UM-4

POLYNOMIAL REGRESSION

SATISH CHANDRA
DIRECTOR

STUDY GROUP

S M SETH

N K GOEL

NATIONAL INSTITUTE OF HYDROLOGY

JAL VIGYAN BHAVAN

ROORKEE-247667

1984-85

CONTENTS

ABSTRACT

For any non linear function Y=f(X) regression may be obtained by fitting a polynomial. The general form of the polynomial regression is as given under:-

$$Y = a_o + a_1X + a_2X^2 + \ldots\ldots\ldots + a_mX^m + \varepsilon$$

where, Y is a dependent variable.

The coefficients a_o, a_1, ..., a_m are the regression coefficients and are determined by the least square method of parameter estimation. The power order m is cho sen so as to minimize the sum of squres of deviations from the line. The user's manual gives the details of a computer programme for polynomial regression. In the programme powers of an independent variable are generated to calculate polynomials of successively increasing degrees. If there is no reduction in the residual sum of squares between two successive degrees of polynimals, the programme terminates the problem before completing the analysis for the highest degree polynomial specified. The output of the program includes regression coefficients for successive degree polynomials, analysis of variance table for successive degree polynomials, table of residuals for the final degree polynomial and plot of Y values and Y estimates versus base variable X.

This manual also describes various statistics given in

the programme output with example, input data specifications and output description. The programme is written in FORTRAN IV. The manual also gives hardware and software requirements of the programme.

1.0 INTRODUCTION

Regression represents a mathematical equation expressing one random variable as being correlatively related to another random variable or several random variables. The regression equation may be any function that can be fitted to a set of observed variables. If the variables are linearaly related, then the regression is called linear regression. In non linear regression, the variables are not linearly related.

For any non linear function Y = F(X) regression may be obtained by fitting a polynomial. The general form of the polynomial regression is as given under :

$$Y = a_o + a_1X + a_2X^2 + \ldots\ldots. + a_mx^m + \varepsilon \qquad \ldots(1)$$

The coefficients $a_o, a_1, \ldots., a_m$ are the regression coefficients and are determined by the least square method of parameter estimation. The power order m is chosen so as to minimize the sum of the squares of deviations from the line. The power m should be much lower than the sample size N in order to have a sufficient number of degrees of freedom (N-m-1) and to have a reliable estimate of the standard deviation. Generally the value of m is between 2 and 4 as it is very difficult to physically explain higher degree of m.

It is preferable to plot the data first to have a preliminary idea about the value of m or the degree of polynomial to be filled to the data.

1.1 Purpose and Capabilities

The programme calls various subroutines to perform the regression analysis. The programme prints the regression coefficients and analysis of variance tables for polynomials of successively increasing degrees. The programme also optionally prints the table of residuals and a plot of Y values and Y. estimates.

The programme for polynomial regression can cater upto 100 observations and 10th degree polynomial. To handle more observations, the dimensions in the main programme should be changed.

1.2 Definitions of the Terminology Related to the Topic

a. Sum of squares due to regression:

This represents the portion of variance due to regression. This is given by following equation:-

$$SSDR = \sum_{i=1}^{N} (Y_i' - \bar{y})^2 \qquad \ldots(2)$$

where,

$SSDR$: sum of squares due to regression

Y_i' : Estimated value of dependent variable

$\bar{Y}$: Mean of dependent variable

N : Total number of observations

The number of degrees of freedom (NDF) will be equal to the degree of polynomial.

b. Mean squares due to regression :

Mean squares due to regression is the ratio of sum of squares due to regression and number of degrees of freedom.

$$MSDR = \frac{SSDR}{NDF} \qquad \ldots(3)$$

where

MSDR : Mean squares due to regression

NDF : Number of degrees of freedom

c. Sum of squares from regression :

Sum of squares from regression is the portion of variance which is due to deviation from regression.

$$SSFR = \sum_{i=1}^{N} (Y_i' - Y_i)^2 \qquad \ldots(4)$$

where,

SSFR : Sum of squares from regression

Y_i : Observed value of dependent variable

Y_i' : Estimated value of dependent variable

The number of degrees of freedom associated with it is equal to (N-m-1). Here m is the degree of polynomial.

d. Mean squares from regression :

Mean squares from regression is the ratio of sum of squares from regression and number of degrees of freedom.

$$MSFR = \frac{SSFR}{N-m-1} \qquad \ldots(5)$$

where,

MSFR : Mean squares from regression.

e. F Value:

F value is the ratio of mean squares due to regression and mean squares from regression.

$$\text{F value} = \frac{\text{MSDR}}{\text{MSFR}} \qquad \text{...(6)}$$

1.3 Scope

The programme for polynomial regression can be used for the analysis of trend in the hydrologic data.

1.4 Hardware and Software Requirements

Fortran compiler and simple fortran instructions are required to run the programme. The programme has been implemented and modified on DEC-2050 and VAX-11/780 system. However, the same programme can be used with little or no modifications on any other computer system also. The memory requirement depends upon the length of the data which will modify the dimension statement of the programme.

2.0 SPECIFIC METHOD

The regression coefficients are estimated by the method of least squares for the sum of squares of residuals. The sum of squares of residuals is given by the following equation :

$$\varepsilon^2 = \sum_{i=1}^{N} (Y_i' - Y_i)^2 \qquad \text{...(7)}$$

where,

Y_i' : Estimated i^{th} value of Y

Y_i : Observed i^{th} value of Y

Estimated i^{th} value of Y is given by the following equation for m degree polynomial.

$$Y_i' = a_o + a_1X_i + a_2X_i^2 + \text{.....} + a_mX_i^m$$

The N equations are :

$$Y_1' = a_o + a_1X_1 + a_2X_1^2 + \text{....} + a_mX_1^m$$

$$Y_2' = a_o + a_1X_2 + a_2X_2^2 + \text{....} + a_mX_2^m \qquad \text{...(8)}$$

$$Y_n' = a_o + a_1X_n + a_2X_n^2 + \text{....} + a_mX_n^m$$

The above equations can be written in matrix form as follows:

$$[Y] = [X] \; [A] \qquad \text{...(9)}$$

where,

$[Y]$ = Nx1 vector of observations

$[X]$ = Nxm matrix

$[A]$ = mx1 vector of unknown parameters

Equation (9) can be written as :

$$[X^T] [Y] = [X^T] [X] [A] \qquad \ldots(10)$$

Solution of equation (10) is obtained by pre-multiplying by $[X^T X]^{-1}$

$$[X^T X]^{-1} [X^T] [Y] = [X^T X]^{-1} [X] [A]$$

$$\text{or } [A] = [X^T X]^{-1} [X^T] [Y] \qquad \ldots(11)$$

Here $[A]$ is a vector of m regression coefficients.

2.1 General Description

2.1.1 Programme description

The computer programme for polynomial regression consists of the main routine named PREG and five other subroutines named GDATA for data matrix generation for polynomial regression, ORDER for rearrangement of intercorrelations, MINV for matrix inversion, MULTR for multiple linear regression and PLOT for plotting. Subroutines are described below:

A. Subroutine GDATA(N,M,X,XBAR,STD,D,SUMSQ)

This subroutine generates independent variables upto the m^{th} power (the highest degree polynomial specified) and calculates means, standard deviations, sums of cross products of deviations from means and product moment correlation coefficients. The calling arguments are :

N : Number of observations

M : The highest degree polynomial to be fitted

X : Input matrix(Nxm+1). When the subroutine is called data for the independent variables are stored in the first column of matrix X and data for the dependent variable are stored in the last column of the matrix. Upon returning to the calling routine, generated powers of the independent variable are stored in columns 2 through m

XBAR : Output vector of length m+1 containing means of independent and dependent variables

STD : Output vector of length m+1 containing standard deviations

D : Output matrix containing correlation coefficients

SUMSQ : Output vector of length m+1 containing sums of products of deviations from means of independent and dependent variables

B. Subroutine ORDER (M,R,NDEP,K, ISAVE,RX,RY)

The purpose of this subroutine is to construct from larger matrix of correlation coeffcients a subset of matrix of intercorrelations among independent variables and a vector of inter-correlations of independent variables with dependent variable. The calling arguments are:

M : Number of variables and order of matrix R

R : Input matrix containing correlation coefficients

NDEP : The subscript number of the dependent variable

K : Number of independent variables to be included in the forthcoming regression, K must be greater than or equal to 1

ISAVE : Input vector of length K+1, containing, in ascending order, the subscript numbers of K independent variables to be included in the forthcoming regression, upon returning to the calling routine, this contains in addition, the subscript number of the dependent variable in K+1 position

RX : Output matrix (K x K) containing intercorrelations among independent variables to be used in forthcoming regression

RY : Output vector of length K containing intercorrelations of independent variables with dependent variable

C. Subroutine MINV (A,N,D,L,M)

The subroutine is used for matrix inversion. Various calling arguments are :

A : Input matrix destroyed in computation and replaced by resultant inverse

N : Order of matrix A

D : Resultant determinant

L : Work vector of length N

M : Work vector of length N

D. Subroutine MULTR(N,K,XBAR,STD,D,RX,RY,ISAVE, B,SB,T,ANS)

This subroutine performs a multiple linear regression analysis for a dependent variable and a set of independent variables. The calling arguments are :

N : Number of observations

K : Number of independent variables in regression

XBAR : Input vector of length m+1 containing means of all the variables

STD : Input vector of length m+1 containing standard deviations of all the variables

D : Input vector of length m+1 containing the diagonal of the matrix of sums of cross products of deviations from means for all variables

RX : Input matrix (K x K) containing the inverse of intercorrelations among independent variables

RY : Input vector of length K containing intercorrelation of independent variables with dependent variable

ISAVE : Input vector of length K+1 containing subscripts of independent variables in ascending order. The subscript of the dependent variable is stored in the last K+1 position

B : Output vector of length K containing regression coefficients

SB : Output vector of length K containing standard deviation of regression coefficients

T : Output vector of length K containing t values

ANS : Output vector of length 10 containing the following information:

ANS(1) : intercept

ANS(2) : multiple correlation coefficient

ANS(3) : standard error of estimate

ANS(4) : sum of squares attributable to regression (SSAR)

ANS(5) : degreesof freedome associated with SSAR

ANS(6) : mean squares of SSAR

ANS(7) : sum of squares of deviations from regression SSDR

ANS(8) : degrees of freedom associated with SSDR

ANS(9) : mean squares of SSDR

ANS(10) : F value

E. Subroutine PLOT (NO,A,N,M,NL,NS)

The purpose of this subroutine is to plot Y values and Y estimates versus base variable X. The calling arguments are:

NO	:	Chart number
A	:	Matrix of data to be plotted. First column represents base variable and successive columns are the cross variables (maximum 9)
N	:	Number of rows in A
M	:	Number of columns in matrix A (This is equal to the total number of variables, maximum is 10)
NL	:	Number of lines in the plot. If 0 is specified 50 lines are used
NS	:	Code for sorting the base variable data in ascending order 0 if sorting is not necessary 1 if sorting is necessary

2.1.2 Programme modifications

The programme capacity can be increased or decreased by making changes in dimension statements. The following are the general rules for the programme modifications.

(a) The dimension of array X must be greater than or equal to the product of N (m+1) where N is the number of observations and m is the highest degree polynomial to be fitted.

(b) The dimension of array DI must be greater than or equal to the product of (m x m).

(c) The dimension of array D must be greater than or equal to (m+2) (m+1)/2.

(d) The dimensions of arrays B,E,SB, and T must be greater than or equal to the highest degree polynomial to be fitted, m.

(e) The dimensions of arrays XBAR,STD,COE,SUMSQ and ISAVE must be greater than or equal to (m+1).

(f) The dimension of array P must be greater than or equal to 3XN.

2.2 Data Requirement

The data of independent and dependent variable for which the regression analysis is to be performed, are required. The pairs of independent and dependent variables should be complete without any gap.

2.3 Analysis

The analysis of 93 years stage data of river Narmada at Broach is given in subsequent sections :

The general form of the regression equation is :

$$Y = a_o + a_1X + a_2X^2 + \dots a_mX^m + \varepsilon$$

The highest degree of polynomial to be fitted to the stage data is 2. So the regression analysis will be performed for two cases.

Case I

$$Y = a_o + a_1X + \varepsilon$$

Case II

$$Y = a_o + a_1X + a_2X^2 + \varepsilon$$

Case I

(a) -Regression coefficients are calculated by method of least squares and are

24.6236 & 0.0611

(b) Sum of squares due to regression and sum of squares from regression are calculated by equation 2 and 4 and are 250.38 and 1615.39 respectively.

(c) Mean squares due to regression and mean squares from regression are calculated by equations 3 & 5 and are 250.38 and 17.75156 respectively.

(d) F value is given by equation 6. The F value is 14.10 .

(e) Improvement in terms of sum of squares is equal to the improvement in sum of squares due to regression, so this is equal to 250.38

For Case II the analysis is done in similar way.

2.4 Advantages and Limitations

2.4.1 Advantages

The program for polynomial regression can deal with upto 100 observations and 10^{th} degree polynomial. In the programme powers of an independent variable are generated to calculate polynomials of successively increasing degrees. If there is no reduction in the residual sum of squares between two successive degrees of polynomials, the programme terminates the problem before completing the analysis for the highest degree polynomial specified.

2.4.2 Limitations

1. The programme cannot deal with variables with missing data.
2. The degree of polynomials will always start from one.
3. The relationship developed should not be used for extrapolation as the confidence intervals on the regression line become very wide as the distance from $\bar{X}$ is increased. Secondly the relationship or the equation of regression line may be different outside the range.

2.5 Programme Details

The listing of the source programme is given in Appendix I. Input data specifications and output details have been described in Appendix II and III. The test data for which the programme has been run and corresponding data file and output file have been given in Appendix IV, V and VI respectively.

3.0 RECOMMENDATIONS

The programme for polynomial regression can deal with upto 100 observations and 10^{th} degree polynomial. If a problem is satisfying above two conditions then the programme can be used as such. However, if a problem is having more than 100 observations or if greater than 10^{th} degree polynomial is desired, the dimension statements must be changed according to the rules given for programme modification .

REFERENCES

1. Haan, C.T. (1977), ' Statistical Methods in Hydrology', Iowa State University Press, Ames, IOWA.

2. Ostle, B. (1963), 'Statistics in Research', Iowa State University Press, Ames, IOWA.

3. Scientific Subroutine Package, International Business Machines, White Plains, N.Y.

APPENDIX I

Polynomial Regression Programme

```
C       MASTER POLYNOMIAL REGRESSION
        DIMENSION X(1100),DI(100),D(66),B(10),E(10),SB(10),
        1T(10),XBAR(11),STD(11),COE(11),SUMSQ(11),ISAVE(11),
        2ANS(10),P(300),TITLE(80)
        OPEN(UNIT=5,FILE='PREG.DAT',STATUS='OLD')
        OPEN(UNIT=6,FILE='PREG.OUT',STATUS='NEW')
  1     FORMAT(I5,I2,I1)
  2     FORMAT(2F6.0)
  3     FORMAT(1X' POLYNOMIAL REGRESSION.......'/)
  4     FORMAT(/,' NUMBER OF OBSERVATIONS',I6/)
  5     FORMAT(/' POLYNOMIAL REGRESSION OF DEGREE',I3)
  6     FORMAT(' INTERCEPT',F20.4)
  7     FORMAT(/,' REGRESSION COEFFICIENTS'/(6F20.4))
  8     FORMAT(1H0/20X,' ANALYSIS OF VARIENCE FOR',I4,
        1'DEGREE OF POLYNOMIAL')
  9     FORMAT(1H0,1X,'SOURCE OF VARIATION',5X,'DEGREE OF'
        1,3X,'SUM OF',3X,4HMEAN,5X,1HF,4X,'IMPROVEMENT IN
        2TERMS'/26X ,'FREEDOM',4X,'SQUARES',3X,'SQUARE',3X,
        35HVALUE,1X,'OF SUM OF SQUARES')
 10     FORMAT('ODUE TO REGRESSION',7X,I6,F13.2
        1,F10.2,F7.2,F12.2)
 11     FORMAT(' DEVIAION ABOUT REGRESSION',I5,F13.2,F10.2)
 12     FORMAT(8X,'TOTAL',11X,I6,F13.2///)
 13     FORMAT('     NO IMPROVEMENT')
 14     FORMAT(1H0//27X,' TABLE OF RESIDUALS'//
        116 HOBSERVATION NO.,5X,7HX VALUE,7X,7HY VALUE,7X,
        210HY ESTIMATE,7X,8HRESIDUAL/)
 15     FORMAT(3X,I6,F18.5,F14.5,F17.5,F15.5)
 16     FORMAT(80A1)
100     READ(5,16)TITLE
        READ(5,1)N,M,NPLOT
        WRITE(6,16)TITLE
        WRITE(6,3)
        WRITE(6,4)N
C       READ INPUT DATA
        L=N*M
        DO 110I=1,N
        J=L+I
 110    READ(5,*) X(I),X(J)
        CALL GDATA(N,M,X,XBAR,STD,D,SUMSQ)
        MM=M+1
        SUM=0.0
        NT=N-1
        DO 200 I=1,M
        ISAVE(I)=I
        CALL ORDER(MM,D,MM,I,ISAVE,DI,E)
        CALL MINV(DI,I,DET,B,T)
        CALL MULTR(N,I,XBAR,STD,SUMSQ,DI,E,ISAVE,B,SB,T,ANS)
        WRITE(6,5) I
        IF(ANS(7)) 140,130,130
```

```
 130    SUMIP=ANS(4)-SUM
        IF(SUMIP) 140,140,150
 140    WRITE(6,13)
        GO TO 210
 150    WRITE(6,6)ANS(1)
        WRITE(6,7)(B(J),J=1,I)
        WRITE(6,8) I
        WRITE(6,9)
        SUM=ANS(4)
        WRITE(6,10) I,ANS(4),ANS(6),ANS(10),SUMIP
        NI=ANS(8)
        WRITE(6,11) NI,ANS(7),ANS(9)
        WRITE(6,12) NT,SUMSQ(MM)
        COE(1)=ANS(1)
        DO 160 J=1,I
 160    COE(J+1)=B(J)
        LA=I
 200    CONTINUE
C       TEST WHETHER PLOT IS REQUIRED OR NOT
 210    IF(NPLOT) 100,100,220
 220    NP3=N+N
        DO 230 I=1,N
        NP3=NP3+1
        P(NP3)=COE(1)
        L=I
        DO 230 J=1,LA
        P(NP3)=P(NP3)+X(L)*COE(J+1)
 230    L=L+N
        N2=N
        L=N*M
        DO 240 I=1,N
        P(I)=X(I)
        N2=N2+1
        L=L+1
 240    P(N2)=X(L)
        WRITE(6,3)
        WRITE(6,5) LA
        WRITE(6,14)
        NP2=N
        NP3=N+N
        DO 250 I=1,N
        NP2=NP2+1
        NP3=NP3+1
        RESID=P(NP2)-P(NP3)
 250    WRITE(6,15) I,P(I),P(NP2),P(NP3),RESID
        CALL PLOT(LA,P,N,3,0,0)
        STOP
        END
        SUBROUTINE ORDER(M,R,NDEP,K,ISAVE,RX,RY)
        DIMENSION R(1),ISAVE(1),RX(1),RY(1)
```

```
      MM=0
      DO 130 J=1,K
      L2=ISAVE(J)
      IF(NDEP-L2)122,123,123
122   L=NDEP+(L2*L2-L2)/2
      GO TO 125
123   L=L2+(NDEP*NDEP-NDEP)/2
125   RY(J)=R(L)
      DO 130 I=1,K
      L1=ISAVE(I)
      IF(L1-L2)127,128,128
127   L=L1+(L2*L2-L2)/2
      GO TO 129
128   L=L2+(L1*L1-L1)/2
129   MM=MM+1
130   RX(MM)=R(L)
      ISAVE(K+1)=NDEP
      RETURN
      END
C
      SUBROUTINE MINV(A,N,D,L,M)
      DIMENSION A(1),L(1),M(1)
C     SEARCH FOR LARGEST ELEMENT
      D=1.0
      NK=-N
      DO 80 K=1,N
      NK=NK+N
      L(K)=K
      M(K)=K
      KK=NK+K
      BIGA=A(KK)
      DO 20 J=K,N
      IZ=N*(J-1)
      DO 20 I=K,N
      IJ=IZ+I
10    IF(ABS(BIGA)-ABS(A(IJ))) 15,20,20
15    BIGA=A(IJ)
      L(K)=I
      M(K)=J
20    CONTINUE
C     INTERCHANGE ROWS
      J=L(K)
      IF(J-K) 35,35,25
25    KI=K-N
      DO 30 I=1,N
      KI=KI+N
      HOLD=-A(KI)
      JI=KI-K+J
      A(KI)=A(JI)
30    A(JI)=HOLD
```

```
C       INTERCHANGE COLUMNS
35      I=M(K)
        IF(I-K)45,45,38
38      JP=N*(I-1)
        DO 40 J=1,N
        JK=NK+J
        JI=JP+J
        HOLD=-A(JK)
        A(JK)=A(JI)
40      A(JI)=HOLD
C       DIVIDE COLUMNS BY MINUS PIVOT
45      IF(BIGA)48,46,48
46      D=0.0
        RETURN
48      DO 55 I=1,N
        IF(I-K)50,55,50
50      IK=NK+I
        A(IK)=A(IK)/(-BIGA)
55      CONTINUE
C       REDUCE MATRIX
        DO 65 I=1,N
        IK=NK+I
        HOLD=A(IK)
        IJ=I-N
        DO 65 J=1,N
        IJ=IJ+N
        IF(I-K)60,65,60
60      IF(J-K)62,65,62
62      KJ=IJ-I+K
        A(IJ)=HOLD*A(KJ)+A(IJ)
65      CONTINUE
C       DIVIDE ROW BY PIVOT
        KJ=K-N
        DO 75 J=1,N
        KJ=KJ+N
        IF(J-K)70,75,70
70      A(KJ)=A(KJ)/BIGA
75      CONTINUE
C       PRODUCT OF PIVOTS
        D=D*BIGA
C       REPLACE PIVOT BY RECIPROCAL
        A(KK)=1.0/BIGA
80      CONTINUE
C       FINAL ROW AND COLUMN INTERCHANGE
        K=N
100     K=(K-1)
        IF(K)150,150,105
105     I=L(K)
        IF(I-K)120,120,108
108     JQ=N*(K-1)
```

```
      JR=N*(I-1)
      DO 110 J=1,N
      JK=JQ+J
      HOLD=A(JK)
      JI=JR+J
      A(JK)=-A(JI)
110   A(JI)=HOLD
120   J=M(K)
      IF(J-K)100,100,125
125   KI=K-N
      DO 130 I=1,N
      KI=KI+N
      HOLD=A(KI)
      JI=KI-K+J
      A(KI)=-A(JI)
130   A(JI)=HOLD
      GO TO 100
150   RETURN
      END
C
      SUBROUTINE MULTR(N,K,XBAR,STD,D,RX,RY,ISAVE,B,SB,T,ANS)
      DIMENSION XBAR(1),STD(1),RX(1),RY(1),D(1),ISAVE(1),
     1B(1),SB(1),T(1),ANS(1)
      MM=K+1
C     BETA WEIGHTS
      DO 100 J=1,K
100   B(J)=0.0
      DO 110 J=1,K
      L1=K*(J-1)
      DO 110 I=1,K
      L=L1+I
110   B(J)=B(J)+RY(I)*RX(L)
      RM=0.0
      BO=0.0
      L1=ISAVE(MM)
C     COEFFICIENTS OF DETERMINATION
      DO 120 I=1,K
      RM=RM+B(I)*RY(I)
C     REGRESSION COEFFICIENTS
      L=ISAVE(I)
      B(I)=B(I)*(STD(L1)/STD(L))
C     INTERCEPT
120   BO=BO+B(I)*XBAR(L)
      BO=XBAR(L1)-BO
C     SUM OF SQUARES ATTRIBUTABLE TO REGRESSION
      SSAR=RM*D(L1)
C     MULTIPLE CORRELATION COEFFICIENT
122   RM=SQRT(ABS(RM))
C     SUM OF SQUARES OF DEVIATIONS FROM REGRESSION
      SSDR=D(L1)-SSAR
```

```
C       VARIANCE OF ESSTIMATE
        FN=N-K-1
        SY=SSDR/FN
C       STANDARD DEVIATIONS OF REGRESSION COEFFICIENTS
        DO 130 J=1;K
        L1=K*(J-1)+J
        L=ISAVE(J)
125     SB(J)=SQRT(ABS((RX(L1)/D(L))*SY))
C       COMPUTED T-VALUES
130     T(J)=B(J)/SB(J)
C       STANDARD ERROR OF ESTIMATE
135     SY=SQRT(ABS(SY))
C       F VALUE
        FK=K
        SSARM=SSAR/FK
        SSDRM=SSDR/FN
        F=SSARM/SSDRM
        ANS(1)=BO
        ANS(2)=RM
        ANS(3)=SY
        ANS(4)=SSAR
        ANS(5)=FK
        ANS(6)=SSARM
        ANS(7)=SSDR
        ANS(8)=FN
        ANS(9)=SSDRM
        ANS(10)=F
        RETURN
        END
        SUBROUTINE PLOT(NO;A;N;M;NL;NS)
        DIMENSION OUT(101);YPR(11);ANS(9);A(300)
 1      FORMAT(1H1;60X;7H CHART ;I3//)
 2      FORMAT(1H ;F11.4;5X;101A1)
3       FORMAT(1H )
 4      FORMAT(10H 123456789)
 5      FORMAT(10A1)
 7      FORMAT(1H ;16X;101H
        1
        1
 8      FORMAT(1H0;9X;11F10.4)
        NLL=NL
        IF(NS)16;16;10
C       SORT BASE VARIABLE IN ASCENDING ORDER
 10     DO 15 I=1;N
        DO 14 J=1;N
        IF(A(I)-A(J)) 14;14;11
 11     L=I-N
        LL=J-N
        DO 12 K=1;M
        L=L+N
```

```
      LL=LL+N
      F=A(L)
      A(L)=A(LL)
 12   A(LL)=F
 14   CONTINUE
 15   CONTINUE
C TEST NLL
 16   IF(NLL) 20,18,20
 18   NLL=50
 20   WRITE(6,1) NO
      REWIND 13
      WRITE(13,4)
      REWIND 13
      READ(13,5) BLANK,(ANG(I),I=1,9)
      REWIND 13
C  FIND SCALE FOR BASE VARIABLE
      XSCAL=(A(N)-A(1))/(FLOAT(NLL-1))
C  FIND SCALE FOR CROSS VARIABLE
      M1=N+1
      YMIN=A(M1)
      YMAX=YMIN
      M2=M*N
      DO 40 J=M1,M2
      IF(A(J)-YMIN) 28,26,26
 26   IF(A(J)-YMAX) 40,40,30
 28   YMIN=A(J)
      GOTO 40
 30   YMAX=A(J)
 40   CONTINUE
      YSCAL=(YMAX-YMIN)/100.0
C  FIND BASE VARIABLE PRINT POSITION
      XB=A(1)
      L=1
      MY=M-1
      I=1
 45   F=I-1
      XPR=XB+F*XSCAL
      IF(A(L)-XPR) 50,50,70
 50   DO 55 IX=1,101
 55   OUT(IX)=BLANK
      DO 60 J=1,MY
      LL=L+J*N
      JP=((A(LL)-YMIN)/YSCAL)+1.0
      OUT(JP)=ANG(J)
 60   CONTINUE
      WRITE(6,2)XPR,(OUT(IZ),IZ=1,101)
      L=L+1
      GOTO 80
 70   WRITE(6,3)
 80   I=I+1
```

```
      IF(I-NLL)45,84,86
 84   XPR=A(N)
      GOTO 50
 86   WRITE(6,7)
      YPR(1)=YMIN
      DO 90 KN=1,9
 90   YPR(KN+1)=YPR(KN)+YSCAL*10.0
      YPR(11)=YMAX
      WRITE(6,8)(YPR(IP),IP=1,11)
      RETURN
      END
      SUBROUTINE GDATA(N,M,X,XBAR,STD,D,SUMSQ)
      DIMENSION X(1),XBAR(1),STD(1),D(1),SUMSQ(1)
      IF(M-1) 105,105,90
 90   L1=0
      DO 100 I=2,M
      L1=L1+N
      DO 100 J=1,N
      L=L1+J
      K=L-N
100   X(L)=X(K)*X(J)
105   MM=M+1
      DF=N
      L=0
      DO 115 I=1,MM
      XBAR(I)=0.
      DO 110 J=1,N
      L=L+1
110   XBAR(I)=XBAR(I)+X(L)
115   XBAR(I)=XBAR(I)/DF
      DO 130 I=1,MM
130   STD(I)=0.
      L=((MM+1)*MM)/2
      DO 150 I=1,L
150   D(I)=0.
      DO 170 K=1,N
      L=0
      DO 170 J=1,MM
      L2=N*(J-1)+K
      T2=X(L2)-XBAR(J)
      STD(J)=STD(J)+T2
      DO 170 I=1,J
      L1=N*(I-1)+K
      T1=X(L1)-XBAR(I)
      L=L+1
170   D(L)=D(L)+T1*T2
      L=0
      DO 175 J=1,MM
      DO 175 I=1,J
      L=L+1
```

```
175   D(L)=D(L)-STD(I)*STD(J)/DF
      L=0
      DO 180 I=1,MM
      L=L+I
      SUMSQ(I)=D(L)
180   STD(I)=SQRT(ABS(D(L)))
C     CALCULATE CORRELATION COEFFICIENT
      L=0
      DO 190 J=1,MM
      DO 190 I=1,J
      L=L+1
190   D(L)=D(L)/(STD(I)*STD(J))
C     CALCULATE STANDARD DEVIATION
      DF=SQRT(DF-1.0)
      DO 200 I=1,MM
200   STD(I)=STD(I)/DF
      RETURN
      END
```

APPENDIX II

Input Specifications

Data file contains control cards and data cards

(i) Control Cards

Card	Variable	Description	Format
First	TITLE	Title of the problem	A
Second	N	Number of observations	I5
	M	Highest degree polynomial to be fitted	I2
	NPLOT	Option code for plotting Y values and Y estimates versus base variable X 0 if plot is not required 1 if plot is required	I1

(ii) Data Cards

Input data are read into the computer one observation at a time i.e. each pair of X and Y values in free format.

APPENDIX III

Output Description

The output for the polynomial regression programme includes

(1) Regression coefficients for successive degree polynomials.

(2) Analysis of variance for successive degree polynomials.

(3) Table of residuals

(4) Plot of Y values and Y estimates versus base variable X.

APPENDIX IV

Test Data

The programme for polynomial regression has been run on 93 years (1887-1979) annual peak stage data of river Narmada at BROACH. This has been done in order to see the trend in the data. The stage data is given below :

Sl. No.	Year	Stage (m)	Sl. No.	Year	Stage (m)	Sl. No.	Year	Stage (m)
1.	1887	26.50	39.	1925	22.00	77.	1963	23.00
2.	1888	21.00	40.	1926	29.50	78.	1964	25.25
3.	1889	23.50	41.	1927	28.00	79.	1965	23.00
4.	1890	25.00	42.	1928	27.50	80.	1966	20.60
5.	1891	22.50	43.	1929	25.00	81.	1967	23.75
6.	1892	27.50	44.	1930	30.50	82.	1968	38.50
7.	1893	24.50	45.	1931	29.50	83.	1969	29.60
8.	1894	30.00	46.	1932	29.00	84.	1970	41.50
9.	1895	21.50	47.	1933	34.00	85.	1971	26.00
10.	1896	30.00	48.	1934	29.50	86.	1972	32.00
11.	1897	27.00	49.	1935	23.50	87.	1973	36.50
12.	1898	30.00	50.	1936	26.50	88.	1974	29.98
13.	1899	18.25	51.	1937	30.50	89.	1975	28.50
14.	1900	28.00	52.	1938	30.50	90.	1976	26.50
15.	1901	26.50	53.	1939	34.00	91.	1977	26.99
16.	1902	20.50	54.	1940	26.50	92.	1978	29.54
17.	1903	23.50	55.	1941	29.00	93.	1979	31.49
18.	1904	21.50	56.	1942	34.50			
19.	1905	30.70	57.	1943	25.50			
20.	1906	32.50	58.	1944	35.00			
21.	1907	25.00	59.	1945	33.00			
22.	1908	24.50	60.	1946	27.50			
23.	1909	17.00	61.	1947	31.50			
24.	1910	25.50	62.	1948	29.50			
25.	1911	20.50	63.	1949	29.00			
26.	1912	28.00	64.	1950	35.00			
27.	1913	33.00	65.	1951	23.50			
28.	1914	21.50	66.	1952	23.50			
29.	1915	24.50	67.	1953	33.00			
30.	1916	27.50	68.	1954	29.50			
31.	1917	27.00	69.	1955	28.50			
32.	1918	19.50	70.	1956	27.00			
33.	1919	32.50	71.	1957	27.50			
34.	1920	22.50	72.	1958	25.50			
35.	1921	24.50	73.	1959	33.00			
36.	1922	24.50	74.	1960	25.50			
37.	1923	30.00	75.	1961	25.20			
38.	1924	27.00	76.	1962	32.30			

APPENDIX V

TEST INPUT

93 YEARS STAGE DATA AT BROACH
00093021

1.	26.50
2.	21.00
3.	23.50
4.	25.00
5.	22.50
6.	27.50
7.	24.50
8.	30.00
9.	21.50
10.	30.00
11.	27.00
12.	30.00
13.	18.25
14.	28.00
15.	26.50
16.	20.50
17.	23.50
18.	21.50
19.	30.70
20.	32.50
21.	25.00
22.	24.50
23.	17.00
24.	25.50
25.	20.50
26.	28.00
27.	33.00
28.	21.50
29.	24.50
30.	27.50
31.	27.00
32.	19.50
33.	32.50
34.	22.50
35.	24.50
36.	24.50
37.	30.00
38.	27.00
39.	22.00
40.	29.50
41.	28.00
42.	27.50
43.	25.00
44.	30.50
45.	29.50
46.	29.00
47.	34.00
48.	29.50

49.	23.50
50.	26.50
51.	30.50
52.	30.50
53.	34.00
54.	26.50
55.	29.00
56.	34.50
57.	25.50
58.	35.00
59.	33.00
60.	27.50
61.	31.50
62.	29.50
63.	29.00
64.	35.00
65.	23.50
66.	23.50
67.	33.00
68.	29.50
69.	28.50
70.	27.00
71.	27.50
72.	25.50
73.	33.00
74.	25.50
75.	25.20
76.	32.30
77.	23.00
78.	25.25
79.	23.00
80.	20.60
81.	23.75
82.	38.50
83.	29.60
84.	41.50
85.	26.00
86.	32.00
87.	36.50
88.	29.98
89.	28.50
90.	26.50
91.	26.99
92.	29.54
93.	31.49

APPENDIX VI

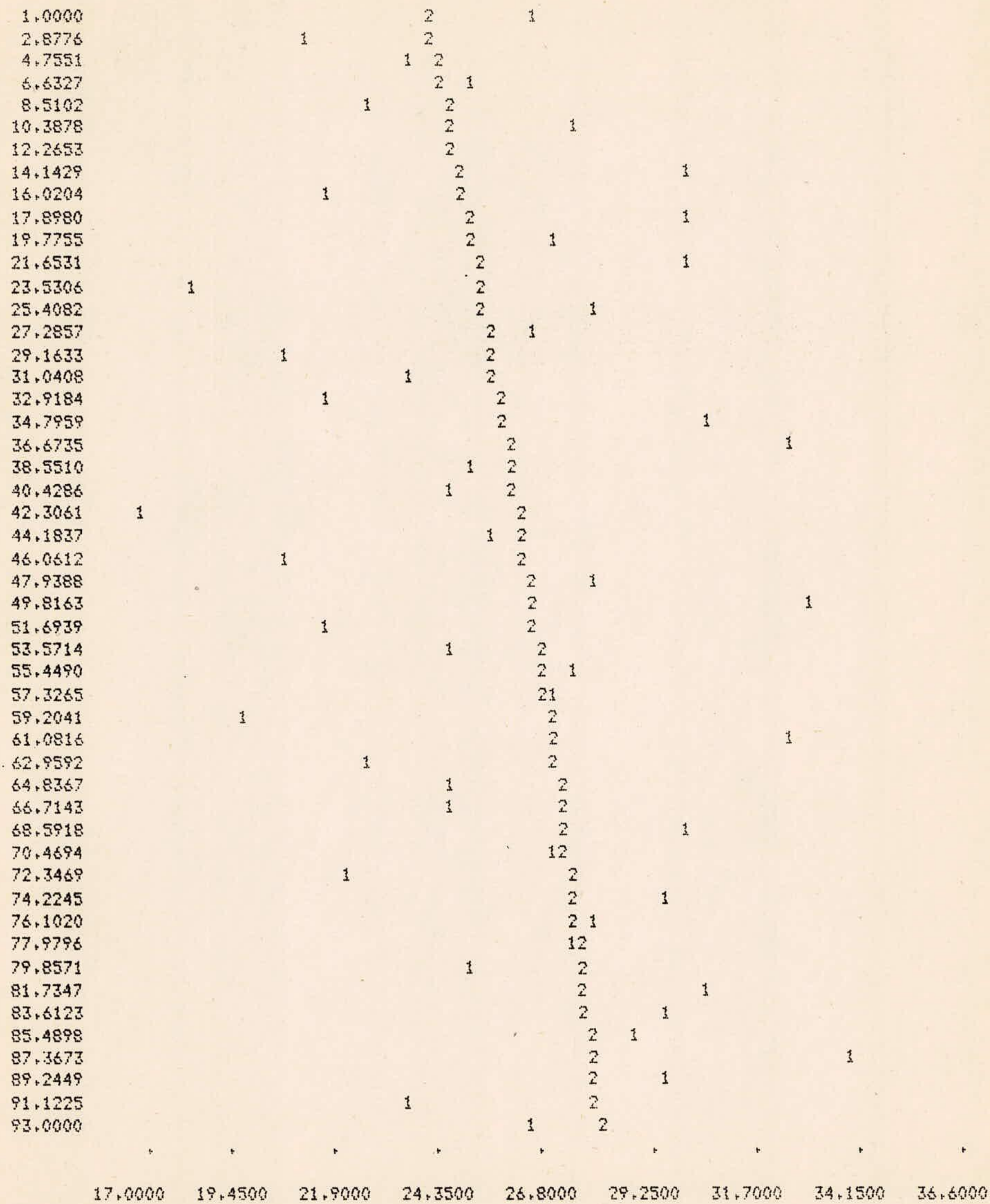
TEST OUTPUT

```
 93 YEARS STAGE DATA AT BROACH
  POLYNOMIAL REGRESSION......


 NUMBER OF OBSERVASIONS    93


 POLYNOMIAL REGRESSION OF DEGREE  1
 INTERCEPT              24.6236

 REGRESSION COEFFICIENTS
              0.0611
0
                     ANALYSIS OF VARIENCE FOR   1DEGREE OF POLYNOMIAL
0 SOURCE OF VARIATION      DEGREE OF    SUM OF    MEAN      F    IMPROVEMENT INTERMS
                           FREEDOM      SQUARES   SQUARE   VALUE OF SUM OF SQUARES
0DUE TO REGRESSION            1          250.38    250.38  14.10       250.38
 DEVIAION ABOUT REGRESSION   91         1615.39     17.75
       TOTAL                 92        1865.77



 POLYNOMIAL REGRESSION OF DEGREE  2
 INTERCEPT              23.8397

 REGRESSION COEFFICIENTS
              0.1106             -0.0005
0
                     ANALYSIS OF VARIENCE FOR   2DEGREE OF POLYNOMIAL
0 SOURCE OF VARIATION      DEGREE OF    SUM OF    MEAN      F    IMPROVEMENT INTERMS
                           FREEDOM      SQUARES   SQUARE   VALUE OF SUM OF SQUARES
0DUE TO REGRESSION            2          261.09    130.55   7.32        10.71
 DEVIAION ABOUT REGRESSION   90         1604.68     17.83
       TOTAL                 92        1865.77


  POLYNOMIAL REGRESSION......


 POLYNOMIAL REGRESSION OF DEGREE  2
0

                             TABLE OF RESIDUALS

OBSERVATION NO.,      X VALUE       Y VALUE        Y ESTIMATE       RESIDUAL

       1              1.00000      26.50000        23.94983         2.55017
       2              2.00000      21.00000        24.05888        -3.05888
```

3	3.00000	23.50000	24.16687	-0.66687
4	4.00000	25.00000	24.27381	0.72619
5	5.00000	22.50000	24.37969	-1.87969
6	6.00000	27.50000	24.48453	3.01547
7	7.00000	24.50000	24.58831	-0.08831
8	8.00000	30.00000	24.69103	5.30897
9	9.00000	21.50000	24.79271	-3.29271
10	10.00000	30.00000	24.89333	5.10667
11	11.00000	27.00000	24.99289	2.00711
12	12.00000	30.00000	25.09141	4.90859
13	13.00000	18.25000	25.18887	-6.93887
14	14.00000	28.00000	25.28527	2.71473
15	15.00000	26.50000	25.38063	1.11937
16	16.00000	20.50000	25.47493	-4.97493
17	17.00000	23.50000	25.56817	-2.06817
18	18.00000	21.50000	25.66037	-4.16037
19	19.00000	30.70000	25.75151	4.94849
20	20.00000	32.50000	25.84159	6.65841
21	21.00000	25.00000	25.93063	-0.93063
22	22.00000	24.50000	26.01861	-1.51861
23	23.00000	17.00000	26.10553	-9.10553
24	24.00000	25.50000	26.19141	-0.69141
25	25.00000	20.50000	26.27623	-5.77623
26	26.00000	28.00000	26.35999	1.64001
27	27.00000	33.00000	26.44271	6.55729
28	28.00000	21.50000	26.52437	-5.02437
29	29.00000	24.50000	26.60498	-2.10498
30	30.00000	27.50000	26.68453	0.81547
31	31.00000	27.00000	26.76303	0.23697
32	32.00000	19.50000	26.84048	-7.34048
33	33.00000	32.50000	26.91687	5.58313
34	34.00000	22.50000	26.99221	-4.49221
35	35.00000	24.50000	27.06650	-2.56650
36	36.00000	24.50000	27.13973	-2.63973
37	37.00000	30.00000	27.21191	2.78809
38	38.00000	27.00000	27.28304	-0.28304
39	39.00000	22.00000	27.35311	-5.35311
40	40.00000	29.50000	27.42213	2.07787
41	41.00000	28.00000	27.49010	0.50990
42	42.00000	27.50000	27.55701	-0.05701
43	43.00000	25.00000	27.62288	-2.62288
44	44.00000	30.50000	27.68768	2.81232
45	45.00000	29.50000	27.75144	1.74856
46	46.00000	29.00000	27.81414	1.18586
47	47.00000	34.00000	27.87578	6.12422
48	48.00000	29.50000	27.93638	1.56362
49	49.00000	23.50000	27.99592	-4.49592
50	50.00000	26.50000	28.05441	-1.55441
51	51.00000	30.50000	28.11184	2.38816
52	52.00000	30.50000	28.16822	2.33178

53	53.00000	34.00000	28.22355	5.77645
54	54.00000	26.50000	28.27782	-1.77782
55	55.00000	29.00000	28.33104	0.66896
56	56.00000	34.50000	28.38321	6.11679
57	57.00000	25.50000	28.43432	-2.93432
58	58.00000	35.00000	28.48438	6.51562
59	59.00000	33.00000	28.53339	4.46661
60	60.00000	27.50000	28.58135	-1.08135
61	61.00000	31.50000	28.62825	2.87175
62	62.00000	29.50000	28.67409	0.82591
63	63.00000	29.00000	28.71889	0.28111
64	64.00000	35.00000	28.76263	6.23737
65	65.00000	23.50000	28.80532	-5.30532
66	66.00000	23.50000	28.84695	-5.34695
67	67.00000	33.00000	28.88753	4.11247
68	68.00000	29.50000	28.92706	0.57294
69	69.00000	28.50000	28.96553	-0.46553
70	70.00000	27.00000	29.00295	-2.00295
71	71.00000	27.50000	29.03932	-1.53932
72	72.00000	25.50000	29.07464	-3.57464
73	73.00000	33.00000	29.10890	3.89110
74	74.00000	25.50000	29.14211	-3.64211
75	75.00000	25.20000	29.17426	-3.97426
76	76.00000	32.30000	29.20536	3.09464
77	77.00000	23.00000	29.23541	-6.23541
78	78.00000	25.25000	29.26440	-4.01440
79	79.00000	23.00000	29.29235	-6.29235
80	80.00000	20.60000	29.31923	-8.71923
81	81.00000	23.75000	29.34507	-5.59507
82	82.00000	38.50000	29.36985	9.13015
83	83.00000	29.60000	29.39358	0.20642
84	84.00000	41.50000	29.41625	12.08375
85	85.00000	26.00000	29.43787	-3.43787
86	86.00000	32.00000	29.45844	2.54156
87	87.00000	36.50000	29.47796	7.02204
88	88.00000	29.98000	29.49642	0.48358
89	89.00000	28.50000	29.51383	-1.01383
90	90.00000	26.50000	29.53018	-3.03018
91	91.00000	26.99000	29.54548	-2.55548
92	92.00000	29.54000	29.55973	-0.01973
93	93.00000	31.49000	29.57292	1.91708

CHART 2


1.0000
2.8776
4.7551
6.6327
8.5102
10.3878
12.2653
14.1429
16.0204
17.8980
19.7755
21.6531
23.5306
25.4082
27.2857
29.1633
31.0408
32.9184
34.7959
36.6735
38.5510
40.4286
42.3061
44.1837
46.0612
47.9388
49.8163
51.6939
53.5714
55.4490
57.3265
59.2041
61.0816
62.9592
64.8367
66.7143
68.5918
70.4694
72.3469
74.2245
76.1020
77.9796
79.8571
81.7347
83.6123
85.4898
87.3673
89.2449
91.1225
93.0000
17.0000
19.4500
21.9000
24.3500
26.8000
29.2500
31.7000
34.1500
36.6000