) EK(L),Z(L),L
  140 FORMAT (1H ,45X,F14.8,6X,F14.8,4X,I3//)
      R=AVEX/(EK(L)+AVEX)
      P=AVEX/EK(L)
      Q=1.0+P
      WRITE (IWR,145)
  145 FORMAT (1H1)
      WRITE (IWR,150) TITLE
  150 FORMAT (1H ,25X,20A4//)
      WRITE (IWR,155)
  155 FORMAT (1H ,41X,14H       R         ,3X,14H       P         ,3X,14H
     1  Q        )
      WRITE (IWR,160) R,P,Q
  160 FORMAT (1H ,41X,F14.8,3X,F14.8,3X,F14.8//)
      J=1
      PHI(J)=(SF/Q**EK(L))
  165 J=J+1
      SPHI=0.0
      DO 170 J=2,NCLAS
      PHI(J)=((((EK(L)+X(J)-1.0)*R)/X(J))*PHI(J-1))
      SPHI=SPHI+PHI(J)
  170 CONTINUE
      SPROB=0.0
      DO 175 J=1,'ICLAS
      PROB(J)=PHI(J)/SF
      SPROB=SPROB+PROB(J)
  175 CONTINUE
      STOT=PHI(1)+SPHI
      DEF1=SF-STOT
      DEF2=1.0-SPROB
      IF (DEF1-0.2) 195,195,180
  180 NOP=NCLAS+5
      LKV=NCLAS+1
      DO 185 J=LKV,NOP
```

```
      Y(J)=0.0
  185 CONTINUE
      NCLAS=NCLAS+5
      DO 190 J=1,NCLAS
      X(J+1)=X(J)+1.0
  190 CONTINUE
      J=1
      GO TO 165
  195 CONTINUE
      AAZ=0.0
      DO 200 J=1,NCLAS
      CHI(J)=(Y(J)-PHI(J))**2/PHI(J)
      AAZ=AAZ+CHI(J)
  200 CONTINUE
      WRITE (IWR,205)
  205 FORMAT (1H ,18X,5HCLASS,8X,18HEXPECTED FREQUENCY,8X,18HOBSERVED FR
     1EQUENCY,8X,10HCHI SQUARE,8X,11HPROBABILITY)
      DO 215 J=1,NCLAS
      WRITE (IWR,210) X(J),PHI(J),Y(J),CHI(J),PROB(J)
  210 FORMAT (1H ,18X,F5.0,12X,F10.4,16X,F10.4,14X,F8.3,8X,F11.8)
  215 CONTINUE
      WRITE (IWR,220) STOT,SF,AAZ,SPROB
  220 FORMAT (1H0,17X,6HTOTALS,12X,F10.4,16X,F10.4,14X,F8.3,8X,F11.8)
      IVAN=NCLAS-3
      WRITE (IWR,225) IVAN
  225 FORMAT (/1H ),17X,4HWITH,I4,1X,18HDEGREES OF FREEDOM)
      WRITE (IWR,145)
      RETURN
      END
```

Input

```
    4
    5    5      4800.          5.
THE DISTRIBUTION OF SPHENE IN THE RENSSELAER GRAYWACKE RENSSELAER COUNTY N.Y.
(2F5.0)
   0.4208.
   1. 474.
   2.  98.
   3.  19.
   4.   1.
    8    6       156.          7.
THE DISTRIBUTION OF QUARTZ IN THE COW RUN SANDSTONE,ST.MARYS WEST VIRGINIA
(2F5.0)
   0.   2.
   1.  10.
   2.  26.
   3.  45.
   4.  39.
   5.  21.
   6.  11.
   7.   2.
    6    4      4800.          5.
THE DISTRIBUTION OF EPIDOTE IN THE RENSSELAER GRAYWACKE RENSSELAER COUNTY N.Y.
(2F10.0)
        0.      2797.
        1.      1288.
        2.       501.
        3.       155.
        4.        47.
        5.        12.
    6    5       401.          5.
   AREAL (GRID) SAMPLING OF THE STATE OF KANSAS FOR NATURAL RESOURCE TARGETS
(2F5.0)
   0. 257.
   1. 104.
   2.  29.
   3.   8.
   4.   2.
   5.   1.
```

Output

**********DISCRETE DISTRIBUTIONS PROGRAM**********

THE DISTRIBUTION OF SPHENE IN THE RENSSELAER GRAYWACKE RENSSELAER COUNTY N.Y.

AVERAGE	VARIANCE	STANDARD DEVIATION
0.152	0.196	0.4429

CALCULATION OF THE EXPECTED POISSON

CLASS	EXPECTED PROBABILITY	EXPECTED FREQUENCY	OBSERVED FREQUENCY
0	0.85875133	4122.0064	4208.0000
1	0.13078067	627.7472	474.0000
2	0.00995840	47.8003	98.0000
3	0.00050552	2.4265	19.0000
4	0.00001924	0.0923	1.0000
TOTALS	1.00001518	4800.0729	4800.0000

CALCULATION OF CHI SQUARE

CLASS	OBSERVED	EXPECTED	CHI SQUARE
0	4208.	4122.006	1.794
1	474.	627.747	37.655
2	98.	47.800	52.719
3	20.	2.518	121.317

CHI SQUARE TO A POISSON FIT = 213.486 WITH 2 DEGREES OF FREEDOM

THE DISTRIBUTION OF SPHENE IN THE RENSSELAER GRAYWACKE RENSSELAER COUNTY N.Y.

*****CALCULATION OF THE EXPECTED NEGATIVE BINOMIAL*****

FIRST K
0.528422

Z VALUE	DZ VALUE
-9.869441	141.61221712
13.035468	348.45330071
1.999516	247.70122355
0.070945	230.34404194
0.000097	229.70763027
0.000002	229.70675677

MAXIMUM LIKELIHOOD K	FINAL Z	L
0.47421228	0.00000286	6

THE DISTRIBUTION OF SPHENE IN THE RENSSELAER GRAYWACKE RENSSELAER COUNTY N.Y.

R	P	Q
0.24308173	0.32114660	1.32114660

CLASS	EXPECTED FREQUENCY	OBSERVED FREQUENCY	CHI SQUARE	PROBABILITY
0.	4206.1502	4208.0000	0.000	0.87628130
1.	484.8527	474.0000	0.242	0.10101099
2.	86.8744	98.0000	1.424	0.01809885
3.	17.4164	19.0000	0.143	0.00362843
4.	3.6771	1.0000	1.949	0.00076606
5.	0.7998	0.0000	0.799	0.00016663
6.	0.1773	0.0000	0.177	0.00003695
7.	0.0398	0.0000	0.039	0.00000830
8.	0.0090	0.0000	0.009	0.00000188
9.	0.0020	0.0000	0.002	0.00000043
TOTALS	4799.9993	4800.0000	4.789	0.99999987

WITH 7 DEGREES OF FREEDOM

**********DISCRETE DISTRIBUTIONS PROGRAM**********

THE DISTRIBUTION OF QUARTZ IN THE COW RUN SANDSTONE,ST.MARYS WEST VIRGINIA

AVERAGE	VARIANCE	STANDARD DEVIATION
3.449	1.965	1.4019

CALCULATION OF THE EXPECTED POSITIVE BINOMIAL

CLASS	EXPECTED PROBABILITY	EXPECTED FREQUENCY	OBSERVED FREQUENCY
0	0.00864988	1.3494	2.0000
1	0.05880042	9.1729	10.0000
2	0.17130664	26.7238	26.0000
3	0.27726526	43.2534	45.0000
4	0.26925760	42.0042	39.0000
5	0.15688873	24.4746	21.0000
6	0.05078588	7.9226	11.0000
7	0.00704559	1.0991	2.0000
TOTALS	1.00000001	156.0000	156.0000

CALCULATION OF CHI SQUARE

CLASS	OBSERVED	EXPECTED	CHI SQUARE
0	2.	1.349	0.314
1	10.	9.173	0.075
2	26.	26.724	0.020
3	45.	43.253	0.071
4	39.	42.004	0.215
5	21.	24.475	0.493
6	11.	7.923	1.195
7	2.	1.099	0.738

CHI SQUARE TO A BINOMIAL FIT = 3.120 WITH 6 DEGREES OF FREEDOM

**********DISCRETE DISTRIBUTIONS PROGRAM**********

THE DISTRIBUTION OF EPIDOTE IN THE RENSSELAER GRAYWACKE RENSSELAER COUNTY N.Y.

AVERAGE	VARIANCE	STANDARD DEVIATION
0.625	0.804	0.8967

THE DISTRIBUTION OF EPIDOTE IN THE RENSSELAER GRAYWACKE RENSSELAER COUNTY N.Y.

*****CALCULATION OF THE EXPECTED NEGATIVE BINOMIAL*****

FIRST K
2.191608

Z VALUE	DZ VALUE
-3.051632	15.51919103
-1.301038	19.65296126
0.915579	24.88831378
0.038703	22.81491234
0.000079	22.72366954
0.000000	22.72348310

MAXIMUM LIKELIHOOD K	FINAL Z	L
2.03009546	0.00000000	6

THE DISTRIBUTION OF EPIDOTE IN THE RENSSELAER GRAYWACKE RENSSELAER COUNTY N.Y.

R	P	Q
0.23557637	0.30817516	1.30817516

CLASS	EXPECTED FREQUENCY	OBSERVED FREQUENCY	CHI SQUARE	PROBABILITY
0.	2782.2639	2797.0000	0.078	0.57963831
1.	1330.5969	1288.0000	1.363	0.27720769
2.	474.9026	501.0000	1.434	0.09893804
3.	150.2901	155.0000	0.147	0.03131043
4.	44.5223	47.0000	0.137	0.00927549
5.	12.6492	12.0000	0.033	0.00263525
6.	3.4914	0.0000	3.491	0.00072738
7.	0.9435	0.0000	0.943	0.00019657
8.	0.2508	0.0000	0.250	0.00005227
9.	0.0658	0.0000	0.065	0.00001372
10.	0.0171	0.0000	0.017	0.00000356
TOTALS	4799.9940	4800.0000	7.963	0.99999876

WITH 8 DEGREES OF FREEDOM

**********DISCRETE DISTRIBUTIONS PROGRAM**********

AREAL GRID SAMPLING OF THE STATE OF KANSAS FOR NATURAL RESOURCE TARGETS

AVERAGE	VARIANCE	STANDARD DEVIATION
0.496	0.624	0.7899

CALCULATION OF THE EXPECTED POSITIVE BINOMIAL

CLASS	EXPECTED PROBABILITY	EXPECTED FREQUENCY	OBSERVED FREQUENCY
0	0.59294832	237.7722	257.0000
1	0.32667972	130.9985	104.0000
2	0.07199254	28.8690	29.0000
3	0.00793273	3.1810	8.0000
4	0.00043704	0.1752	2.0000
5	0.00000963	0.0038	1.0000
TOTALS	1.00000000	401.0000	401.0000

CALCULATION OF CHI SQUARE

CLASS	OBSERVED	EXPECTED	CHI SQUARE
0	257.	237.772	1.554
1	104.	130.998	5.564
2	29.	28.869	0.000
3	11.	3.360	17.370

CHI SQUARE TO A BINOMIAL FIT = 24.490 WITH 2 DEGREES OF FREEDOM

AREAL GRID SAMPLING OF THE STATE OF KANSAS FOR NATURAL RESOURCE TARGETS

*****CALCULATION OF THE EXPECTED NEGATIVE BINOMIAL*****

FIRST K
1.927149

Z VALUE	DZ VALUE
0.083118	1.69696223
-0.067420	1.32904809
0.004351	1.50424724
0.000015	1.49365490
0.000000	1.49361795
-0.000000	1.49361777

MAXIMUM LIKELIHOOD K	FINAL Z	L
1.97932439	-0.00000011	6

AREAL GRID SAMPLING OF THE STATE OF KANSAS FOR NATURAL RESOURCE TARGETS

R	P	Q
0.20046154	0.25072158	1.25072158

CLASS	EXPECTED FREQUENCY	OBSERVED FREQUENCY	CHI SQUARE	PROBABILITY
0.	257.5324	257.0000	0.001	0.64222552
1.	102.1833	104.0000	0.032	0.25482123
2.	30.5139	29.0000	0.075	0.07609471
3.	8.1136	8.0000	0.001	0.02023362
4.	2.0246	2.0000	0.000	0.00504911
5.	0.4853	1.0000	0.545	0.00121039
TOTALS	400.8534	401.0000	0.656	0.99963461

WITH 3 DEGREES OF FREEDOM

KANSAS GEOLOGICAL SURVEY COMPUTER PROGRAM
THE UNIVERSITY OF KANSAS, LAWRENCE

PROGRAM ABSTRACT

Title (If subroutine state in title):

FORTRAN IV Computer Program for Fitting Observed Count Data to Discrete Distribution Models of Binomial, Poisson and Negative Binomial

Date: December 1968

Author, organization: Charles W. Ondrick, Kansas Geological Survey, and John C. Griffiths
The Pennsylvania State University

Direct inquiries to: Authors or

Name: D.F.Merriam Address: Kansas Geological Survey
Kansas University, Lawrence, Kansas 66044

Purpose/description: Fits observed count data to Poisson, positive binomial and negative binomial discrete distribution models. The observed distribution is compared to the theoretical by the method of Chi square.

Mathematical method:

Restrictions, range:

Computer manufacturer: GE or IBM Model: 635 or 360/67

Programming language: FORTRAN IV

Memory required: K Approximate running time:

Special peripheral equipment required: none

Remarks (special compilers or operating systems, required word lengths, number of successful runs, other machine versions, additional information useful for operation or modification of program)

Program has been run successfully on the GE635, IBM 360/Models 50 and 67 and IBM 1130 computers.

COMPUTER CONTRIBUTIONS

1. Mathematical simulation of marine sedimentation with IBM 7090/7094 computers, by J.W.Harbaugh, 1966.
2. A generalized two-dimensional regression procedure, by J.R.Dempsey, 1966.
3. FORTRAN IV and MAP program for computation and plotting of trend surfaces for degree 1 through 6, by Mont O'Leary, R.H.Lippert, and O.T.Spitz, 1966.
4. FORTRAN II program for multivariate discriminant analysis using an IBM 1620 computer, by J.C.Davis and R.J.Sampson, 1966.
5. FORTRAN IV program using double Fourier series for surface fitting of irregularly spaced data, by W.R. James, 1966.
6. FORTRAN IV program for estimation of cladistic relationships using the IBM 7040, by R.L.Bartcher, 1966.
7. Computer applications in the earth sciences: Colloquium on classification procedures, edited by D.F. Merriam, 1966.
8. Prediction of the performance of a solution gas drive reservoir by Muskat's Equation, by Apolonio Baca, 1967.
9. FORTRAN IV program for mathematical simulation of marine sedimentation with IBM 7040 or 7094 computers, by J.W.Harbaugh and W.J.Wahlstedt, 1967.
10. Three-dimensional response surface program in FORTRAN II for the IBM 1620 computer, by R.J.Sampson and J.C.Davis, 1967.
11. FORTRAN IV program for vector trend analyses of directional data, by W.T.Fox, 1967.
12. Computer applications in the earth sciences: Colloquium on trend analysis, edited by D.F.Merriam and N.C.Cocke, 1967.
13. FORTRAN IV computer programs for Markov chain experiments in geology, by W.C.Krumbein, 1967.
14. FORTRAN IV programs to determine surface roughness in topography for the CDC 3400 computer, by R.D. Hobson, 1967.
15. FORTRAN II program for progressive linear fit of surfaces on a quadratic base using an IBM 1620 computer, by A.J.Cole. C. Jordan, and D.F.Merriam, 1967.
16. FORTRAN IV program for the GE 625 to compute the power spectrum of geological surfaces, by J.E.Esler and F.W.Preston, 1967.
17. FORTRAN IV program for Q-mode cluster analysis of nonquantitative data using IBM 7090/7094 computers, by G.F.Bonham-Carter, 1967.
18. Computer applications in the earth sciences: Colloquium on time-series analysis, D.F.Merriam, editor, 1967.
19. FORTRAN II time-trend package for the IBM 1620 computer, by J.C.Davis and R.J.Sampson, 1967.
20. Computer programs for multivariate analysis in geology, D.F.Merriam, editor, 1968.
21. FORTRAN IV program for computation and display of principal components, by W.J.Wahlstedt and J.C. Davis, 1968.
22. Computer applications in the earth sciences: Colloquium on simulation, D.F.Merriam and N.C.Cocke, editors, 1968.
23. Computer programs for automatic contouring, by D.B.McIntyre, D.D.Pollard, and R. Smith, 1968.
24. Mathematical model and FORTRAN IV program for computer simulation of deltaic sedimentation, by G.F. Bonham-Carter and A.J.Sutherland, 1968.
25. FORTRAN IV CDC 6400 computer program for analysis of subsurface fold geometry, by E.H.T.Whitten, 1968.
26. FORTRAN IV computer program for simulation of transgression and regression with continuous-time Markov models, by W.C.Krumbein, 1968.
27. Stepwise regression and nonpolynomial models in trend analysis, by A.T.Miesch and J.J.Connor, 1968.
28. KWIKR8 a FORTRAN IV program for multiple regression and geologic trend analysis, by J.E.Esler, P.F. Smith, and J.C.Davis, 1968.
29. FORTRAN IV program for harmonic trend analysis using double Fourier series and regularly gridded data for the GE 625 computer, by J.W.Harbaugh and M.J.Sackin, 1968.
30. Sampling a geological population (workshop on experiment in sampling), by J.C.Griffiths and C.W.Ondrick, 1968.
31. Multivariate procedures and FORTRAN IV program for evaluation and improvement of classifications, by Ferruh Demirmen, 1969.
32. FORTRAN IV programs for canonical correlation and canonical trend-surface analysis, by P.J.Lee, 1969.
33. FORTRAN IV program for construction of Pi diagrams with the Univac 1108 computer, by Jeffrey Warner, 1969.
34. FORTRAN IV program for nonlinear estimation, by R.B.McCammon, 1969.
35. FORTRAN IV computer program for fitting observed count data to discrete distribution models of binomial, Poisson and negative binomial, by C.W.Ondrick and J.C.Griffiths, 1969.