DANIEL F. MERRIAM, Editor

# STEPWISE REGRESSION AND NONPOLYNOMIAL MODELS IN TREND ANALYSIS


By

**A. T. MIESCH**

and

**J. J. CONNOR**

U. S. Geological Survey



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

**COMPUTER CONTRIBUTION 27**

State Geological Survey
The University of Kansas, Lawrence
1968

## Editor's Remarks

Just in case our "new" cover design was missed, note that the American Association of Petroleum Geologists now is assisting our effort in computer applications. With support of our activities by the AAPG, we enter a new era. Although the COMPUTER CONTRIBUTIONS have been enthusiastically received, we believe the series can be even better and more widely distributed.

An important area of computer applications in the earth sciences is trend analysis. One reason is that geologists have long been concerned with mapping trends in connection with locating mineral deposits. Petroleum geologists especially have been interested in "trendology" in defining structures or other features where petroleum might be trapped. The program presented here is another approach to trend analysis.

Even though many trend-surface programs are available, each one has merit. Choice of which one to use is left to the investigator and will depend on the problem involved. No one program is "better" than another, but may be "better" applied to a particular problem. As the authors state in this report "Selection of the best trend surface equation for separating and describing components of variation in the map data is ... difficult, and criteria that can be used to select the best equation consist mostly of geologic factors pertinent to the particular problem. This may require that a number of surfaces (models) be fitted and examined."

The program described here will be made available on magnetic tape for a limited time by the Geological Survey for $15.00. An extra $10.00 is charged if punched cards are required. Other programs described in the COMPUTER CONTRIBUTION series are listed on the inside back cover.

Users may find the following table of help in selecting a program appropriate to their problem.

METHOD

| Contouring | Time trend | Trend Analysis | Harmonic Analysis | Classification | Simulation | Other |
|---|---|---|---|---|---|---|
| CC15 | SDP12 | SDP3 | SDP24 | SDP4 | CC1 | B170-3 |
| CC23 | CC19 | SDP11 | CC5 | SDP9* | CC9 | SDP28 |
| | | SDP14 | CC16 | SDP13 | CC13 | CC2 |
| | | SDP26 | | SDP15 | CC24 | CC8 |
| | | B171 | | SDP23 | CC26 | CC11 |
| | | CC3 | | SDP27 | | CC14 |
| | | CC10 | | CC4 | | CC25 |
| | | CC27 | | CC6 | | |
| | | | | CC17 | | |
| | | | | CC20 | | |
| | | | | CC21 | | |

*out-of-print; SDP = Special Distribution Publications; CC = Computer Contributions; B = Bulletins.

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

# STEPWISE REGRESSION AND NONPOLYNOMIAL MODELS IN TREND ANALYSIS[1/]

by

A.T. Miesch and J.J. Connor

## INTRODUCTION

Techniques of fitting regression surfaces to geologic map data have been applied widely during the last decade, since the work of Krumbein (1956) and Miller (1956). The procedures, as used in geology, generally are referred to as trend analysis. Trend analysis of map data is performed primarily to separate and describe various components of variation that might be present, thereby facilitating geologic interpretation. Another purpose of trend analysis is to derive a regression equation that can be used for interpolation or prediction between control points on the map. Although there have been numerous applications of trend analysis to geologic problems since 1956, the general procedure has remained essentially the same as described by Krumbein (1959), except for some recent applications of Fourier series as an alternative to the polynomial models generally used (Preston and Harbaugh, 1965; James, 1966; Krumbein, 1966). This paper describes a somewhat modified approach to the general procedures of trend analysis in which a stepwise regression method is used. Stepwise regression has been used in trend analysis previously by Agterberg (1964), and a brief description of the technique used here was given previously by Miesch and Connor (1967).

Trend analysis directed toward interpolating or predicting values between map control points can, in many cases, be relatively straightforward. At least the criteria for judging the suitability of the regression surface can be clearly stated. The surface should account for a high proportion of the total sum of squares in the dependent variable, the deviations of the observed values of the dependent variable should not be autocorrelated, and the surface should "behave well" between the map control points. That is, there should be no more maxima or minima in the surface than are called for by the data being analyzed. The best polynomial equation that can be used for prediction is the lowest order polynomial leading to nonsignificant autocorrelation in the trend deviations.

Selection of the best trend surface equation for separating and describing components of variation in the map data is considerably more difficult, and criteria that can be used to select the best equation consist mostly of geologic factors pertinent to the particular problem. This may require that a number of surfaces (models) be fitted and examined. The general mathematical forms of the equations must be chosen on substantive grounds, but following this they can be refined using either geologic criteria (if this is possible) or various statistical tests. Most trend analysis work so far has involved refinement of polynomial equations up to about fifth degree by dropping polynomial terms in groups (for example, dropping fifth degree terms, then fourth degree, and so forth). By using stepwise regression methods, the terms to be used in the equation can be selected individually according to their potential effectiveness in reducing the total sum of squares in the dependent variable - and dropped from the equation individually if they are not effective or if they are redundant. The terms under examination need not be restricted to polynomials, however; other terms may be better in particular problems. Terms selected frequently depend on the X-Y coordinate system used and by changing the coordinate system, by entering different kinds of terms into the stepwise procedure, and by changing the significance level at which terms are to be entered into or deleted from the regression equation, a number of different trend surfaces of about equally good fit to the observed data can be derived. The relative geologic significance of these surfaces remains a matter of subjective judgment.

Stepwise regression generally leads to trend-surface equations which are considerably more efficient than conventional polynomial equations in that they may account for large proportions of the total sum of squares with many fewer terms. An attractive consequence of this is that the matrix operations performed to derive the coefficients are less affected by roundoff errors. Moreover, the derived coefficients are more stable - or less sensitive to small errors in the data.

## CONVENTIONAL TECHNIQUES OF TREND ANALYSIS

Trend analysis is an empirical method for examining and interpreting the variation of numerical geologic map data. The total variation is viewed as having three types of components which are of varying relative magnitude. One of the components is a trend which extends over at least the major part of the study area, and may be thought of as having resulted from one or more geologic processes which acted across the major part of the area. A second component consists of local deviations from the trend that have resulted from geologic processes

1/Publication authorized by the Director, U.S. Geological Survey.

which occurred within some limited part of the area of study, but over areas broader than the average interval between data points on the map. A third component consists of what has been referred to as "noise," and includes all variation having resulted from local geologic factors and from data errors of various kinds. The local geologic factors give rise to sampling errors which, in many instances, may form the major portion of the total noise in the data.

The trend component is estimated by fitting a regression equation to the observed map data. Except for early work based on orthogonal polynomials (Krumbein, 1956) and some recent work employing Fourier series which has already been cited, nearly all trend analysis applications known to us employ polynomial equations containing terms such as $X$, $X^2$, $X^3$, $\ldots X^n$, or where map data are used the terms based on two independent variables are $X$, $Y$, $X^2$, $XY$, $Y^2$, $X^3$, ... etc. Using these terms in the general linear model (Krumbein and Graybill, 1965, p. 301), the trend function becomes

$$T = b_0 + b_1X + b_2Y + b_3X^2 + b_4XY + b_5Y^2 + b_6X^3 \ldots . \qquad (1)$$

Polynomial equations as in (1) are well suited to geologic trend-surface studies because they are appropriate for defining a wide variety of smooth surfaces that seem intuitively to be reasonable representations of trend components. The general concept of a trend component is a smooth continuous surface, which may have gentle flexure, and which extends across the entire area of investigation. However, a great many surfaces of this type might be poorly approximated by low-degree polynomial equations and other equally simple terms can be useful for this purpose. This is demonstrated in Figure 1 which shows the best fit of a quadratic equation to a hypothetical trend defined by the equation

$$T = 100X^{-1} + 0.2e^X . \qquad (2)$$

The quadratic equation, fitted by the method of least squares, accounts for 97.7 percent of the total sum of squares in the dependent variable, but the deviation of the quadratic from the correct trend could cause erroneous geologic interpretation in an actual trend analysis problem.

A number of methods and tests are available for selecting the particular polynomial terms to use in defining the trend component. In most applications the terms are added to or deleted from the model in groups, such as the second-degree group of terms ($X^2$, $XY$, $Y^2$) or the third-degree group ($X^3$, $X^2Y$, $XY^2$, $Y^3$). The nth degree terms are all those having exponents which sum to n. The selection of terms for polynomial models has been done in many different ways, including inspection of the proportion of the total sum of squares in the dependent variable that can be accounted for by the terms (cf. Whitten, 1959, p. 839), testing the statistical significance of the proportion of the total sum of squares accounted for (cf. Allen and Krumbein, 1962, p. 522), observation of the effect of the terms on the computed variance of the trend residuals (Mandelbaum, 1963, p. 509), estimating the autocorrelation in trend residuals (cf. Connor and Miesch, 1964, p. 121), and consideration of geologic factors alone (cf. Read and Merriam, 1966, p. 97).

Advantages of polynomial terms are their convenience in computation and their independence of the units of the particular coordinate scheme used to define X and Y. Either the scale or the origin of the coordinate scheme may be changed without causing changes in either the form of the computed surface or the degree to which the surface fits the observed data (i.e., the proportion of the total sum of squares accounted for by the surface). Changes in scale and origin, however, do affect the polynomial coefficients, and, therefore, the selection of terms by stepwise regression. This will be demonstrated in a later section.

Another advantage of polynomial terms, particularly the lower degree terms, is that they generally define surfaces which are "well behaved" between map control points. If higher degree polynomial terms, or some other types of terms in X and Y, are used it is possible, in some instances, to find on evaluation of the trend equation between map control points that the surface is far beyond the general realm of the data. Maxima and minima may be present which are uncalled for by either the data or the underlying geology.

After the trend component has been estimated the differences between the fitted surface and the observed map value are determined for each map control point. These differences are termed the trend residuals and contain estimates of the second and third components of variation - the local deviations and noise - referred to previously. Where the residuals are autocorrelated, indicating that adjacent values on the map tend to be similar (e.g., clusters of positive or negative values), the second component is presumed to be dominant over the third. Where autocorrelation is low, indicating that adjacent trend residuals tend to be unrelated, the third component is presumed to be dominant over the second. Autocorrelation in trend residuals has been discussed in several papers by Agterberg (1964, 1966) and by Connor and Miesch (1964, p. 123).

## USE OF NONPOLYNOMIAL TERMS IN THE GENERAL LINEAR MODEL

The procedure of trend analysis is empirical primarily because too little is known about the quantitative aspects of geologic processes that may produce trends. Only rarely, if ever, does the geologist

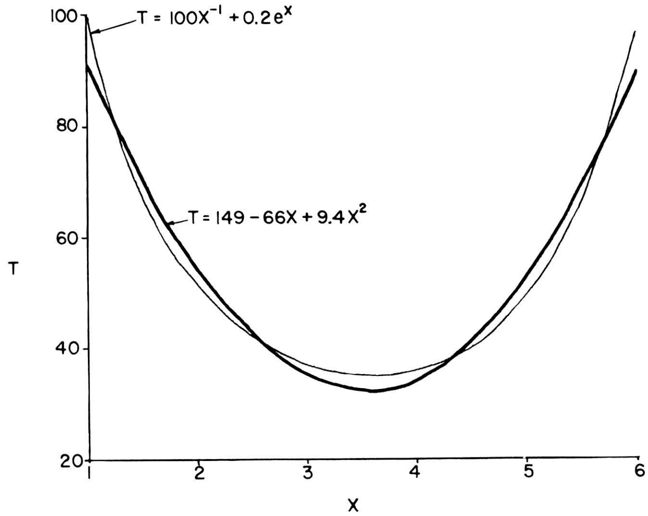



Figure 1.- Least-squares fit of a quadratic equation to the curve $T = 100X^{-1} + 0.2e^{X}$.

have reason to suspect that the variable he is studying will vary across the region of investigation as, for instance, the square of the distance from some point, or the log of the distance. If the proper form of the function were known, it would be a relatively simple matter to determine the coefficients for the function by least-squares methods, and no search for an otherwise adequate or approximate function would be necessary. Because the correct functional form is almost never known in real geologic problems, polynomial equations have been widely employed as approximations and while it is probably true that such equations are adequate for many of the problems to which the techniques of trend analysis have been applied, there is no reason to expect that they are fundamentally correct or that their use can be justified in many cases on any sound theoretical basis. Neither do we have reason to believe that they are the best or most efficient that can be used. Polynomial terms have been adequate in many problems because various combinations could be used to approximate a wide variety of other smooth continuous functions that may be more fundamentally correct.

The empirical nature of trend analysis is further evidenced by the fact that even if the correct functional form of the trend were known, it could not be properly fitted to the data by least-squares methods if local geologic effects, distributed over the area in a nonrandom fashion, were present. The effects cause the estimates of the trend coefficient to be biased. The computed trends, in such cases, are partially dependent on the local components of the variation, and the degree and nature of these effects can be neither estimated nor reduced.

There are an almost unlimited number of simple terms in X and Y, aside from polynomial terms, that could be used in the general linear equation for trend analysis procedures. Those with which we have experimented are listed in Table 1, along with polynomial terms through fifth degree. The approach we have used is to enter all terms of Table 1 into a stepwise regression procedure, allowing any of them to be entered into the trend-surface equation as independent variables according to the stepwise test criteria. In this manner, the most efficient terms are entered into the equation and all terms which either account for a nonsignificant portion of the variance in the dependent variable (at a specified confidence level) or are linearly dependent on terms already in the equation are excluded. The trend-surface equation obtained through the stepwise regression procedure, therefore, is generally more efficient than one obtained through ordinary polynomial regression in regard to the number of terms and the portion of the total variance accounted for. Most polynomial regression equations, arrived at by adding and testing the terms in groups, contain individual terms which are superfluous. A second advantage that may be gained by including nonpolynomial terms in the trend model is that the trend equation may be derived from a better conditioned coefficient matrix in the normal equations and, therefore, will be less sensitive to sampling and analytical errors in the data or to roundoff errors in computation.

## STEPWISE REGRESSION

The stepwise regression procedure used in this investigation was programmed in ALGOL for the B5500 computer (U.S.G.S. program no. W0005), by D.S. Handwerker of the U.S. Geological Survey, and follows, in major part, the general technique described by Efroymson (1960). The algorithm includes procedures similar to those described by Fisher (1950, p. 156-166), Anderson and Bancroft (1952, p. 176-182), and Ostle (1954, p. 202-227). Part of the ALGOL program was rewritten in FORTRAN IV for the IBM System 360, by G.I. Selner, and is included in a general system of statistical programs used by the U.S. Geological Survey. A modification of the latter program that can be used independently of the general system has been prepared by Robert Terrazas and is included here in appendix B.

In the application of stepwise regression methods in trend analysis, we attempt to derive a regression equation containing terms that are each significant at some prescribed level of confidence. Significant terms are those which, when included in the equation, account for sufficiently large portions of the total variance in the dependent variable so that the relationship is unlikely to have resulted from chance alone.

The only major difference between the stepwise procedure in our program and the algorithm given by Efroymson (1960) is that we specify a probability level, Q, on input rather than the critical values of F. Efroymson specified values of $F_1$ and $F_2$

Table 1.- Thirty-eight terms in X and Y used in stepwise regression.

| | |
|---|---|
| Polynomial terms: | |
| Linear . . . . . . . . . . . . . . . | $X, Y$ |
| Quadratic . . . . . . . . . . . . . | $X^2, XY, Y^2$ |
| Cubic . . . . . . . . . . . . . . . | $X^3, X^2Y, XY^2, Y^3$ |
| Quartic . . . . . . . . . . . . . . | $X^4, X^3Y, X^2Y^2, XY^3, Y^4$ |
| Quintic . . . . . . . . . . . . . . | $X^5, X^4Y, X^3Y^2, X^2Y^3, XY^4, Y^5$ |
| Nonpolynomial terms: | |
| Square root . . . . . . . . . . . | $\sqrt{X}, \sqrt{XY}, \sqrt{Y}$ |
| Exponential . . . . . . . . . . . | $e^X, e^Y, e^{2X}, e^{X+Y}, e^{2Y}$ |
| Logarithmic . . . . . . . . . . . | $\log X, \log Y, (\log X)^2, \log X \cdot \log Y, (\log Y)^2$ |
| Reciprocal . . . . . . . . . . . | $X^{-1}, Y^{-1}, X^{-2}, (XY)^{-1}, Y^{-2}$ |

on input and these are used in adding independent terms to and deleting them from the regression equation. In the program we used, the probability of computed F values' occurring by chance is estimated using approximation techniques and compared with the specified value of Q (generally 0.05 or 0.01). Initially, none of the terms is considered to be in the regression equation. By means of simple linear correlation coefficients, each is tested then for the proportion of the total sum of squares of the dependent variable that it explains. The most significant term in this respect is then entered into the regression equation. Because the significance of an independent term in the equation will change with the addition of new terms, each term already in the equation is tested for significance immediately after the addition of each new term. Such terms in the equation shown to be nonsignificant are deleted from the equation.

The basic method consists of solving a set of simultaneous equations by Gaussian elimination and using results obtained at each stage of the elimination as stepwise test criteria (Efroymson, 1960, p. 192). Equations are formed from simple correlation coefficients derived from the normal equations as shown by Ostle (1954, p. 202-205); they are discussed more fully in the section on matrix condition.

The essential features of the procedure are given here for the convenience of those who may wish to prepare their own computer programs or as an aid in implementing the program given in appendix B.

1. Read N, number of rows in data matrix, and Q, probability level at which terms are to be added to or deleted from regression (usually 0.05 or 0.01).
2. Read data matrix, $X_{r,c}$
   $1 \leqslant r \leqslant N$
   $0 \leqslant c \leqslant 2$
   $X_{r,0}$ = observed value of dependent variable
   $X_{r,1}$ = X map coordinate
   $X_{r,2}$ = Y map coordinate
3. Generate n - 2 functions of the map coordinates, $X = (X_{r,1})$ and $Y = (X_{r,2})$, similar to the terms listed in Table 1. Augmented data matrix, $X_{r,c}$, now is N by (n + 1) with $1 < r \leqslant N$ and $0 \leqslant c \leqslant n$.
4. Generate means, $\overline{X}_c$, and standard deviations, $s_c$, for all columns in augmented data matrix. $0 \leqslant c \leqslant n$.
5. Generate simple correlation coefficient matrix, $r_{i,j}$. $0 \leqslant i,j \leqslant n$.
6. Set NDF = N - 1, where NDF indicates degrees of freedom.
7. Set VAR = 1, where VAR is proportion of variance in dependent variable not accounted for by the regression.
8. Set array $C_i = -1$ $(1 \leqslant i \leqslant n)$. If $C_i = -1$, then the ith term is not in the regression equation. If $C_i = +1$, then the ith term is in the regression equation.

9. Generate array $V_i = r^2_{i,0}/r_{i,i}$ $(1 \leqslant i \leqslant n)$.
   a. Find minimum $V_i$ among all values where corresponding $C_i = +1$ and $r_{i,i} > 0.00001$. If none of the variables satisfy the criteria with respect to $C_i$ and $r_{i,i}$, go to 9b. (NOTE: The limitation on $r_{i,i}$ reduces the possibility of entering linearly dependent variables into the regression equation (Efroymson, 1960, p. 195 ). Set VMIN = minimum $V_i$. F = (VMIN · NDF)/ VAR. Determine QF, the probability of F for 1 and NDF degrees of freedom (see text). If <u>QF ⩾ Q</u>, set k = i of minimum $V_i$, increase <u>VAR by</u> VMIN, increase NDF by 1, and go to 10. <u>If QF < Q</u>, go to 9b.
   b. Find maximum $V_i$ among all values where corresponding C = −1 and $r_{i,i} > 0.00001$. Set VMAX = maximum $V_i$. F = (VMAX · NDF)/ (VAR − VMAX). Determine QF, the probability of F for 1 and NDF degrees of freedom (see text). If <u>QF ⩽ Q</u>, set k = i of maximum $V_i$, decrease <u>VAR by</u> VMAX, decrease NDF by 1, and go to 10. If <u>QF > Q</u>, go to 11.
10. Apply the following transformations to those parts of the $r_{i,j}$ matrix in Figure 2 having the corresponding equation numbers.

Equation (1) $$r_{i,j} \leftarrow r_{i,j} - \frac{C_k \cdot C_j \cdot r_{i,k} \cdot r_{j,k}}{r_{k,k}}$$

(2) $$r_{i,k} \leftarrow -\frac{r_{i,k}}{r_{k,k}}$$

(3) $$r_{i,j} \leftarrow r_{i,j} - \frac{r_{k,j} \cdot r_{i,k}}{r_{k,k}}$$

(4) $$r_{k,k} \leftarrow \frac{1}{r_{k,k}}$$

(5) $$r_{k,j} \leftarrow \frac{r_{k,j}}{r_{k,k}}$$

(6) $$r_{i,j} \leftarrow r_{i,j} - \frac{C_k \cdot C_i \cdot r_{k,j} \cdot r_{k,i}}{r_{k,k}}$$

(7) $$r_{i,0} \leftarrow r_{i,0} - \frac{r_{k,0} \cdot r_{i,k}}{r_{k,k}}$$

(8) $$r_{k,0} \leftarrow \frac{r_{k,0}}{r_{k,k}}$$

(9) $$r_{i,0} \leftarrow r_{i,0} - \frac{C_k \cdot C_i \cdot r_{k,0} \cdot r_{k,i}}{r_{k,k}}$$

Change sign of $C_k$ $(C_k \leftarrow -C_k)$. Go to 9.

11. The column vector, $r_{i,0}$, now contains the standardized partial regression coefficients for all variables $(1 \leqslant i \leqslant n)$ which reduce the variance in the dependent variable by a significant amount. $C_i$ for these variables is equal to +1. The regression coefficients are computed from:

    $$b_i = \frac{r_{i,0} \cdot s_0}{s_i}$$ , for all values of i where $C_i = +1$

    and the regression equation constant from

    $$b_0 = \bar{X}_0 - \sum_i (b_i \cdot \bar{X}_i)$$, for all values of i where $C_i = +1$.
12. The percent of the total sum of squares in the dependent variable accounted for by the regression equation is

    $$PSS = 100\sum_i (r_{i,0} \cdot r_{0,i}),$$ where $C_i = +1$.
13. Compute regression residuals,

    $$d_r = X_{r,0} - T_r,$$
    where $T_r$ is computed from the regression equation for the rth row of the data matrix.
14. Print
    $r_{i,j}$ $(0 \leqslant i,j \leqslant n)$
    $C_i$ $(1 \leqslant i \leqslant n)$
    $b_i$ $(1 \leqslant i \leqslant n)$, where $C_i = +1$
    $b_0$
    PSS
    $T_r$ $(1 \leqslant r \leqslant N)$
    $d_r$ $(1 \leqslant r \leqslant N)$

A sample calculation is given in Appendix A.

QF in step 9 of the procedure is approximated by techniques given by Abramowitz and Stegun (1964, p. 932, 946-947), as follows. If NDF, the number of degrees of freedom for the variance estimate in the denominator of F, is greater than 100, the variable Z is derived from

$$Z = \frac{\left(F^{1/3}\left(1 - \frac{2}{9NDF}\right)\right) - \frac{7}{9}}{\left(\frac{2}{9} + \frac{2F^{2/3}}{9NDF}\right)^{1/2}} . \qquad (3)$$

Then, because the F distribution is approximately normal for large numbers of degrees of freedom, the area under the F distribution curve above Z is estimated from

$$QF = \frac{1}{\sqrt{2\pi}} \int_Z^{\infty} e^{-t^2/2}\, dt, \quad (4)$$

which is approximated by

$$QF \approx \frac{e^{-Z^2/2}}{\sqrt{2\pi}} \left(a_1 t + a_2 t^2 + a_3 t^3 + a_4 t^4 + a_5 t^5\right) \quad (5)$$

with

$$t = 1/(1 + 0.2316419|Z|)$$
$$a_1 = +0.31938\ 1530$$
$$a_2 = -0.35656\ 3782$$
$$a_3 = +1.78147\ 7937$$
$$a_4 = -1.82125\ 5978$$
$$a_5 = +1.33027\ 4429$$

If Z as computed from (3) is negative, QF as computed from (5) is replaced by 1.0 − QF. If NDF $\leqslant$ 100 and is an even number

$$Z = \frac{NDF}{NDF + F} \quad (6)$$

$$QF = 1 - \sqrt{1 - Z}\left(1 + \frac{Z}{2} + \frac{1\cdot 3}{2\cdot 4} Z^2 + \ldots + \frac{1\cdot 3 \ldots (NDF - 3)}{2\cdot 4 \ldots (NDF - 2)} Z^{(NDF - 2)/2}\right) \quad (7)$$

If NDF $\leqslant$ 100 and is an odd number,

$$QF = 1 - \alpha(NDF) \quad (8)$$

where, if NDF = 1

$$\alpha(NDF) = \frac{2\theta}{\pi} \quad (9)$$

or where, if NDF > 1

$$\alpha(NDF) = \frac{2}{\pi}\left\{\theta + \sin\theta\left(\cos\theta + \frac{2}{3}\cos^3\theta + \ldots + \frac{2\cdot 4 \ldots (NDF - 3)}{1\cdot 3 \ldots (NDF - 2)} \cos^{(NDF - 2)}\theta\right)\right\} \quad (10)$$

and

$$\theta = \arctan (F/NDF)^{1/2} \quad (11)$$

At the conclusion of the stepwise regression procedure a subset of the upper triangle of the $r_{i,j}$ matrix contains the elements of the inverse of a subset of the original $r_{i,j}$ matrix of correlation coefficients. The subset consists of all correlation coefficients among independent variables in the regression equation (i.e., variables for which $C_i = +1$ at the end of the stepwise regression procedure).

The $r_{i,0}$ column of the $r_{i,j}$ matrix, at the end of the procedure, contains the standardized regression coefficients for all variables with $C_i = +1$; in some problems a variable may account for a statistically significant portion of the variance in the dependent variable, and therefore be included in the regression equation, but the corresponding standardized regression coefficient may indicate that its contribution to the regression is extremely small. In such cases, the user may choose to drop this variable and recompute a regression equation containing the other selected independent variables by conventional regression methods.

## MATRIX CONDITION

The algorithm given by Efroymson (1960) contains shortcut procedures for estimation of the regression coefficients, but the end result of the method is equivalent to solving the following set of equations for the β's (standardized partial regression coefficients) after the variables to be included in the regression equation have been selected.

$$\begin{aligned} \beta_1 + r_{12}\beta_2 + \ldots + r_{1m}\beta_m &= r_{10} \\ r_{21}\beta_1 + \beta_2 + \ldots + r_{2m}\beta_m &= r_{20} \\ &\vdots \\ r_{m1}\beta_1 + r_{m2}\beta_2 + \ldots + \beta_m &= r_{m0} \end{aligned} \quad (12)$$

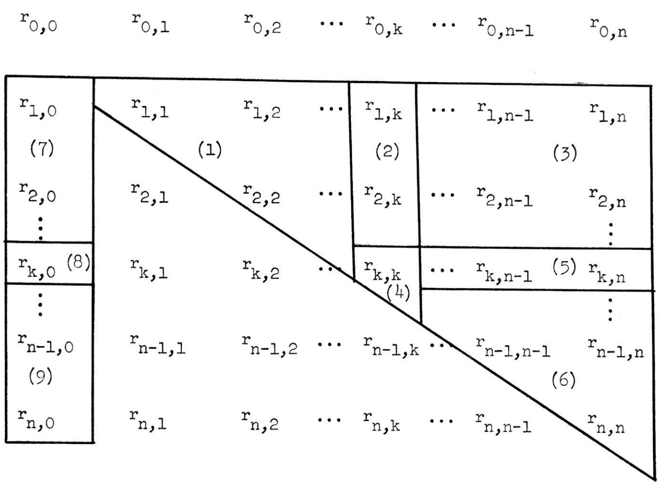
The subscript numbers refer to the m selected independent variables and the zero subscript refers to the dependent variable. The equations in (12) are represented in matrix notation as

$$RB = C \quad (13)$$

where R is the matrix of simple correlation coefficients among the independent variables and B and C are vectors representing the unknown coefficients and the correlations of the dependent variable with the independent variables, respectively. Solution of the equations is attempted by the matrix operation

$$B = R^{-1}C \quad (14)$$

where $R^{-1}$ is the inverse of the R matrix. However, if the R matrix is singular (having no inverse) the equations cannot be solved. Moreover, if the R matrix is ill conditioned the solutions for the β's (represented by B) may be very sensitive to errors in the r's. Also, if R is highly ill conditioned, although not singular, derivation of the inverse,




Equation (1) $1 \leq i \leq k-1$ and $i \leq j \leq k-1$

(2) $1 \leq i \leq k-1, \quad j = k$

(3) $1 \leq i \leq k-1$ and $k+1 \leq j \leq n$

(4) $i = k$ and $j = k$

(5) $i = k, \quad k+1 \leq j \leq n$

(6) $k+1 \leq i \leq n$ and $i \leq j \leq n$

(7) $1 \leq i \leq k-1, \quad j = 0$

(8) $i = k, \quad j = 0$

(9) $k+1 \leq i \leq n, \quad j = 0$

Figure 2.- Partitions of the $r_{i,j}$ matrix to which equations 1 to 9 of step 10 in the stepwise regression procedure are applied.

$R^{-1}$, may require carrying a very large number of significant figures in computation. Computer programming in double precision arithmetic alleviates this roundoff problem to some extent, but it may be encountered again with a still more ill-conditioned matrix. Krumbein (1959, p. 828) discussed the limitations associated with ill-conditioned matrices in trend analysis and noted that matrix condition is poorer where a limited number of control points are irregularly scattered or clustered on the map. Mandelbaum (1963, p. 506-508) suggested a method for improving the condition of a matrix formed from sums of squares and cross products in the map coordinates.

A number of measures can be used to evaluate matrix condition, but for convenience we have adopted one recommended by Booth (1957, p. 85) and Macon (1963, p. 66). This is the determinant of the matrix after it has been normalized by dividing each element in a row by the row sum of squares. We refer to the determinant of the normalized matrix as the "condition value." Condition values may range from zero for a singular matrix to $\pm$ 1.0 for one that is ideally conditioned.

An alternative and equally good measure of matrix condition is the ratio of the largest to the smallest eigenvalue of the R matrix (cf. Fox, 1965, p. 142). The only advantage of the condition value is in ease of computation, particularly the fact that the condition value is less sensitive than the eigenvalue ratio to the number of significant figures carried in R.

Two factors determine the condition value of the R matrix in any particular trend analysis problem employing a given map coordinate system. These are (1) the distribution of X-Y control points on the map, and (2) the particular mathematical terms present in the trend equation. Some elementary examples are given in Figure 3. At the extreme, the distribution of map control points is entirely inadequate (Fig. 3A) for fitting even a first-degree polynomial surface; condition values for the R matrices of polynomial surfaces of degree one through three are zero. The R matrices for the linear through cubic polynomial trends are (to two significant figures) shown in Table 2. Singularity of this matrix is obvious from the fact that each subset (representing linear, quadratic, or cubic terms) contains duplicate rows and columns. Ill conditioning occurs in other matrices where rows or columns tend to be linearly correlated. The degree of linear correlation can be high where more than one high-degree polynomial term is used.

In Figure 3B, the single outlying point provides some control for fitting a linear surface, but the control is poor. The low condition values for the quadratic and cubic surfaces indicate that the R matrices will probably be quite ill conditioned. The distribution of control points is somewhat better in Figure 3C where three points occur as outliers. The condition values here indicate that linear and probably quadratic surfaces can be fitted without roundoff problems. The clustered distribution of control points

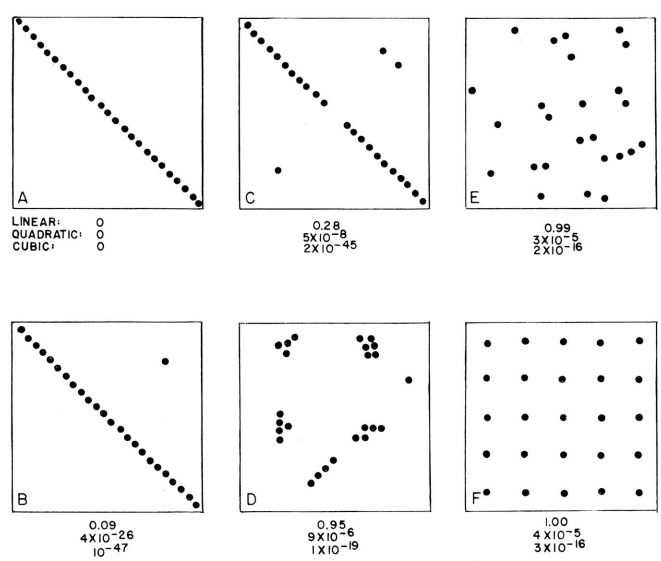



Figure 3.- Condition values of R matrices that would be used in fitting polynomial surfaces to various distributions of map control points.

Table 2.- R matrices for linear through cubic polynomial trends.

| | Linear | | Quadratic | | | Cubic | | | |
|---|---|---|---|---|---|---|---|---|---|
| | X | Y | $X^2$ | XY | $Y^2$ | $X^3$ | $X^2Y$ | $XY^2$ | $Y^3$ |
| X | 1.0 | 1.0 | 0.97 | 0.97 | 0.97 | 0.92 | 0.92 | 0.92 | 0.92 |
| Y | 1.0 | 1.0 | 0.97 | 0.97 | 0.97 | 0.92 | 0.92 | 0.92 | 0.92 |
| $X^2$ | 0.97 | 0.97 | 1.00 | 1.00 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 |
| XY | 0.97 | 0.97 | 1.00 | 1.00 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 |
| $Y^2$ | 0.97 | 0.97 | 1.00 | 1.00 | 1.00 | 0.99 | 0.99 | 0.99 | 0.99 |
| $X^3$ | 0.92 | 0.92 | 0.99 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 |
| $X^2Y$ | 0.92 | 0.92 | 0.99 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 |
| $XY^2$ | 0.92 | 0.92 | 0.99 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 |
| $Y^3$ | 0.92 | 0.92 | 0.99 | 0.99 | 0.99 | 1.00 | 1.00 | 1.00 | 1.00 |

shown in Figure 3D provides a good basis for fitting linear and quadratic surfaces, but computation of coefficients for a cubic surface may be difficult. The control points in Figure 3E are randomly distributed over the map area, and lead to slightly better conditioned R matrices. With an increase in the number of random control points, the condition values of the R matrices would converge toward those for points on a grid, as in Figure 3F. The similarity of the condition values for the distributions of 3D, 3E, and 3F (particularly the linear and quadratic values) suggests that, at least for low-order polynomial surfaces, the map control point distribution is critical mostly to the extent that all major parts of the map should be represented. In particular, Figure 3D strongly indicates that the "outcrop type" of control point distributions will probably not adversely affect the condition of the R matrix for this type of trend analysis.

A high condition value for an R matrix is by no means the only test of the similarity of map data for trend analysis. Clustered map control points, as in Figure 3D, introduce redundancy in data and may affect determination of the number of degrees of freedom available for statistical tests. Also, even though the condition value of the R matrix used to estimate a linear trend for Figure 1B is high, it is obvious that such a surface may be controlled by one data point only, and therefore may be highly unstable.

## Significance of the Condition Value

Condition values of R matrices, or of equivalent matrices (e.g., matrices of sums of squares and cross products), can serve as indicators of possible sampling or mathematical problems in trend-analysis investigations. Among the causes contributing to low condition values are a poor distribution of control points on the map or a near linear dependence among any two or more terms in X and Y used in the regression surface equation.

The principal consequence of ill-conditioned matrices is that roundoff in computation is more likely to result in erroneous regression coefficients. The minimum condition value that can be tolerated depends on the computer used, the technique used in computing the coefficients, and the method of programming (Fox, 1965, p. 96-97, 136ff). We have used several methods for deriving the regression coefficient estimates, using a B5500 computer which carries approximately 11 decimal figures in single precision computations. Matrix computations have been checked by examining the symmetry of R matrix inverses and by comparing inverses derived by different methods of computation. It has been noted that at least one (commonly five or more) significant figure is retained in matrix computations by several different procedures where the condition value of the original R matrix is greater than about $10^{-10}$, in absolute value. Where the condition value is lower than $10^{-10}$, in absolute quantity, serious errors in matrix computations due to roundoff may or may not occur, but the probability of such errors appears to increase with decreasing condition value. The exact point at which the last significant figure is lost in matrix computations depends again on the particular matrix, the computer used, and the procedures and programming techniques used in the computations.

Another effect of ill conditioning in the R matrix, as indicated by a low condition value, is that the estimated coefficients, the β's, are unstable. That is, when the original values of X and Y are perturbed slightly by adding to them small random normal deviates (with zero mean), the estimates of the coefficients change rapidly. However, the changes in the coefficients are apparently compensating because changes in the form of the fitted surfaces are related only to the magnitude of the perturbations in X and Y, not to the condition value. This consequence of low condition values, therefore, is of little importance unless some use were to be made of individual coefficients.

It should be noted that the condition value is by no means a complete measure or indicator of error in the estimated regression coefficients. If the condition value is high, roundoff errors in computation will probably not cause error in the coefficients, but the estimates will not necessarily be close to the true coefficients. The standard error of a regression coefficient estimate is given by (See Ostle, 1954, p. 213-216.):

$$sb_i = \left( \frac{s_0^2 \cdot (1 - \frac{PSS}{100}) \cdot r^{i,i}}{s_i^2 \cdot (N - m - 1)} \right)^{1/2} \qquad (15)$$

where, after the stepwise procedure, $r^{i,i}$ is the ith diagonal element in the inverse of the matrix of correlation coefficients among the n independent variables in the regression. If the condition value of R is small, $r^{i,i}$ may be large causing $sb_i$ to be large, but $sb_i$ depends also on the other variables in (15).

EXAMPLE I - Lost Springs Area, East-Central Kansas

The location of the Lost Springs area is shown in Figure 4. The structure contour map is drawn on the top of the "Mississippi chat," an interval of weathered Mississippian chert generally considered to be the basal deposit of the Pennsylvanian system in this part of Kansas (Shenkel, 1955, p. 176). An oil pool in the "chat" and associated beds (as shown

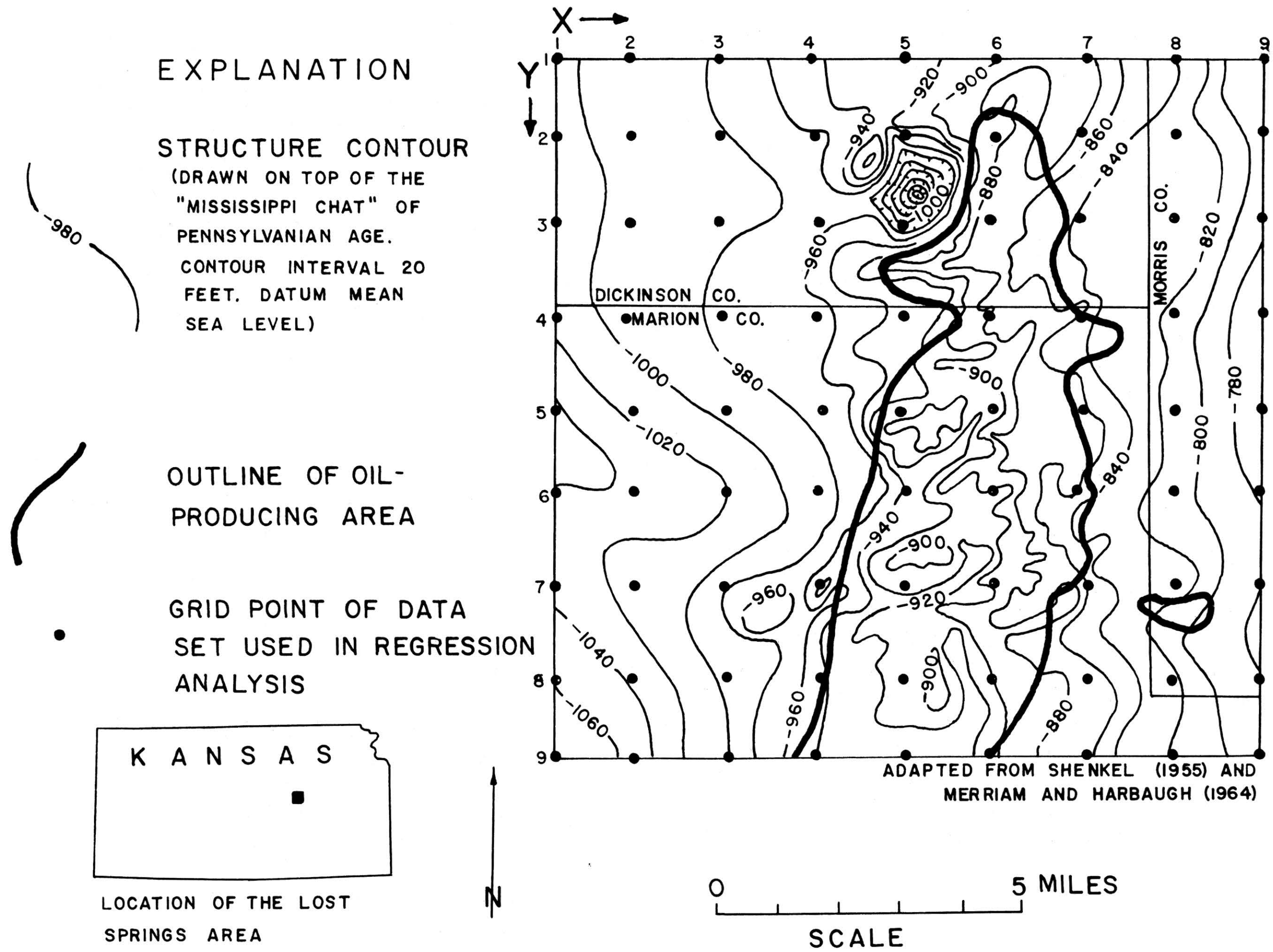



Figure 4.- Structure contour map of the top of the "Mississippi chat" in the Lost Springs area, Kansas.

by Merriam and Harbaugh, 1964) is outlined by the solid black line. The greater detail shown by the contours within the area underlain by the oil pool is due to the more dense well control in this area. The data set used for analysis in this example consists of 81 values of the dependent variable (structural elevation) read at the X and Y grid locations. The origin of the grid is arbitrarily placed off the northwest corner of the map; values of the X and Y coordinates each range from 1 to 9, increasing from west to east and from north to south, respectively.

Conventional polynomial regression surfaces of degree 1 to 5 were fitted to the observed data and accounted for 94.7, 96.7, 97.2, 97.3, and 97.6 percent of the total sum of squares in the dependent variable, respectively. The percentage sums of squares accounted for and analysis of variance tests both indicate that polynomial terms beyond degree 2 do little to improve the regression, even though the third-degree polynomial terms, as a group, are significant at the 0.05 level of probability. The first- and second-degree polynomial surfaces and the residuals from each are shown by the contour maps in Figure 5 (A and B).

Condition values of the R matrices used to compute the polynomial regression surfaces, degree 1 through 5, are 1.0, $3 \cdot 10^{-5}$, $2 \cdot 10^{-16}$, $5 \cdot 10^{-37}$, and $< 10^{-46}$, respectively. On entering the stepwise regression procedure with the linear through quartic terms, it was found that the only terms significant at the 0.05 probability level are X and $X^2$ (Fig. 5C). These two terms account for 95.8 percent of the total sum of squares in the dependent variable, in contrast to 97.3 percent accounted for by the 14 terms in the complete quartic equation. The condition value of the R matrix used to compute the surface based on X and $X^2$ only was 0.025, in contrast to $5 \cdot 10^{-37}$ for the complete quartic surface, and $3 \cdot 10^{-5}$ for the complete quadratic surface.

Although translation of the grid coordinate system will not affect the form of a polynomial surface fitted to the data, nor the proportion of the total sum of squares accounted for by the surface, large changes can occur in the coefficients of the regression equation (the b's of equation 1 as well as the β's of equation 12). Translation of the coordinate system, therefore, causes changes in the relative significance of various terms in the regression equation. On adding a constant of 10 to each X and Y coordinate value and entering the stepwise procedure with newly derived polynomial terms (again degree 1 through 5) the only term significant at the 0.05 probability level is $X_t^3$ where in this case $X_t = X + 10$. A surface based on $X_t^3$ alone accounts for 95.7 percent of the total sum of squares in the dependent variable (Fig. 5D). The R matrix based on this single term, of course, has a condition value of 1.0. The stepwise procedure was repeated after adding a constant of 30, rather than 10, to all X and Y coordinate values. The only significant term in this case was $X_t^4$ where $X_t = X + 30$, and a surface based on this single term accounted for 95.8 percent of the total sum of squares.

Trend and residuals maps derived by means of stepwise regression, before and after translation of the coordinate system, are roughly the same as those derived from fitting a second-degree polynomial surface (Fig. 5B to 5D). The differences are certainly not sufficient to cause widely differing geologic interpretations. In each case the oil pool, as outlined by Merriam and Harbaugh (1964, p. 23) occurs almost entirely within an area of positive trend residuals.

A stepwise regression analysis was made of the Lost Springs structural data (Fig. 4) using all polynomial and other terms listed in Table 1. Three terms were selected by the procedure as significant at the 0.05 probability level. One of the terms, $X^2$, is a polynomial; the other two, $\sqrt{Y}$ and log X·log Y, are not. The surface (Fig. 5E) accounts for about the same percentage of the total sum of squares as a second-degree polynomial (Fig. 5B), but it is derived from an R matrix with a much higher condition value (0.12 vs. $3 \cdot 10^{-5}$). The principal features of the trend and residual maps are about the same.

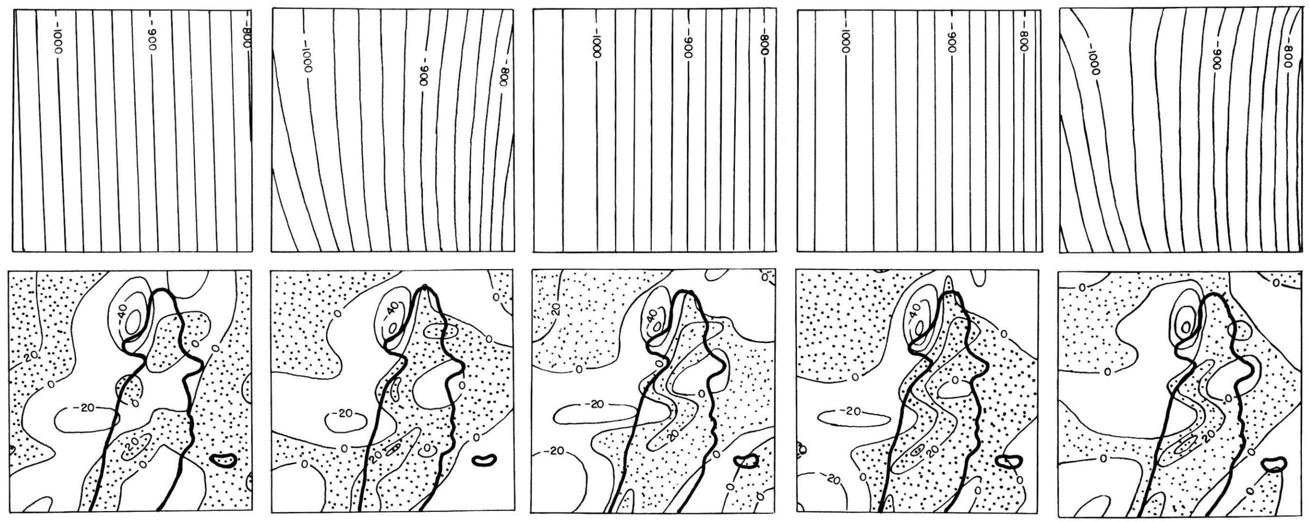
The stepwise regression analysis was repeated using all terms of Table 1, but by selecting those terms significant at the 0.25, rather than the 0.05, level of probability. The trend-surface equation is:

$$t = -974 + 2.43X^2 + 2.62XY - 0.023Y^3 - 33\sqrt{Y} - 0.0018X^3Y^2 \qquad (16)$$

and accounts for 97.2 percent of the total sum of squares in the dependent variable, identical to that accounted for by a third-degree polynomial. However, the condition value of the R matrix used to derive equation (16) is $5 \cdot 10^{-4}$, whereas that corresponding to the third-degree polynomial is $2 \cdot 10^{-16}$.

The stepwise regression analysis using all terms of Table 1 was repeated after translation of the coordinate system and again after changing the scale of the coordinate system. Translation was accomplished by adding a constant of 10 to each value of X and Y. Scale changes were made by multiplying X and Y by 1.5 and 0.75, respectively. These changes in the coordinate system are only two examples of the kinds of arbitrary changes that could be made.

On translation of the coordinate system, in




| | A | B | C | D | E |
|---|---|---|---|---|---|
| KIND OF REGRESSION: | A. LINEAR | B. QUADRATIC | C. STEPWISE USING LINEAR THROUGH QUARTIC POLYNOMIAL TERMS | D. STEPWISE USING LINEAR THROUGH QUARTIC POLYNOMIAL TERMS | E. STEPWISE USING ALL TERMS OF TABLE I |
| TREND EQUATION: | $T = -1076 + 31.5X - 0.95Y$ | $T = -1024 + 9.96X - 3.84Y + 1.6X^2 + 1.1XY - 0.27Y^2$ | $T = -1052 + 15.5X + 1.60X^2$ | $T = -1091 + 0.046(X+10)^3$ | $T = -975 + 2.53X^2 - 29.7\sqrt{Y} + 91.9 \text{Log}X \text{Log}Y$ |
| SUM OF SQUARES ACCOUNTED FOR BY TREND: | 94.7% | 96.7% | 95.8% | 95.7% | 96.8% |
| CONDITION VALUE OF R MATRIX: | 1.0 | $3 \times 10^{-5}$ | 0.025 | 1.0 | 0.12 |

CONTOUR INTERVAL 20 FEET — POSITIVE RESIDUALS (stippled) — NEGATIVE RESIDUALS (blank) — OUTLINE OF OIL POOL (heavy line)

Figure 5.- Trend and residual maps of structure on top of the "Mississippi chat", Lost Springs area, Kansas.

this problem, none of the terms selected prior to translation (i.e., $X^2$, $\sqrt{Y}$, and log X· logY) were significant at the 0.05 probability level. Instead, the same surface was obtained that resulted previously when the stepwise regression procedure was applied to linear through quartic polynomial terms (Fig. 5D). That is, the only significant term was $X_t^3$, where $X_t = X + 10$.

After changing the scale on X and Y the same terms selected prior to the scale change (i.e., $X^2$, $\sqrt{Y}$, and log X· log Y) were again selected in the stepwise regression procedure. The regression coefficients were different, but the condition value of the R matrix was nearly the same (0.09 vs 0.12). The sum of squares accounted for by the surface and the trend and residual maps were also essentially the same.

Translation of the coordinate system had large effects on the condition values of R matrices used to fit quadratic through quintic polynomial surfaces. These are summarized below:

| Polynomial surface | Before translation | After translation |
|---|---|---|
| Quadratic | $3 \cdot 10^{-5}$ | $4 \cdot 10^{-8}$ |
| Cubic | $2 \cdot 10^{-16}$ | $4 \cdot 10^{-26}$ |
| Quartic | $4 \cdot 10^{-35}$ | $<10^{-46}$ |
| Quintic | $<10^{-46}$ | $<10^{-46}$ |

These results are in accord with the conclusion by Mandelbaum (1963, p. 507) that matrix condition is improved where the origin of the coordinate system is placed nearer the center of the map area.

The principal conclusion reached after trend analysis of the structural data on the "Mississippi chat" using stepwise regression procedures is that the residual maps are not highly sensitive to either the mathematical functions used to estimate the trends or the coordinate system used to define X and Y. Moreover, although the R matrices used to derive the stepwise regression surfaces (Figs. 5C to 5E) have higher condition values than the R matrix used to derive the quadratic surface (Fig. 5B), the condition value for the latter seems satisfactorily high. The condition value of an R matrix used to derive a third-degree polynomial surface, however, is considerably lower than the surface of equally good fit to the observed data derived by means of stepwise regression (Equation 16).

EXAMPLE II - The Lansing Group, Kansas

Structure contour maps drawn on the top of the Lansing Group of Pennsylvanian age in Kansas are shown in Figure 6. The map in Figure 6A, which shows the entire State of Kansas, is adapted from Merriam, Winchell, and Atkinson (1958) and is based on elevation data from thousands of wells. The map in Figure 6B is based on data from 200 wells used in trend-analysis studies by Merriam and Harbaugh (1964) who also made the data available for the study described here.

Initially, conventional polynomial surfaces of first through fifth degree were fitted to the Lansing data. The percentage sum of squares accounted for and the condition values of the R matrices corresponding to the linear through quintic surfaces are as follows:

| Surface | Percent of total sum of squares | Condition value |
|---|---|---|
| Linear | 42.4 | 0.97 |
| Quadratic | 83.7 | $8.6 \cdot 10^{-6}$ |
| Cubic | 96.2 | $2.7 \cdot 10^{-18}$ |
| Quartic | 96.9 | $6.5 \cdot 10^{-41}$ |
| Quintic | 97.1 | $<10^{-46}$ |

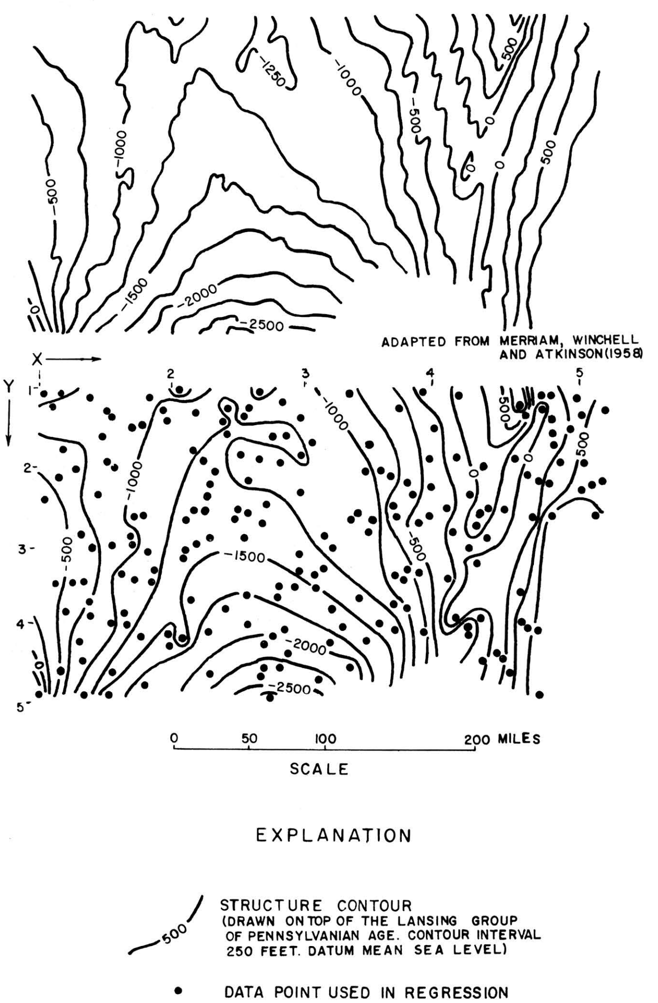
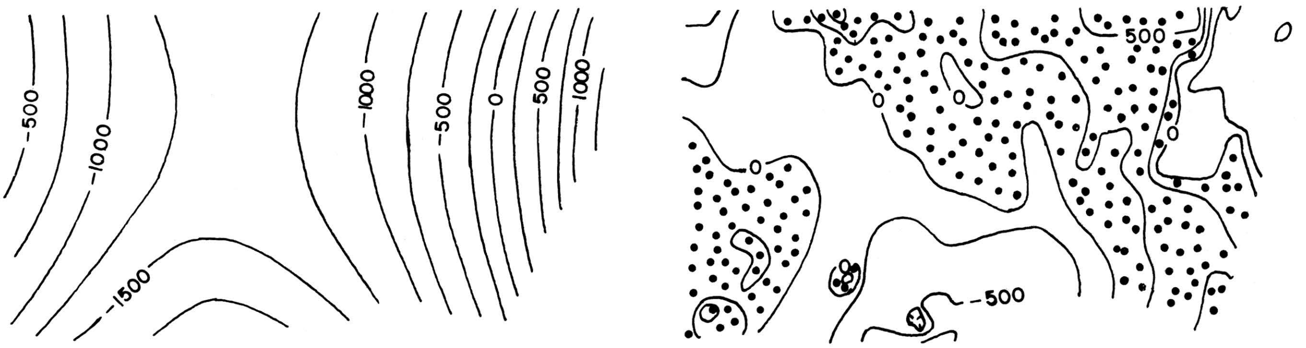



Figure 6.- Structure contour maps of the top of the Lansing Group in Kansas.

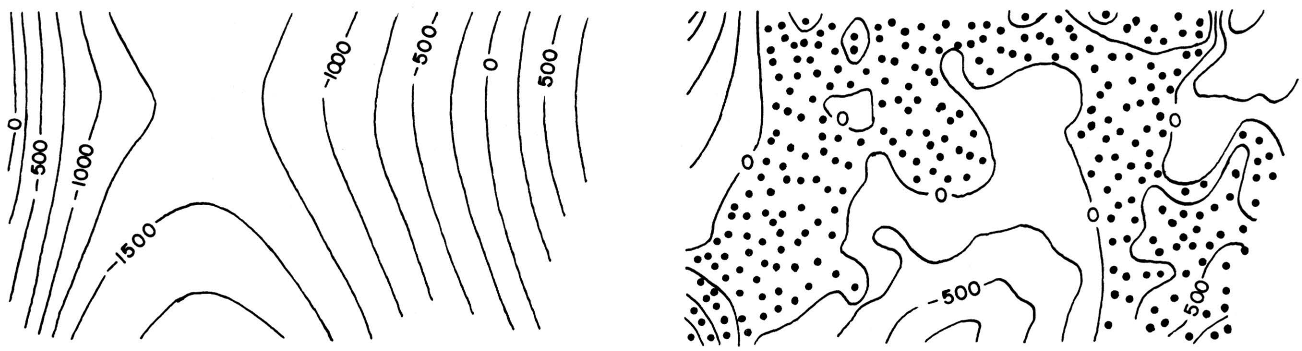
The cubic terms account for 12.5 percent of the total sum of squares in the dependent variable beyond the 83.7 percent accounted for by the linear plus quadratic terms. However, the condition values of the R matrices corresponding to the cubic and higher order surfaces are extremely small. Maps of the quadratic surface and residuals are given in Figure 7A.

A large number of other trend-surface functions were fitted to the Lansing data, most of which were determined by the stepwise regression technique using various levels of probability for entering and deleting terms, and using various coordinate schemes to define the X-Y map locations. Some functions were selected arbitrarily, however, and were tested by fitting them to the observed map data. One arbi-

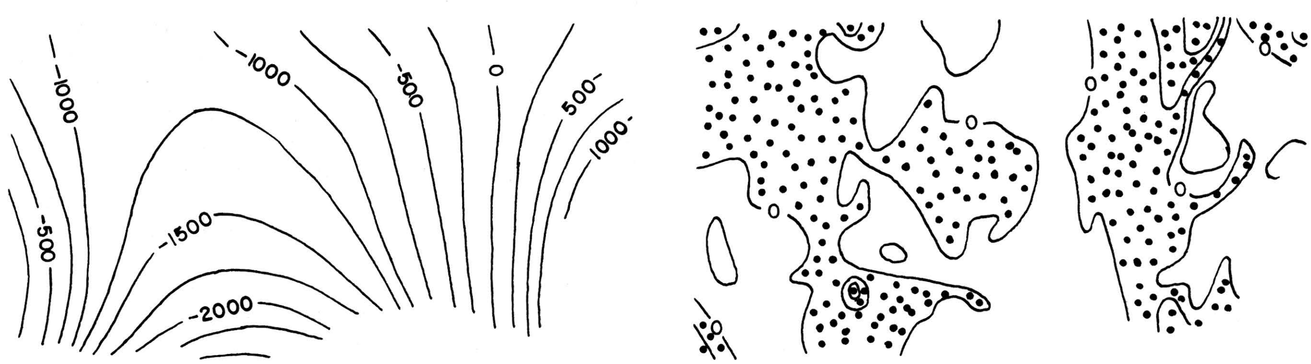
TREND MAPS RESIDUAL MAPS




A. Quadratic: $\text{Trend} = 1237 - 2179X + 280.9Y + 404.5X^2 + 29.41XY - 82.12Y^2$. 83.7% sum of squares accounted for by trend. Condition value of R matrix = $8.6 \times 10^{-6}$.




B. Arbitrary: $\text{Trend} = -1897 + 2457X - 515.5Y - 13{,}213\,\text{LOG}X + 2342\,\text{LOG}Y$. 84.4% sum of squares accounted for by trend. Condition value of R matrix = $4.0 \times 10^{-4}$.




C. Stepwise: $\text{Trend} = -1037 + 562.9Y + 230.8X^2 - 1126XY + 355.0Y^2 - 35.91X^3 + 190.8X^2Y - 49.26Y^3$. 96.0% sum of squares accounted for by trend. Condition value of R matrix = $6.0 \times 10^{-12}$.

Figure 7.- Trend and residual maps of the structure on top of the Lansing Group. Areas of positive residuals shown by dots. Contour interval is 250 feet.

trary function containing X, Y, log X, and log Y is shown in Figure 7B; a map of the residuals about the trend surface is also shown. The surface in Figure 7B accounts for slightly more of the total sum of squares than does the quadratic polynomial surface, but the function contains one less independent term and the R matrix has a higher condition value ( $4 \cdot 10^{-4}$). Residual maps derived from the quadratic polynomial surface (Fig. 7A) and the arbitrary surface (Fig. 7B) are different in one important respect; the former displays a northwestward trending "low" in the western part of the State and the latter displays a northeastward trending "high" in the same area. Both maps are equally valid in a mathematical sense, and although neither may reflect the underlying geologic structure correctly, both are worthy of geologic appraisal.

Stepwise regression performed at various levels of significance using all or part of the terms in Table 1 and changing either the origin or the scale of the coordinate system resulted in surfaces accounting for either a high percentage of the total sum of squares - generally greater than 96 percent-or a more moderate percentage - less than 90 percent. Those maps for which less than 90 percent of the total sum of squares is accounted for by the surface show residual configurations similar to either those of Figure 7A or Figure 7B, and, therefore, convey no additional possibilities for geologic evaluation. Also, because they are based on trend equations containing at least five independent terms, they are less efficient than the conventional quadratic or the arbitrarily selected surfaces (Figs. 7A, 7B).

Those residual maps derived from trend surfaces which account for high percentages of the total sum of squares show configurations that are not too dissimilar among themselves but that are different from those in Figures 7A and 7B. Maps based on one of the more efficient of the surfaces are shown in Figure 7C. The condition value of the R matrix is low but probably acceptable; the seven independent terms selected from the linear through cubic group of polynomial terms by the stepwise technique at the 0.01 level of significance account for only 0.2 percent less of the sum of squares than the entire cubic trend. Although the surfaces shown in Figure 7 account for 84 to 96 percent of the total sum of squares in the dependent variable, the configurations of the surfaces are all grossly similar. However, the minor differences among the trends are sufficient to cause large differences among the corresponding residual maps.

The positive and negative values in the residual maps in Figure 7 are probably clustered strongly enough to indicate at least moderate amounts of autocorrelation. The residuals, therefore, cannot be ascribed entirely to noise or to extremely local variation in the data. They appear to reflect either local structural features of the Lansing Group or the inadequacy of the regression equations in estimating the actual underlying regional trends. It is probable that they reflect both to varying degrees. The residual map in Figure 7C appears to show the locations of the Central Kansas uplift and the Nemaha ridge (Fig. 8) more clearly than the residual maps in Figures 7A and 7B, but the northward- or northwestward-trending "high" in the western part of the State does not correspond to any known structural feature in this region, and we have no good reason to propose that a new structural feature has been discovered. It seems more likely that the "high" is a result of the particular mathematical function used to define the surface, particularly because other functions (Figs. 7A, 7B) result in residual maps displaying a northeastward-trending "high," or a northwestward-trending "low" in the same area.

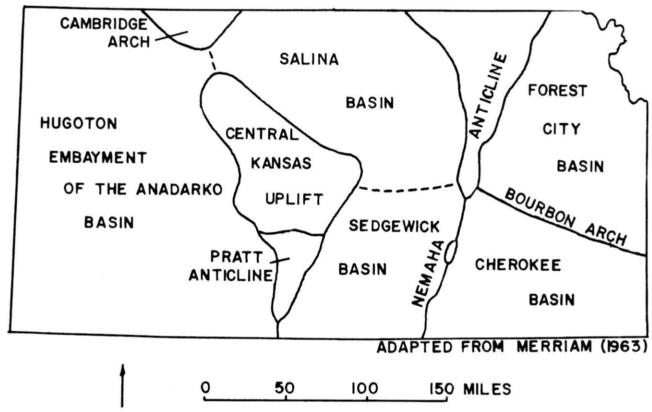



Figure 8.- Major pre-Des Moines and post-Mississippian structural features in Kansas.

## CONCLUSIONS

Trend analysis has proved to be a useful tool in the examination and interpretation of geologic map data and has been applied widely in many different kinds of geologic problems during the past 10 years. The method is comparatively straightforward where the purpose is to predict or interpolate values between the original control points on the map. However, where the purpose is to separate and describe regional and local components of variation, results can be sensitive to the mathematical function selected to estimate the regional trend. This is particularly true where a large proportion of the variance in the data can be described by a smooth continuous surface having only gentle flexure. Small differences among many such surfaces that may fit the data equally well can result in large differences in the configurations of contour maps of the residuals. If residual maps are examined for relation to local geologic factors, then the geologic conclusions may depend in part on the mathematical form of the trend function and, hence, might reflect subjective decisions made at the onset of the analysis.

This circumstance is caused by two factors: (1) too little is currently known about the quantitative nature of the geologic processes to select the proper form of a trend function on subject matter grounds, and (2) even if the true form of the trend function were known, its fit to the observed data by least-squares methods could easily be biased by the presence of local nonrandom variation. Because of these factors the procedures of trend analysis may remain empirical for some time and should be properly used in an exploratory manner with geologic factors playing a major part in evaluation of both the trend and residual maps. The use of trend analysis as a search technique was emphasized recently by Krumbein (1967, p. 42).

Use of polynomial equations in trend analysis offers several important advantages over some of the numerous other models that could be used. The principal advantage of polynomial terms is that they can approximate a wide variety of other functions. However, where local variation about the trend is small, the approximations may not be sufficiently good. Other configurations of trend residuals may be obtained using different, but equally valid, functions to estimate the trend. As an alternative to conventional use of polynomial equations, and current methods of adding or deleting terms in groups according to degree, we have used a stepwise regression method for selecting terms individually. Among the possible terms to be entered into the trend equation we include not only polynomial terms up to fifth degree, but also a number of others (Table 1). The stepwise regression can be performed at different probability levels and using different coordinate systems to define the X and Y map locations. Using these procedures it is possible to estimate the trend using a number of different surfaces of almost equally good fit to the observed data and to compile a number of different residual maps for any one set of map data. These, then, can be appraised in geologic terms.

A secondary, but important, benefit to be gained from the use of stepwise regression as outlined here is that fewer difficulties may arise in matrix computations. This results from both the elimination of terms in the regression equation which do little to reduce the total sum of squares in the dependent variable and the use of alternative terms that have lower degrees of linear correlation with others included in the trend equation.

The trend surfaces derived by stepwise regression methods can, in no sense, be regarded as more fundamentally correct than conventional polynomial or Fourier surfaces which may fit the observed data equally well. They are only alternatives which deserve equal consideration in trend analysis problems where the form of the trend function might make a difference in the final geologic analysis. One precaution that must be taken when using some nonpolynomial functions to describe geologic trends is the examination of their behavior between control points. Unlike low-order polynomial functions, they can exhibit unrealistic maxima or minima in such regions.

Of the two examples used in this paper, the first shows that the mathematical form of the trend function is of little or no importance in determination of the residual configurations. In the second example the configurations of the trend residuals are, in part, extremely sensitive to the type of trend function fitted to the data. Comparison and evaluation of the residual maps in this case can only be based on a priori knowledge of the structural features being searched for. The proper use of trend analysis in most geologic investigations, therefore, appears to be as a method of exploring numerical map data rather than as a routine analytical procedure.

## REFERENCES


Abramowitz, M., and Stegun, I.A., eds., 1964, Handbook of mathematical functions with formulas, graphs, and mathematical tables: Natl. Bur. Standards, Applied Math. Ser. 55, 1046 p.

Agterberg, F.P., 1964, Methods of trend surface analysis: Colorado Sch. Mines Quart., v. 59, no. 4, p. 111-130.

Agterberg, F.P., 1966, Trend surfaces with autocorrelated residuals: Pennsylvania State Univ., Mineral Industries, v. 1, p. H-1 - H-19.

Allen, P., and Krumbein, W.C., 1962, Secondary trend components in the top Ashdown Pebble Bed: A case history: Jour. Geology, v. 70, no. 5, p. 507-538.

Anderson, R.L., and Bancroft, T.A., 1952, Statistical theory in research: McGraw-Hill Book Co., Inc., New York, 399 p.

Booth, A.D., 1957, Numerical methods, 2nd ed: Academic Press, Inc., New York, 195 p.

Connor, J.J., and Miesch, A.T., 1964, Application of trend analysis to geochemical prospecting data from Beaver County, Utah: Stanford Univ. Publ., Geol. Sci., v. 9, no. 1, p. 110-125.

Efroymson, M.A., 1960, Multiple regression analysis, in Ralston, A., and Wilf, H.S., eds., Mathematical methods for digital computers: John Wiley & Sons, New York, p. 191-203.

Fisher, R.A., 1950, Statistical methods for research workers: Hafner Publishing Co., Inc., New York, 354 p.

Fox, L., 1965, An introduction to numerical linear algebra: Oxford Univ. Press, New York, 327 p.

James, W.R., 1966, FORTRAN IV program using double Fourier series for surface fitting of irregularly spaced data: Kansas Geol. Survey Computer Contr. 5, 19 p.

Krumbein, W.C., 1956, Regional and local components in facies maps: Am. Assoc. Petroleum Geologists Bull., v. 40, no. 9, p. 2163-2194.

Krumbein, W.C., 1959, Trend surface analysis of contour-type maps with irregular control-point spacing: Jour. Geophysical Res., v. 64, no. 7, p. 823-834.

Krumbein, W.C., 1966, A comparison of polynomial and Fourier models in map analysis: Office of Naval Research, Geography Branch, Tech. Rept. No. 2, ONR Task No. 388-078, 45 p.

Krumbein, W.C., 1967, The general linear model in map preparation and analysis, in Computer applications in the earth sciences: Colloquium on trend analysis, D.F. Merriam and N.C. Cocke, eds.: Kansas Geol. Survey Computer Contr. 12, p. 38-44.

Krumbein, W.C., and Graybill, F.A., 1965, An introduction to statistical models in geology: McGraw-Hill Book Co., Inc., New York, 475 p.

Macon, N., 1963, Numerical analysis: John Wiley & Sons, New York, 161 p.

Mandelbaum, H., 1963, Statistical and geological implications of trend mapping with nonorthogonal polynomials: Jour. Geophysical Res., v. 68, no. 2, p. 505-519.

Merriam, D.F., 1963, The geologic history of Kansas: Kansas Geol. Survey Bull. 162, 317 p.

Merriam, D.F., and Harbaugh, J.W., 1964, Trend-surface analysis of regional and residual components of geologic structure in Kansas: Kansas Geol. Survey Sp. Dist. Publ. 11, 27 p.

Merriam, D.F., Winchell, R.L., and Atkinson, W.R., 1958, Preliminary regional structural contour map on top of the Lansing Group (Pennsylvanian) in Kansas: Kansas Geol. Survey Oil and Gas Inv. Map no. 19.

Miesch, A.T., and Connor, J.J., 1967, Stepwise regression in trend analysis, in Computer applications in the earth sciences: Colloquium on trend analysis, D.F. Merriam and N.C. Cocke, eds.: Kansas Geol. Survey Computer Contr. 12, p. 16-18.

Miller, R.L., 1956, Trend surfaces: their application to analysis and description of environments of sedimentation: Jour. Geology, v. 64, p. 425-446.

Ostle, B., 1954, Statistics in research: The Iowa State Coll. Press, Ames, Iowa, 487 p.

Preston, F.W., and Harbaugh, J.W., 1965, BALGOL programs and geologic application for single and double Fourier series using IBM 7090/7094 computers: Kansas Geol. Survey Sp. Dist. Publ. 24, 72 p.

Read, W.A., and Merriam, D.F., 1966, Trend-surface analysis of stratigraphic thickness data from some Namurian rocks east of Stirling, Scotland: Scottish Jour. Geology, v. 2, pt. 1, p. 96-100.

Shenkel, C.W., Jr., 1955, Geology of the Lost Springs Pools area, Marion and Dickinson Counties, Kansas: Kansas Geol. Survey Bull. 114, pt. 6, p. 168-193.

Whitten, E.H.T., 1959, Composition trends in a granite: Modal variation and ghost stratigraphy in part of the Donegal Granite, Eire: Jour. Geophysical Res., v. 64, no. 7, p. 835-848.

APPENDIX A.- Sample calculation using the stepwise regression procedure.

The raw data used in this example are artificial, and consist of a matrix of pseudorandom normal deviates with 1000 rows and 5 columns augmented by an additional column generated as a function of $X_3$, $X_4$, and $X_5$ according to

$$X_0 = 6X_3 + 7X_4 + 8X_5$$

1., 2., and 3.[1/] The complete data matrix, $X_{r,c}$, consists of 1000 rows and 6 columns ( $1 \leqslant N \leqslant 1000$, $0 \leqslant c \leqslant 5$). The population means and standard deviations of $X_1$ through $X_5$ are all zero and one, respectively. Population correlation coefficients among pairs of $X_1$ through $X_5$ are as follows:

| | $X_1$ | $X_2$ | $X_3$ | $X_4$ | $X_5$ |
|---|---|---|---|---|---|
| $X_1$ | 1.000 | 0.500 | 0.995 | 1.000 | 0.000 |
| $X_2$ | ----- | 1.000 | 0.500 | 0.500 | 0.000 |
| $X_3$ | ----- | ----- | 1.000 | 0.995 | 0.000 |
| $X_4$ | ----- | ----- | ----- | 1.000 | 0.000 |
| $X_5$ | ----- | ----- | ----- | ----- | 1.000 |

Note that the correlation coefficient between $X_1$ and $X_4$ is 1.000, and as the means and standard deviations are the same, $X_{r,1} = X_{r,4}$ for all r. The probability level, Q, at which terms are to be selected or deleted is set at 0.05.

4. Starting at step 4 of the stepwise regression procedure, as outlined in the text, the means ($\bar{X}_c$) and the standard deviations ($s_c$) are estimated as:

| c | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|
| $\bar{X}_c$ | -0.13451 | -0.0093740 | 0.024055 | -0.012381 | -0.0093740 | 0.00067395 |
| $s_c$ | 15.637 | 0.993 | 0.998 | 0.997 | 0.993 | 1.035 |

5. The part of the matrix of estimated simple correlation coefficients ($r_{i,j}$) of concern is:

| | j = | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|---|
| i= | 0 | 1.00000 | 0.84671 | 0.43406 | 0.84871 | 0.84671 | 0.56457 |
| | 1 | 0.84671 | 1.00000 | 0.54278 | 0.99502 | 1.00000 | 0.04054 |
| | 2 | 0.43406 | ------- | 1.00000 | 0.53576 | 0.54278 | -0.02307 |
| | 3 | 0.84871 | ------- | ------- | 1.00000 | 0.99502 | 0.04490 |
| | 4 | 0.84671 | ------- | ------- | ------- | 1.00000 | 0.04054 |
| | 5 | 0.56457 | ------- | ------- | ------- | ------- | 1.00000 |

6. NDF = N - 1 = 999

7. VAR = 1

1/ Numbers correspond to those in text section on stepwise regression.

8. $C_i = -1$ for $i = 1, 2, \ldots, 5$.

9. $V_1 = (0.84671)^2/1 = 0.71692$; $C_1 = -1$; $r_{11} = 1.000$

$V_2 = (0.43406)^2/1 = 0.18841$; $C_2 = -1$; $r_{22} = 1.000$

$V_3 = (0.84871)^2/1 = 0.72031$; $C_3 = -1$; $r_{33} = 1.000$

$V_4 = (0.84671)^2/1 = 0.71692$; $C_4 = -1$; $r_{44} = 1.000$

$V_5 = (0.56457)^2/1 = 0.31874$; $C_5 = -1$; $r_{55} = 1.000$

9a. As all $C_i = -1$ at this stage, step 9a is skipped.

9b. Maximum $V_i = V_3 = VMAX = 0.72031$.

$F = (VMAX \cdot NDF)/(VAR - VMAX) = (0.72031 \cdot 999)/(1 - 0.72031) = 2572.81$

for 1 and 999 degrees of freedom.

QF estimated using equations 3 through 5 in the text is extremely small ($QF < Q = 0.05$). Therefore,

$k = 3$ and

$VAR = 1 - 0.72031 = 0.27969$

$NDF = 999 - 1 = 998$

10. (1) $r_{11} = 1.00000 - \dfrac{-1 \cdot -1 \cdot 0.99502 \cdot 0.99502}{1.00000} = 0.0099352$

$r_{12} = 0.54278 - \dfrac{-1 \cdot -1 \cdot 0.99502 \cdot 0.53576}{1.00000} = 0.0096881$

$r_{22} = 1.00000 - \dfrac{-1 \cdot -1 \cdot 0.53576 \cdot 0.53576}{1.00000} = 0.71296$

(2) $r_{13} = -\dfrac{0.99502}{1.00000} = -0.99502$

$r_{23} = -\dfrac{0.53576}{1.00000} = -0.53576$

(3) $r_{14} = 1.00000 - \dfrac{0.99502 \cdot 0.99502}{1.00000} = 0.0099352$

$r_{15} = 0.04054 - \dfrac{0.04490 \cdot 0.99502}{1.00000} = -0.0041364$

$r_{24} = 0.54278 - \dfrac{0.99502 \cdot 0.53576}{1.00000} = 0.0096881$

$r_{25} = -0.02307 - \dfrac{0.04490 \cdot 0.53576}{1.00000} = -0.047126$

(4) $r_{33} = \dfrac{1.00000}{1.00000} = 1.00000$

(5) $r_{34} = \dfrac{0.99502}{1.00000} = 0.99502$

$r_{35} = \dfrac{0.04490}{1.00000} = 0.04490$

(6) $r_{44} = 1.00000 - \frac{-1 \cdot -1 \cdot 0.99502 \cdot 0.99502}{1.00000} = 0.0099352$

$r_{45} = 0.04054 - \frac{-1 \cdot -1 \cdot 0.04490 \cdot 0.99502}{1.00000} = -0.0041364$

$r_{55} = 1.00000 - \frac{-1 \cdot -1 \cdot 0.04490 \cdot 0.04490}{1.00000} = 0.99798$

(7) $r_{10} = 0.84671 - \frac{0.84871 \cdot 0.99502}{1.00000} = 0.0022266$

$r_{20} = 0.43406 - \frac{0.84871 \cdot 0.53576}{1.00000} = -0.020645$

(8) $r_{30} = \frac{0.84871}{1.00000} = 0.84871$

(9) $r_{40} = 0.84671 - \frac{-1 \cdot -1 \cdot 0.84871 \cdot 0.99502}{1.00000} = 0.0022266$

$r_{50} = 0.56457 - \frac{-1 \cdot -1 \cdot 0.84871 \cdot 0.04490}{1.00000} = 0.52646$

$C_3 = +1$

9. $r_{i,j}$ is now:

| | j = | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|---|
| i = | 0 | 1.00000 | 0.84671 | 0.43406 | 0.84871 | 0.84671 | 0.56457 |
| | 1 | 0.0022266 | 0.0099352 | 0.0096881 | -0.99502 | 0.0099352 | -0.0041364 |
| | 2 | -0.020645 | --------- | 0.71296 | -0.53576 | 0.0096881 | -0.047126 |
| | 3 | 0.84871 | --------- | --------- | 1.00000 | 0.99502 | 0.04490 |
| | 4 | 0.0022266 | --------- | --------- | ------- | 0.0099352 | -0.0041364 |
| | 5 | 0.52646 | --------- | --------- | ------- | --------- | 0.99798 |

$V_1 = (0.0022266)^2/0.0099352 = 0.00049901$; $C_1 = -1$; $r_{11} = 0.0099352$

$V_2 = (-0.020645)^2/0.71296 = 0.00059781$; $C_2 = -1$; $r_{22} = 0.71296$

$V_3 = (0.84871)^2/1.00000 = 0.72031$; $C_3 = +1$; $r_{33} = 1.00000$

$V_4 = (0.0022266)^2/0.0099352 = 0.00049901$; $C_4 = -1$; $r_{44} = 0.0099352$

$V_5 = (0.52646)^2/0.99798 = 0.27772$; $C_5 = -1$; $r_{55} = 0.99798$

9a. Minimum $V_i = V_3 = VMIN = 0.72031$

$F = (VMIN \cdot NDF)/VAR = (0.72031 \cdot 998)/0.27969 = 2570.24$ for 1 and 998 degrees of freedom.

QF estimated using equations 3 through 5 in the text is extremely small, ($QF < Q = 0.05$). Therefore, go to 9b.

9b. Maximum $V_i = V_5 = VMAX = 0.27772$

$F = (VMAX \cdot NDF)/(VAR - VMAX) = (0.27772 \cdot 998)/(0.27969 - 0.27772) = 140,692$

for 1 and 998 degrees of freedom.

QF estimated using equations 3 through 5 in the text is extremely small ($QF < Q = 0.05$). Therefore

$k = 5$

$VAR = 0.27969 - 0.27772 = 0.00197$

$NDF = 998 - 1 = 997$

10. (1) $r_{11} = 0.0099352 - \dfrac{-1 \cdot -1 \cdot -0.0041364 \cdot -0.0041364}{0.99798} = 0.0099181$

$r_{12} = 0.0096881 - \dfrac{-1 \cdot -1 \cdot -0.0041364 \cdot -0.047126}{0.99798} = 0.0094928$

$r_{13} = -0.99502 - \dfrac{-1 \cdot 1 \cdot -0.0041364 \cdot 0.04490}{0.99798} = -0.99521$

$r_{14} = 0.0099352 - \dfrac{-1 \cdot -1 \cdot -0.0041364 \cdot -0.0041364}{0.99798} = 0.0099181$

$r_{22} = 0.71296 - \dfrac{-1 \cdot -1 \cdot -0.047126 \cdot -0.047126}{0.99798} = 0.71073$

$r_{23} = -0.53576 - \dfrac{-1 \cdot 1 \cdot -0.047126 \cdot 0.04490}{0.99798} = -0.53788$

$r_{24} = 0.0096881 - \dfrac{-1 \cdot -1 \cdot -0.047126 \cdot -0.0041364}{0.99798} = 0.0094928$

$r_{33} = 1.00000 - \dfrac{-1 \cdot 1 \cdot 0.04490 \cdot 0.04490}{0.99798} = 1.00202$

$r_{34} = 0.99502 - \dfrac{-1 \cdot -1 \cdot 0.04490 \cdot -0.0041364}{0.99798} = 0.99521$

$r_{44} = 0.0099352 - \dfrac{-1 \cdot -1 \cdot -0.0041364 \cdot -0.0041364}{0.99798} = 0.0099181$

(2) $r_{15} = - \dfrac{-0.0041364}{0.99798} = 0.0041448$

$r_{25} = - \dfrac{-0.047126}{0.99798} = 0.047221$

$r_{35} = - \dfrac{0.04490}{0.99798} = -0.044991$

$r_{45} = - \dfrac{-0.0041364}{0.99798} = 0.0041448$

(4) $r_{55} = \dfrac{1}{0.99798} = 1.00202$

(7) $r_{10} = 0.0022266 - \frac{0.52646 \cdot -0.0041364}{0.99798} = 0.0044087$

$r_{20} = -0.020645 - \frac{0.52646 \cdot -0.047126}{0.99798} = 0.0042152$

$r_{30} = 0.84871 - \frac{0.52646 \cdot 0.04490}{0.99798} = 0.82502$

$r_{40} = 0.0022266 - \frac{0.52646 \cdot -0.0041364}{0.99798} = 0.0044087$

(8) $r_{50} = \frac{0.52646}{0.99798} = 0.52753$

$C_5 = +1$

9. $r_{i,j}$ is now

| | j = | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|---|
| i = | 0 | 1.00000 | 0.84671 | 0.43406 | 0.84871 | 0.84671 | 0.56457 |
| | 1 | 0.0044087 | 0.0099181 | 0.0094928 | -0.99521 | 0.0099181 | 0.0041448 |
| | 2 | 0.0042152 | --------- | 0.71073 | -0.53788 | 0.0094928 | 0.047221 |
| | 3 | 0.82502 | --------- | --------- | 1.00202 | 0.99521 | -0.044991 |
| | 4 | 0.0044087 | --------- | --------- | ------- | 0.0099181 | 0.0041448 |
| | 5 | 0.52753 | --------- | --------- | ------- | --------- | 1.00202 |

$V_1 = (0.0044087)^2/0.0099181 = 0.0019597$; $C_1 = -1$; $r_{11} = 0.0099181$

$V_2 = (0.0042152)^2/0.71073 = 0.00002500$; $C_2 = -1$; $r_{22} = 0.71073$

$V_3 = (0.82502)^2/1.00202 = 0.67929$; $C_3 = +1$; $r_{33} = 1.00202$

$V_4 = (0.0044087)^2/0.0099181 = 0.0019597$; $C_4 = -1$; $r_{44} = 0.0099181$

$V_5 = (0.52753)^2/1.00202 = 0.27773$; $C_5 = +1$; $r_{55} = 1.00202$

9a. Minimum $V_i = V_5 = \text{VMIN} = 0.27773$

$F = (\text{VMIN} \cdot \text{NDF})/\text{VAR} = (0.27773 \cdot 997)/0.00197 = 140{,}557$ for 1 and 997 degrees of freedom.

QF estimated using equations 3 through 5 in the text is extremely small ($QF < Q = 0.05$). Therefore, go to 9b.

9b. Because $V_1 = V_4$, and the fact that $V_1$ precedes $V_4$, the maximum $V_i$ where $C_i = -1$ and $r_{i,i} >$ 0.00001 is taken as $V_1$.

Maximum $V_i = V_1 = \text{VMAX} = 0.0019597$

$F = (\text{VMAX} \cdot \text{NDF})/(\text{VAR} - \text{VMAX}) = (0.0019597 \cdot 997)/(0.00197 - 0.00196) = 195{,}382$ 2/ for 1 and 997 degrees of freedom.

2/ If more decimal places are carried in the demoninator variables this value is much larger.

QF estimated using equations 3 through 5 in the text is extremely small ($QF < Q = 0.05$). Therefore

$k = 1$

$VAR = 0.00197 - 0.00196 = 0.00001$

$NDF = 997 - 1 = 996$

10. (4) $r_{11} = \frac{1.0}{0.0099181} = 100.83$

(5) $r_{12} = \frac{0.0094928}{0.0099181} = 0.95712$

$r_{13} = \frac{-0.99521}{0.0099181} = -100.34$

$r_{14} = \frac{0.0099181}{0.0099181} = 1.00000$

$r_{15} = \frac{0.0041448}{0.0099181} = 0.41790$

(6) $r_{22} = 0.71073 - \frac{-1 \cdot -1 \cdot 0.0094928 \cdot 0.0094928}{0.0099181} = 0.70164$

$r_{23} = -0.53788 - \frac{-1 \cdot -1 \cdot -0.99521 \cdot 0.0094928}{0.0099181} = 0.41465$

$r_{24} = 0.0094928 - \frac{-1 \cdot -1 \cdot 0.0099181 \cdot 0.0094928}{0.0099181} = 0.00000$

$r_{25} = 0.047221 - \frac{-1 \cdot -1 \cdot 0.0041448 \cdot 0.0094928}{0.0099181} = 0.043254$

$r_{33} = 1.00202 - \frac{-1 \cdot 1 \cdot -0.99521 \cdot -0.99521}{0.0099181} = 100.86$

$r_{34} = 0.99521 - \frac{-1 \cdot 1 \cdot 0.0099181 \cdot -0.99521}{0.0099181} = 0.00000$

$r_{35} = -0.044991 - \frac{-1 \cdot 1 \cdot 0.0041448 \cdot -0.99521}{0.0099181} = -0.46089$

$r_{44} = 0.0099181 - \frac{-1 \cdot -1 \cdot 0.0099181 \cdot 0.0099181}{0.0099181} = 0.00000$

$r_{45} = 0.0041448 - \frac{-1 \cdot -1 \cdot 0.0041448 \cdot 0.0099181}{0.0099181} = 0.00000$

$r_{55} = 1.00202 - \frac{-1 \cdot 1 \cdot 0.0041448 \cdot 0.0041448}{0.0099181} = 1.00375$

(8) $r_{10} = \frac{0.0044087}{0.0099181} = 0.44451$

(9) $r_{20} = 0.0042152 - \frac{-1 \cdot -1 \cdot 0.0044087 \cdot 0.0094928}{0.0099181} = 0.00000$

$r_{30} = 0.82502 - \frac{-1 \cdot 1 \cdot 0.0044087 \cdot -0.99521}{0.0099181} = 0.38264$

$$r_{40} = 0.0044087 - \frac{-1 \cdot -1 \cdot 0.0044087 \cdot 0.0099181}{0.0099181} = 0.00000$$

$$r_{50} = 0.52753 - \frac{-1 \cdot 1 \cdot 0.0044087 \cdot 0.0041448}{0.0099181} = 0.52937$$

$C_1 = +1$

9. $r_{i,j}$ is now

| | j = | 0 | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|---|
| i= | 0 | 1.00000 | 0.84671 | 0.43406 | 0.84871 | 0.84671 | 0.56457 |
| | 1 | 0.44451 | 100.83 | 0.95712 | -100.34 | 1.00000 | 0.41790 |
| | 2 | 0.00000 | ------- | 0.70164 | 0.41465 | 0.00000 | 0.043254 |
| | 3 | 0.38264 | ------- | ------- | 100.86 | 0.00000 | -0.46089 |
| | 4 | 0.00000 | ------- | ------- | ------- | 0.00000 | 0.00000 |
| | 5 | 0.52937 | ------- | ------- | ------- | ------- | 1.00375 |

$V_1 = (0.44451)^2/100.83 = 0.00196$; $C_1 = +1$; $r_{11} = 100.83$

$V_2 = (0.00)^2/0.70164 = 0.00$; $C_2 = -1$; $r_{22} = 0.70164$

$V_3 = (0.38264)^2/100.86 = 0.00145$; $C_3 = +1$; $r_{33} = 100.86$

$V_4 = (0.00)^2/0.00 =$ undefined[3/] ; $C_4 = -1$; $r_{44} = 0.00000$

$V_5 = (0.52937)^2/1.00375 = 0.27919$; $C_5 = +1$; $r_{55} = 1.00375$

9a. Minimum $V_i = V_3 = 0.00145$

$F = (VMIN \cdot NDF)/VAR = (0.00145 \cdot 996)/0.00001 = 144{,}420$ for 1 and 996 degrees of freedom.

QF estimated using equations 3 through 5 in the text is extremely small ($QF < Q = 0.05$). Therefore, go to 9b.

9b. The only $V_i$ for which $C_i = -1$ and $r_{i,i} > 0.00001$ is $V_2$

Maximum $V_i = V_2 = VMAX = 0.0$

$F = (VMAX \cdot NDF)/(VAR - VMAX) = (0.0 \cdot 996)/(0.00001 - 0.0) = 0.0$ for 1 and 996 degrees of freedom.

QF estimated using equations 3 through 5 in the text is near 1.0 ($QF > Q = 0.05$). Therefore, go to 11.

11. The standardized partial regression coefficients are contained in the first column of the final $r_{i,j}$ matrix and are $\beta_1 = 0.44451$, $\beta_3 = 0.38264$, and $\beta_5 = 0.52937$. The final regression coefficients are:

$$b_1 = \frac{0.44451 \cdot 15.637}{0.993} = 7.00,$$

[3/] Division by zero is undefined and results here because of the "built-in" perfect linear correlation of $X_1$ and $X_4$.

$$b_3 = \frac{0.38264 \cdot 15.637}{0.997} = 6.00,$$

$$b_5 = \frac{0.52937 \cdot 15.637}{1.035} = 8.00,$$

and

$$b_0 = -0.135 - (7.00 \cdot -0.0093740) - (6.00 - 0.012381) - (8.00 \cdot 0.00067395) = 0.00.$$

12. $PSS = 100\,[(0.44451 \cdot 0.84671) + (0.38264 \cdot 0.84871) + (0.52937 \cdot 0.56457)] = 100.00.$

13. $T_r = 7.00X_1 + 6.00X_3 + 8.00X_5$, which in this case, because $X_1 = X_4$, is equivalent to

$$T_r = 6.00X_3 + 7.00X_4 + 8.00X_5.$$

The residuals are then derived from

$$d_r = X_{r,0} - T_r,$$

where $X_{r,0}$ is the column of the original data matrix containing the observed values of the dependent variable. As PSS = 100.00, all values of $d_r = 0$. The part of the original $r_{i,j}$ for which $C_i = +1$ is:

| | j= | 1 | 3 | 5 |
|---|---|---|---|---|
| i= | 1 | 1.00000 | 0.99502 | 0.04054 |
| | 3 | 0.99502 | 1.00000 | 0.04490 |
| | 5 | 0.04054 | 0.04490 | 1.00000 |

and the inverse is

| | | |
|---|---|---|
| 100.83 | -100.34 | 0.41790 |
| -100.34 | 100.86 | -0.46089 |
| 0.41790 | -0.46089 | 1.00375 |

## APPENDIX B – Computer Program Listing

Card deck set up (all of the following cards must be present)

Header card
Probability level card
Polynomial term card
Square root term card
Exponential term card
Logarithmic term card
Reciprocal term card
CN data cards

Header card

| | | |
|---|---|---|
| Columns | 1 - 30 | Format A30 Title |
| | 39 - 43 | Format I5 CN = number of data cards |
| | 44 - 46 | Format I3 CM = 1 + number of independent terms on table 1 to be used. |
| | 49 | Format I1 Option. If 0, trend residuals will not be computed. If 1, residuals will be computed. |

Probability level card: Probability level, Q, usually given as 0.01 (01 in columns 5 and 6) or 0.05 (05 in columns 5 and 6). Format F7.3. Set Q = 1.0 if selected terms are not to be eliminated by stepwise procedure.

Polynomial terms card: Punch the word LINEAR, QUADRATIC, CUBIC, QUARTIC, or QUINTIC (beginning in column 1) depending on highest order polynomial terms desired. If polynomial terms are not to be used, leave this card blank.

Square root terms card: Punch the word ROOT (beginning in column 1) if square root terms in Table 1 are desired. If not desired, leave this card blank.

Exponential terms card: Punch the word EXPONENTIAL (beginning in column 1 ) if exponential terms in Table 1 are desired. If not desired, leave this card blank.

Logarithmic terms card: Punch the word LOGARITHMIC (beginning in column 1) if the logarithmic terms in Table 1 are desired. If not desired, leave this card blank.

Reciprocal terms card: Punch the word RECIPROCAL (beginning in column 1) if reciprocal terms in Table 1 are desired. If not desired, leave this card blank.

Data cards: Format 3G7.0. Each of CN data cards contains observed values of the dependent variable, and the X and Y map coordinates, in that sequence.

NOTE: All polynomial and nonpolynomial terms that are selected to be entered and tested for use as independent variables are numbered consecutively from 2 to CM according to their sequence in Table 1, and are referred to by this number on the output. This is different, therefore, from the numbering scheme used in the foregoing text.

```
C       STEPWISE REGRESSION PROGRAM AFTER EFROYMSOM(1960).                    A  20
C       WRITTEN IN FORTRAN IV BY ROBERT TERRAZAS AFTER PROGRAMS BY
C       D.S. HANDWERKER AND G.I.SELNER, U.S.GEOLOGICAL SURVEY
C       THE PROGRAM IN ITS PRESENT STATE REQUIRES APPROXIMATELY
C       200 K BYTES OF CORE, DUE TO THE STORAGE OF THE Y(500,39) MATRIX
C       USED IN SUBSEQUENT COMPUTATION OF RESIDUALS.  THE STORAGE
C       REQUIREMENT MAY BE REDUCED SOME 78 BYTES BY USING AN EXTERNAL
C       DATA STORAGE FILE (TAPE OR DISK).
C       COMMENTS IN PROGRAM REFER TO DESCRIPTIONS OF PROCEDURE IN              A  40
C       FOREGOING TEXT.                                                        A  50
C       NOTE THAT SUBSCRIPT OF DEPENDENT VARIABLE IS ZERO IN TEXT, BUT         A  60
C       ONE IN THE PROGRAM.                                                    A  70
C                                                                              A  80
        REAL*8 B,Q,Y,CMR,REC,VMAX,SSQRS,R,Z,FR,QF,DEN,RES,CORR,VMIN,
       1STDEV,F,V,FAC,SUM,X,PSS,VAR,FNDF,PROD,SUMW,SEOUW,MEANS
        INTEGER TITLE,CN,CM,OPTION,RDIN,PROUT,C,CC,DEGF,BOOL                   A 100
        DIMENSION X(39),CORR(39,39),MEANS(39),STDEV(39),TITLE(8),
       1C(39),CC(39),B(39),R(39),Y(500,39),ITRANS(5),JTRANS(9),Z(3),V(39)
        DATA IBLANK/'    '/                                                    A 160
        DATA JTRANS/'LINE','QUAD','CUBI','QUAR','QUIN','ROOT','EXPO','LOGA    A 170
       1','RECI'/                                                              A 180
        RDIN=5                                                                 A 190
        PROUT=6                                                                A 200
C       STEP 1 STARTS HERE                                                     A 210
    110 READ (RDIN,660,END=650) TITLE,CN,CM,OPTION                             A 220
        WRITE (PROUT,680) TITLE,CN,CM                                          A 230
        READ (RDIN,670) Q                                                      A 240
        READ (RDIN,890) ITRANS                                                 A 250
        DO 120 I=1,CM                                                          A 260
        MEANS(I)=0.0                                                           A 270
        STDEV(I)=0.0                                                           A 280
        DO 120 J=1,CM                                                          A 290
    120 CORR(I,J)=0.0                                                          A 300
        K=4                                                                    A 310
        NSV=CM                                                                 A 320
        II=1                                                                   A 330
        SUMW=0.0                                                               A 340
C       STEPS 2 AND 3 START HERE                                               A 350
    130 READ (RDIN,900) (X(J),J=1,3)                                           A 360
C       IF SCALE AND ORIGIN OF COORDINATE SYSTEM ARE TO BE CHANGED,
C       INSERT TRANSFORMATIONS HERE.  TRANSFORMATIONS OF THE DEPENDENT
C       VARIABLE MAY ALSO BE DONE HERE.
        Z(2)=X(2)                                                              A 370
        Z(3)=X(3)                                                              A 380
        SUMW=SUMW+1.                                                           A 390
        IF (ITRANS(1).EQ.IBLANK) GO TO 140                                     A 400
        IF (ITRANS(1).EQ.JTRANS(1)) GO TO 150                                  A 410
C       THE GENERATION OF INDEPENDENT TERMS FOR THE STEPWISE REGRESSION
C       PROCEDURE BEGINS HERE.  THEY ARE GENERATED IN THE ORDER GIVEN
C       IN TABLE 1 OF THE TEXT.  OTHER TERMS WILL BE MORE APPROPRIATE
C       FOR SPECIFIC REGRESSION PROBLEMS.
        X(K)=X(2)**2                                                           A 420
        K=K+1                                                                  A 430
        X(K)=X(2)*X(3)                                                         A 440
        K=K+1                                                                  A 450
        X(K)=X(3)**2                                                           A 460
        K=K+1                                                                  A 470
        IF (ITRANS(1).EQ.JTRANS(2)) GO TO 150                                  A 480
        X(K)=X(4)*X(2)                                                         A 490
        K=K+1                                                                  A 500
```

```
      X(K)=X(4)*X(3)
      K=K+1
      X(K)=X(2)*X(6)
      K=K+1
      X(K)=X(3)*X(6)
      K=K+1
      IF (ITRANS(1).EQ.JTRANS(3)) GO TO 150
      X(K)=X(7)*X(2)
      K=K+1
      X(K)=X(7)*X(3)
      K=K+1
      X(K)=X(4)*X(6)
      K=K+1
      X(K)=X(2)*X(10)
      K=K+1
      X(K)=X(10)*X(3)
      K=K+1
      IF (ITRANS(1).EQ.JTRANS(4)) GO TO 150
      X(K)=X(11)*X(2)
      K=K+1
      X(K)=X(11)*X(3)
      K=K+1
      X(K)=X(7)*X(6)
      K=K+1
      X(K)=X(4)*X(10)
      K=K+1
      X(K)=X(2)*X(15)
      K=K+1
      X(K)=X(15)*X(3)
      K=K+1
      GO TO 150
  140 K=K-2
  150 IF (ITRANS(2).NE.JTRANS(6)) GO TO 160
      X(K)=Z(2)**.5
      K=K+1
      X(K)=X(K-1)*Z(3)**.5
      K=K+1
      X(K)=Z(3)**.5
      K=K+1
  160 IF (ITRANS(3).NE.JTRANS(7)) GO TO 170
      X(K)=DEXP(Z(2))
      K=K+1
      X(K)=DEXP(Z(3))
      K=K+1
      X(K)=DEXP(2.*Z(2))
      K=K+1
      X(K)=DEXP(Z(2)+Z(3))
      K=K+1
      X(K)=DEXP(2.*Z(3))
      K=K+1
  170 IF (ITRANS(4).NE.JTRANS(8)) GO TO 180
      X(K)=DLOG10(Z(2))
      K=K+1
      X(K)=DLOG10(Z(3))
      K=K+1
      X(K)=X(K-2)**2
      K=K+1
      X(K)=DLOG10(Z(2))*DLOG10(Z(3))
      K=K+1
      X(K)=(DLOG10(Z(3)))**2
```

```
      K=K+1
  180 IF (ITRANS(5).NE.JTRANS(9)) GO TO 190
      X(K)=1./Z(2)
      K=K+1
      X(K)=1./Z(3)
      K=K+1
      X(K)=1./Z(2)**2
      K=K+1
      X(K)=1./(Z(2)*Z(3))
      K=K+1
      X(K)=1./Z(3)**2
      K=K+1
C     STEPS 4 AND 5 START HERE
  190 DO 200 I=1,CM
      MEANS(I)=MEANS(I)+X(I)
      STDEV(I)=STDEV(I)+X(I)**2
      DO 200 J=1,CM
  200 CORR(I,J)=CORR(I,J)+X(I)*X(J)
      IF (OPTION.EQ.0) GO TO 220
      DO 210 J=1,CM
  210 Y(II,J)=X(J)
  220 II=II+1
      K=4
      IF (II.LE.CN) GO TO 130
      DO 230 I=1,CM
  230 STDEV(I)=DSQRT((SUMW*STDEV(I)-MEANS(I)**2)/(SUMW*(SUMW-1)))
      DO 240 I=1,CM
      DO 240 J=1,CM
  240 CORR(I,J)=(SUMW*CORR(I,J)-MEANS(I)*MEANS(J))/(STDEV(I)*STDEV(J)*SU
     1MW*(SUMW-1.))
      DO 250 I=1,CM
  250 MEANS(I)=MEANS(I)/SUMW
      WRITE (PROUT,810)
      WRITE (PROUT,800) (MEANS(I),I=1,NSV)
      WRITE (PROUT,790)
      WRITE (PROUT,800) (STDEV(I),I=1,NSV)
C     STEPS 6,7, AND 8 START HERE
C     THE MATRIX OF CORRELATION COEFFICIENTS (CORR) MAY BE PRINTED
C     HERE.  THE SUBSET OF THIS MATRIX CORRESPONDING TO THE INDEPENDENT
C     TERMS FINALLY SELECTED MAY THEN BE EXTRACTED AND NORMALIZED.
C     THE DETERMINANT OF THE NORMALIZED MATRIX IS THE CONDITION VALUE.
      DEGF=SUMW-1
      VAR=1.0
      C(1)=999999
      DO 260 I=2,NSV
  260 C(I)=-1
  270 DO 280 I=2,NSV
  280 V(I)=(CORR(I,1)**2)/CORR(I,I)
C     STEP 9A STARTS HERE
      VMIN=0.9999E49
      DO 300 I=2,NSV
      IF ((C(I).NE.1).OR.(CORR(I,I).LE.0.00001)) GO TO 290
      IF (V(I).GE.VMIN) GO TO 290
      K=I
      VMIN=V(I)
  290 CONTINUE
  300 CONTINUE
      IF (VMIN.EQ.0.9999E49) GO TO 310
      F=(VMIN*DEGF)/VAR
C     INSERT  COMP  OF OF
```

```
      IF (DEGF.GT.100) CALL QOFX (F,1,DEGF,QF)
      IF (DEGF.LE.100) CALL QOFF (F,1,DEGF,QF)
      IF (QF.LT.Q) GO TO 310
      VAR=VAR+VMIN
      DEGF=DEGF+1
      WRITE (PROUT,840) K,F,DEGF,QF
      GO TO 350
C     START OF 9B
  310 VMAX=0.0
      DO 330 I=2,NSV
      IF ((C(I).NE.-1).OR.(CORR(I,I).LE.0.00001)) GO TO 320
      IF (V(I).LE.VMAX) GO TO 320
      K=I
      VMAX=V(I)
  320 CONTINUE
  330 CONTINUE
      IF (VMAX) 640,460,340
  340 F=(VMAX*DEGF)/(VAR-VMAX)
C     COMP OF QF
      IF (DEGF.GT.100) CALL QOFX (F,1,DEGF,QF)
      IF (DEGF.LE.100) CALL QOFF (F,1,DEGF,QF)
      IF (QF.GT.Q) GO TO 460
      VAR=VAR-VMAX
      DEGF=DEGF-1
      WRITE (PROUT,830) K,F,DEGF,QF
C     START STEP 10
  350 REC=1.0/CORR(K,K)
      IK=K-1
      KI=K+1
      IF (IK.LT.2) GO TO 400
      DO 390 I=2,IK
      FAC=CORR(I,K)*REC
      CORR(I,1)=CORR(I,1)-CORR(K,1)*FAC
      DO 360 J=I,IK
      PROD=CORR(J,K)*FAC
      IFAC=C(K)*C(J)
      IF (IFAC.NE.1) PROD=-PROD
      CORR(I,J)=CORR(I,J)-PROD
  360 CONTINUE
      IF (KI.GT.NSV) GO TO 380
      DO 370 J=KI,NSV
      PROD=CORR(K,J)*FAC
      CORR(I,J)=CORR(I,J)-PROD
  370 CONTINUE
  380 CORR(I,K)=-FAC
  390 CONTINUE
  400 IF (KI.GT.NSV) GO TO 430
      DO 420 I=KI,NSV
      FAC=CORR(K,I)*REC
      IFAC=C(K)*C(I)
      IF (IFAC.NE.1) FAC=-FAC
      CORR(I,1)=CORR(I,1)-CORR(K,1)*FAC
      DO 410 J=I,NSV
      PROD=CORR(K,J)*FAC
      CORR(I,J)=CORR(I,J)-PROD
  410 CONTINUE
      CORR(K,I)=CORR(K,I)*REC
  420 CONTINUE
  430 CORR(K,1)=CORR(K,1)*REC
      CORR(K,K)=REC
```

```
      IF (C(K).EQ.-1) GO TO 440
      C(K)=-1
      GO TO 450
  440 C(K)=1
  450 CONTINUE
      GO TO 270
C     START STEP 11
  460 J=2
      DO 480 I=2,NSV
      IF (C(I).NE.1) GO TO 470
      B(J)=(CORR(I,1)*STDEV(1))/STDEV(I)
      CC(J-1)=I
      J=J+1
  470 CONTINUE
  480 CONTINUE
      K=J-1
      KK=K-1
      WRITE (PROUT,850) (CC(I),I=1,KK)
      WRITE (PROUT,860) (C(I),I=1,NSV)
      WRITE (PROUT,870)
      DO 510 II=1,KK
      I=CC(II)
      DO 500 JJ=II,KK
      J=CC(JJ)
      IF (II.EQ.1) GO TO 500
      III=II-1
      DO 490 L=1,III
  490 X(L)=0.0
  500 X(JJ)=CORR(I,J)
      WRITE (PROUT,880) (X(L),L=1,KK)
  510 CONTINUE
      SUM=0
      J=2
      DO 530 I=2,NSV
      IF (C(I).NE.1) GO TO 520
      SUM=SUM+(B(J)*MEANS(I))
      J=J+1
  520 CONTINUE
  530 CONTINUE
      B(1)=MEANS(1)-SUM
      WRITE (PROUT,820)
      WRITE (PROUT,800) (B(I),I=1,K)
      IF (K.EQ.1) GO TO 110
      PSS=0
      DO 550 I=2,NSV
      IF (C(I).NE.1) GO TO 540
      PSS=PSS+(CORR(I,1)*CORR(1,I))
  540 CONTINUE
  550 CONTINUE
      IDFN=J-2
      IK1=SUMW
      IDFD=IK1-1-IDFN
      NDF=IK1-1
      FAC=1.0-PSS
      DEN=FAC*IDFN
      FR=PSS*IDFD/DEN
      CMR=DSQRT(PSS)
      IF (CMR.GT.1.0) CMR=1.0
      WRITE (PROUT,720) CMR,IDFN,IDFD,FR
      SSQRS=NDF*FAC*STDEV(1)**2
```

```
      WRITE (PROUT,730) SSQRS
      SEOUW=SQRT(SSQRS/IDFD)
      FNDF=SUMW-1
      FAC=SEOUW/(STDEV(1)*SQRT(FNDF))
      J=2
      DO 570 I=2,NSV
      IF (C(I).LT.1) GO TO 560
      R(J)=CORR(I,1)
      J1=J+99
      R(J1)=FAC*SQRT(CORR(I,I))
      J=J+1
  560 CONTINUE
  570 CONTINUE
      IDFN=J-1
      WRITE (PROUT,770)
      WRITE (PROUT,800) (R(I),I=2,IDFN)
      WRITE (PROUT,760)
      I1=IDFN+99
      WRITE (PROUT,800) (R(I),I=101,I1)
      SUM=0.0
      J=2
      DO 610 I=2,NSV
      IF (C(I).LT.1) GO TO 600
      I1=J+99
      R(J)=STDEV(1)*R(I1)/STDEV(I)
      J=J+1
      SUM=SUM+CORR(I,I)*(MEANS(I)/STDEV(I))**2
      I1=I+1
      IF (I.EQ.NSV) GO TO 600
      DO 590 L=I1,NSV
      IF (C(L).LT.1) GO TO 580
      FAC=CORR(I,L)*MEANS(I)*MEANS(L)/(STDEV(I)*STDEV(L))
      SUM=SUM+FAC+FAC
  580 CONTINUE
  590 CONTINUE
  600 CONTINUE
  610 CONTINUE
      SUM=1.0/SUMW     +SUM/FNDF
      R(1)=SEOUW*DSQRT(SUM)
      WRITE (PROUT,780)
      WRITE (PROUT,800) (R(I),I=1,IDFN)
      PSS=100.0*PSS
      WRITE (PROUT,690) PSS
      IF (OPTION.EQ.0) GO TO 110
      WRITE (PROUT,700)
      DO 630 N=1,CN
      SUM=0.0
      J=2
      DO 620 I=2,NSV
      IF (C(I).NE.1) GO TO 620
      SUM=SUM+(B(J)*Y(N,I))
      J=J+1
  620 CONTINUE
      SUM=SUM+B(1)
      RES=Y(N,1)-SUM
      WRITE (PROUT,710) N,Y(N,1),SUM,RES
  630 CONTINUE
      GO TO 110
  640 WRITE (PROUT,910)
  650 STOP
```

```
C
  660 FORMAT (7A4,A2,8X,I5,I3,2X,I1)
  670 FORMAT (F7.3)
  680 FORMAT ('1',8X,'T I T L E',23X,6X,'N',4X,'M',2X/1X,7A4,A2,12X,I5,2
     1X,I3)
  690 FORMAT ('0','PERCENT OF TOTAL SUMS OF SQUARES OF DEPENDENT ','VARI
     1ABLE EXPLAINED = ',F7.3/)
  700 FORMAT ('1','  ROW                OBSERVED        COMPUTED    ','    RE
     1SIDUAL    '/)
  710 FORMAT (' ',I5,10X,2X,1PD13.5,2X,1PD13.5,2X,1PD13.5)
  720 FORMAT ('0','MULTIPLE CORRELATION COEFFICIENT = ',F11.7,'   F VALU
     1E FOR ',I3,' AND ',I5,' DEG OF FR = ',F11.3)
  730 FORMAT ('0','WTD SUM OF RESIDUALS SQUARED = ',D14.7)
  740 FORMAT ('0','PARTIAL CORRELATION COEFFICIENTS'/)
  750 FORMAT (' ',10F13.5/)
  760 FORMAT ('0','STANDARD ERROR OF REGRESSION WEIGHTS'/)
  770 FORMAT ('0','REGRESSION WEIGHTS'/)
  780 FORMAT ('0','STANDARD ERROR OF REGRESSION CONSTANT AND ','REGRESSI
     1ON COEFFICIENTS'/)
  790 FORMAT ('0','STANDARD DEVIATIONS')
  800 FORMAT (10(1X,    1PD12.5))
  810 FORMAT ('0','MEANS')
  820 FORMAT ('OREGRESSION CONSTANT AND COEFFICIENTS ','(CONSTANT FIRST)
     1'/)
  830 FORMAT ('OVARIABLE ',I5,' ADDED   F = ',D14.6,' FOR 1 AND ',I5,2X,
     1'DEGREES OF FREEDOM',8X,'Q COMPUTED = ',D14.6)
  840 FORMAT ('OVARIABLE ',I5,' DELETED F = ',D14.6,' FOR 1 AND ',I5,2X,
     1'DEGREES OF FREEDOM',8X,'Q COMPUTED = ',D14.6)
  850 FORMAT ('OINDEPENDENT VARIABLES IN THE REGRESSION EQUATION'/(10(1X
     1,I3)))
  860 FORMAT ('OC(I) ARRAY'/(10(1X,I3)))
  870 FORMAT ('OINVERSE MATRIX. (UPPER TRIANGLE AND DIAGONAL ONLY)')
  880 FORMAT (10(1X,1PD14.6))
  890 FORMAT (A4)
  900 FORMAT (3(G7.0))
  910 FORMAT ('1','VMAX LT 0.0'/)
      END
      SUBROUTINE QOFF (F,M,N,Q)
C     COMPUTES AREA OF F DISTRIBUTION WITH M AND N DEGREES OF FREEDOM
C     FROM F TO INFINITY
      INTEGER M,N,I,J,K,L,INUM,IDEN
      REAL*8 A,R,FAC,SINSQ,B,Q,SUM,SINTH,F,  TH,FDEN,COSSQ,
     1S,DEN,FNUM,TERM,COSTH
      IF ((F.GE.0.0).OR.(M.GE.1).OR.(N.GE.1)) GO TO 10
      Q=0.0
      GO TO 190
   10 R=DFLOAT(M)
      S=DFLOAT(N)
      IF ((MOD(M,2).NE.0).AND.(MOD(N,2).NE.0)) GO TO 20
      X=S/(S+R*F)
      IF (MOD(N,2).EQ.0) GO TO 170
      IF (MOD(N,2).EQ.0) GO TO 150
C     COMPUTE Q FOR M AND N BOTH ODD
   20 TH=DATAN(DSQRT(R*F/S))
      IF ((M.LE.0).OR.(N.LE.0)) GO TO 30
      SINTH=DSIN(TH)
      COSTH=DCOS(TH)
   30 IF (N-1) 50,40,50
   40 A=0.636620*TH
      GO TO 70
```

```
   50 SUM=COSTH
      TERM=COSTH
      FNUM=0.0
      FDEN=1.0
      COSSQ=COSTH*COSTH
      I1=N-2
      DO 60 I=3,I1,2
      FNUM=FNUM+2.0
      FDEN=FDEN+2.0
      TERM=TERM*FNUM*COSSQ/FDEN
      SUM=SUM+TERM
   60 CONTINUE
      A=0.636620*(TH+SINTH*SUM)
   70 IF (M-1) 90,80,90
   80 B=0.0
      GO TO 140
   90 INUM=(N+1)/2.0
      DEN=0.5*S
      FAC=1.27324
      K=INUM-1
      DO 100 J=2,K
      INUM=INUM-1
      DEN=DEN-1.0
      FAC=FAC*FLOAT(INUM)/DEN
  100 CONTINUE
      IF (N-1) 120,110,120
  110 FAC=0.5*FAC
  120 SUM=1.0
      TERM=1.0
      SINSQ=SINTH*SINTH
      FNUM=N-1
      FDEN=1.0
      K=M-3
      DO 130 I=2,K,2
      FNUM=FNUM+2
      FDEN=FDEN+2
      TERM=TERM*FNUM*SINSQ/FDEN
      SUM=SUM+TERM
  130 CONTINUE
      B=COSTH*N*FAC*SUM*SINTH
  140 Q=1.0-A+B
      GO TO 190
C     COMPUTATION OF Q FOR M EVEN FOLLOWS
  150 SUM=1.0
      TERM=1.0
      FNUM=N-2
      FDEN=0.0
      FAC=1.0-X
      K=(M-2)/2
      DO 160 I=1,K
      FNUM=FNUM+2.0
      FDEN=FDEN+2.0
      TERM=TERM*FNUM*FAC/FDEN
      SUM=SUM+TERM
  160 CONTINUE
      Q=X**(0.5*N)*SUM
      GO TO 190
C     COMPUTATION FOR N EVEN FOLLOWS
  170 SUM=1.0
      TERM=1.0
```

```
      FNUM=M-2.0
      FDEN=0.0
      K=(N-2)/2
      DO 180 I=1,K
      FNUM=FNUM+2.0
      FDEN=FDEN+2.0
      TERM=TERM*FNUM*X/FDEN
      SUM=SUM+TERM
  180 CONTINUE
      Q=1.0-(1.0-X)**(0.5*M)*SUM
  190 RETURN
      END
      SUBROUTINE QOFX (F,M,N,Q)
C     COMPUTES Q BY COMPUTING X THE NO OF STD DEV FROM THE MEAN
C     IN THE NORMAL DIST AND THEN COMPUTES THE AREA OF NORMAL
C     PROBABILITY CURVE FROM X TO INFINITY
      REAL*8 A,ZX,B,Q,FAC,F,T,X
      INTEGER M,N
      IF ((M.GT.0).AND.(N.GT.0).AND.(F.GT.0)) GO TO 10
      Q=0.0
      GO TO 20
   10 FAC=F**0.3333333333
      A=0.2222222222/M
      B=0.2222222222/N
      X=(FAC*(1.0-B)-(1.0-A))/DSQRT(A+B*FAC**2)
      T=1.0/(1.0+0.2316419*DABS(X))
      ZX=0.3989422804*DEXP(-0.5*X*X)
      FAC=((((1.330274429*T-1.821255978)*T+1.781477937)*T-0.356563782)*T
     1+0.319381530)*T
      FAC=FAC*ZX
      IF (X.LT.0) Q=1.0-FAC
      IF (X.GE.0) Q=FAC
   20 RETURN
      END
```

CONTROL CARDS FOR SAMPLE RUN NUMBER 1

```
LOST SPRINGS AREA, KANSAS                    81 39  1
    05
QUINTIC
ROOT
EXPONENTIAL
LOGARITHMIC
RECIPROCAL
```

CONTROL CARDS FOR SAMPLE RUN NUMBER 2

```
LOST SPRINGS AREA, KANSAS                    81  6  1
   1.0
QUADRATIC
              BLANK CARD
              BLANK CARD
              BLANK CARD
              BLANK CARD
```

DATA FOR SAMPLE RUNS

| | | |
|---|---|---|
| -1010. | 1 | 1 |
| -995. | 2 | 1 |
| -980. | 3 | 1 |
| -945. | 4 | 1 |
| -930. | 5 | 1 |
| -905. | 6 | 1 |
| -890. | 7 | 1 |
| -840. | 8 | 1 |
| -780. | 9 | 1 |
| -1010. | 1 | 2 |
| -998. | 2 | 2 |
| -980. | 3 | 2 |
| -960. | 4 | 2 |
| -970. | 5 | 2 |
| -900. | 6 | 2 |
| -880. | 7 | 2 |
| -835. | 8 | 2 |
| -790. | 9 | 2 |
| -1010. | 1 | 3 |
| -995. | 2 | 3 |
| -985. | 3 | 3 |
| -965. | 4 | 3 |
| -1000. | 5 | 3 |
| -880. | 6 | 3 |
| -840. | 7 | 3 |
| -830. | 8 | 3 |
| -790. | 9 | 3 |
| -1020. | 1 | 4 |
| -995. | 2 | 4 |
| -980. | 3 | 4 |
| -965. | 4 | 4 |
| -940. | 5 | 4 |
| -880. | 6 | 4 |
| -865. | 7 | 4 |
| -820. | 8 | 4 |
| -780. | 9 | 4 |
| -1045. | 1 | 5 |
| -1020. | 2 | 5 |
| -995. | 3 | 5 |
| -970. | 4 | 5 |
| -910. | 5 | 5 |
| -910. | 6 | 5 |
| -865. | 7 | 5 |
| -820. | 8 | 5 |
| -770. | 9 | 5 |
| -1040. | 1 | 6 |
| -1035. | 2 | 6 |
| -1020. | 3 | 6 |
| -990. | 4 | 6 |
| -940. | 5 | 6 |
| -880. | 6 | 6 |
| -860. | 7 | 6 |
| -810. | 8 | 6 |
| -780. | 9 | 6 |
| -1030. | 1 | 7 |
| -1015. | 2 | 7 |
| -990. | 3 | 7 |
| -960. | 4 | 7 |
| -890. | 5 | 7 |
| -900. | 6 | 7 |
| -860. | 7 | 7 |
| -830. | 8 | 7 |
| -780. | 9 | 7 |
| -1060. | 1 | 8 |
| -1040. | 2 | 8 |
| -1000. | 3 | 8 |
| -950. | 4 | 8 |
| -930. | 5 | 8 |
| -900. | 6 | 8 |
| -870. | 7 | 8 |
| -820. | 8 | 8 |
| -780. | 9 | 8 |
| -1070. | 1 | 9 |
| -1050. | 2 | 9 |
| -1000. | 3 | 9 |
| -940. | 4 | 9 |
| -930. | 5 | 9 |
| -910. | 6 | 9 |
| -865. | 7 | 9 |
| -820. | 8 | 9 |
| -790. | 9 | 9 |

# OUTPUT FROM SAMPLE RUN NO. 1

```
      T I T L E                              N     M
LOST SPRINGS AREA, KANSAS                   81    39

MEANS
-9.23185D 02  5.00000D 00  5.00000D 00  3.16667D 01  2.50000D 01  3.16667D 01  2.25000D 02  1.58333D 02  1.58333D 02  2.25000D 02
 1.70367D 03  1.12500D 03  1.00278D 03  1.12500D 03  1.70367D 03  1.34250D 04  8.51833D 03  7.12500D 03  7.12500D 03  8.51833D 03
 1.34250D 04  2.14511D 00  4.60150D 00  2.14511D 00  1.42415D 03  1.42415D 03  8.43743D 06  2.02819D 06  8.43743D 06  6.17751D-01
 6.17751D-01  4.68351D-01  3.81617D-01  4.68351D-01  3.14330D-01  3.14330D-01  1.71085D-01  9.88032D-02  1.71085D-01

STANDARD DEVIATIONS
 8.42142D 01  2.59808D 00  2.59808D 00  2.66393D 01  1.95576D 01  2.66393D 01  2.42513D 02  1.71000D 02  1.71000D 02  2.42513D 02
 2.16357D 03  1.48463D 03  1.38587D 03  1.48463D 03  2.16357D 03  1.92453D 04  1.29547D 04  1.16672D 04  1.16672D 04  1.29547D 04
 1.92453D 04  6.35200D-01  1.96825D 00  6.35200D-01  2.54742D 03  2.54742D 03  2.05422D 07  8.24107D 06  2.05422D 07  2.96342D-01
 2.96342D-01  3.05015D-01  2.73208D-01  3.05015D-01  2.70528D-01  2.70528D-01  3.03450D-01  1.40542D-01  3.03450D-01
```

| | | |
|---|---|---|
| VARIABLE 4 ADDED F = 0.143767D 04 FOR 1 AND 79 DEGREES OF FREEDOM | Q COMPUTED = -0.226934D-07 | $V_4 = X^2$ |
| VARIABLE 32 ADDED F = 0.215154D 02 FOR 1 AND 78 DEGREES OF FREEDOM | Q COMPUTED = 0.133650D-04 | $V_{32} = (\log X)^2$ |
| VARIABLE 38 ADDED F = 0.774788D 01 FOR 1 AND 77 DEGREES OF FREEDOM | Q COMPUTED = 0.674363D-02 | $V_{38} = (XY)^{-1}$ |
| VARIABLE 33 ADDED F = 0.586042D 01 FOR 1 AND 76 DEGREES OF FREEDOM | Q COMPUTED = 0.178398D-01 | $V_{33} = \log X \cdot \log Y$ |
| VARIABLE 24 ADDED F = 0.857454D 01 FOR 1 AND 75 DEGREES OF FREEDOM | Q COMPUTED = 0.449579D-02 | $V_{24} = \sqrt{Y}$ |
| VARIABLE 38 DELETED F = 0.949314D 00 FOR 1 AND 76 DEGREES OF FREEDOM | Q COMPUTED = 0.333026D 00 | |
| VARIABLE 32 DELETED F = 0.828271D 00 FOR 1 AND 77 DEGREES OF FREEDOM | Q COMPUTED = 0.365649D 00 | |

```
INDEPENDENT VARIABLES IN THE REGRESSION EQUATION
  4  24  33

C(I) ARRAY
***  -1  -1   1  -1  -1  -1  -1  -1  -1
 -1  -1  -1  -1  -1  -1  -1  -1  -1  -1
 -1  -1  -1   1  -1  -1  -1  -1  -1  -1
 -1  -1   1  -1  -1  -1  -1  -1  -1

INVERSE MATRIX. (UPPER TRIANGLE AND DIAGONAL ONLY)
  2.528475D 00   1.734184D 00  -2.621470D 00
  0.0            2.967578D 00  -2.974279D 00
  0.0            0.0            4.496053D 00

REGRESSION CONSTANT AND COEFFICIENTS (CONSTANT FIRST)

-9.74558D 02  2.52720D 00 -2.97100D 01  9.19153D 01

MULTIPLE CORRELATION COEFFICIENT =    0.9837603    F VALUE FOR    3 AND    77 DEG OF FR =     771.048

WTD SUM OF RESIDUALS SQUARED =  0.1827793D 05

REGRESSION WEIGHTS

 7.99423D-01 -2.24093D-01  2.98192D-01

STANDARD ERROR OF REGRESSION WEIGHTS

 3.25250D-02  3.52362D-02  4.33714D-02

STANDARD ERROR OF REGRESSION CONSTANT AND REGRESSION COEFFICIENTS

 8.49739D 00  1.02821D-01  4.67158D 00  1.33689D 01

PERCENT OF TOTAL SUMS OF SQUARES OF DEPENDENT VARIABLE EXPLAINED =  96.778
```

| ROW | OBSERVED | COMPUTED | RESIDUAL |
|---|---|---|---|
| 1 | -1.01000D 03 | -1.00174D 03 | -8.25878D 00 |
| 2 | -9.95000D 02 | -9.94160D 02 | -8.40380D-01 |
| 3 | -9.80000D 02 | -9.81524D 02 | 1.52362D 00 |
| 4 | -9.45000D 02 | -9.63833D 02 | 1.88332D 01 |
| 5 | -9.30000D 02 | -9.41088D 02 | 1.10884D 01 |
| 6 | -9.05000D 02 | -9.13289D 02 | 8.28921D 00 |
| 7 | -8.90000D 02 | -8.80436D 02 | -9.56439D 00 |
| 8 | -8.40000D 02 | -8.42528D 02 | 2.52761D 00 |
| 9 | -7.80000D 02 | -7.99565D 02 | 1.95652D 01 |
| 10 | -1.01000D 03 | -1.01405D 03 | 4.04749D 00 |
| 11 | -9.98000D 02 | -9.98137D 02 | 1.36607D-01 |
| 12 | -9.80000D 02 | -9.80628D 02 | 6.28291D-01 |
| 13 | -9.60000D 02 | -9.59481D 02 | -5.19073D-01 |
| 14 | -9.70000D 02 | -9.34055D 02 | -3.59453D 01 |
| 15 | -9.00000D 02 | -9.04065D 02 | 4.06461D 00 |
| 16 | -8.80000D 02 | -8.69359D 02 | -1.06414D 01 |
| 17 | -8.35000D 02 | -8.29846D 02 | -5.15396D 00 |
| 18 | -7.90000D 02 | -7.85468D 02 | -4.53172D 00 |
| 19 | -1.01000D 03 | -1.02349D 03 | 1.34904D 01 |
| 20 | -9.95000D 02 | -1.00271D 03 | 7.70722D 00 |
| 21 | -9.85000D 02 | -9.82349D 02 | -2.65122D 00 |
| 22 | -9.65000D 02 | -9.59179D 02 | -5.82077D 00 |
| 23 | -1.00000D 03 | -9.32184D 02 | -6.78155D 01 |
| 24 | -8.80000D 02 | -9.00913D 02 | 2.09128D 01 |
| 25 | -8.40000D 02 | -8.65123D 02 | 2.51232D 01 |
| 26 | -8.30000D 02 | -8.24672D 02 | -5.32798D 00 |
| 27 | -7.90000D 02 | -7.79466D 02 | -1.05337D 01 |
| 28 | -1.02000D 03 | -1.03145D 03 | 1.14512D 01 |
| 29 | -9.95000D 02 | -1.00721D 03 | 1.22110D 01 |
| 30 | -9.80000D 02 | -9.84830D 02 | 4.83038D 00 |
| 31 | -9.65000D 02 | -9.60226D 02 | -4.77394D 00 |
| 32 | -9.40000D 02 | -9.32118D 02 | -7.88160D 00 |
| 33 | -8.80000D 02 | -8.99937D 02 | 1.99374D 01 |
| 34 | -8.65000D 02 | -8.63379D 02 | -1.62092D 00 |
| 35 | -8.20000D 02 | -8.22262D 02 | 2.26189D 00 |
| 36 | -7.80000D 02 | -7.76469D 02 | -3.53122D 00 |
| 37 | -1.04500D 03 | -1.03846D 03 | -6.53526D 00 |
| 38 | -1.02000D 03 | -1.01154D 03 | -8.45684D 00 |
| 39 | -9.95000D 02 | -9.87594D 02 | -7.40601D 00 |
| 40 | -9.70000D 02 | -9.61877D 02 | -8.12323D 00 |
| 41 | -9.10000D 02 | -9.32906D 02 | 2.29059D 01 |
| 42 | -9.10000D 02 | -9.00020D 02 | -9.98040D 00 |
| 43 | -8.65000D 02 | -8.62865D 02 | -2.13507D 00 |
| 44 | -8.20000D 02 | -8.21231D 02 | 1.23118D 00 |
| 45 | -7.70000D 02 | -7.74982D 02 | 4.98243D 00 |
| 46 | -1.04000D 03 | -1.04481D 03 | 4.80549D 00 |
| 47 | -1.03500D 03 | -1.01569D 03 | -1.93070D 01 |
| 48 | -1.02000D 03 | -9.90462D 02 | -2.95377D 01 |
| 49 | -9.90000D 02 | -9.63836D 02 | -2.61643D 01 |
| 50 | -9.40000D 02 | -9.34160D 02 | -5.84045D 00 |
| 51 | -8.80000D 02 | -9.00697D 02 | 2.06970D 01 |
| 52 | -8.60000D 02 | -8.63055D 02 | 3.05509D 00 |
| 53 | -8.10000D 02 | -8.20999D 02 | 1.09993D 01 |
| 54 | -7.80000D 02 | -7.74378D 02 | -5.62177D 00 |
| 55 | -1.03000D 03 | -1.05064D 03 | 2.06364D 01 |
| 56 | -1.01500D 03 | -1.01967D 03 | 4.67157D 00 |
| 57 | -9.90000D 02 | -9.93357D 02 | 3.35725D 00 |
| 58 | -9.60000D 02 | -9.65962D 02 | 5.96193D 00 |
| 59 | -8.90000D 02 | -9.35689D 02 | 4.56894D 01 |
| 60 | -9.00000D 02 | -9.01740D 02 | 1.73961D 00 |
| 61 | -8.60000D 02 | -8.63686D 02 | 3.68575D 00 |
| 62 | -8.30000D 02 | -8.21273D 02 | -8.72692D 00 |
| 63 | -7.80000D 02 | -7.74337D 02 | -5.66272D 00 |
| 64 | -1.06000D 03 | -1.05606D 03 | -3.93630D 00 |
| 65 | -1.04000D 03 | -1.02349D 03 | -1.65057D 01 |
| 66 | -1.00000D 03 | -9.96241D 02 | -3.75868D 00 |
| 67 | -9.50000D 02 | -9.68180D 02 | 1.81800D 01 |
| 68 | -9.30000D 02 | -9.37391D 02 | 7.39094D 00 |
| 69 | -9.00000D 02 | -9.03019D 02 | 3.01908D 00 |
| 70 | -8.70000D 02 | -8.64608D 02 | -5.39162D 00 |
| 71 | -8.20000D 02 | -8.21887D 02 | 1.88659D 00 |
| 72 | -7.80000D 02 | -7.74678D 02 | -5.32187D 00 |
| 73 | -1.07000D 03 | -1.06116D 03 | -8.83888D 00 |
| 74 | -1.05000D 03 | -1.02718D 03 | -2.28237D 01 |
| 75 | -1.00000D 03 | -9.99095D 02 | -9.04538D-01 |
| 76 | -9.40000D 02 | -9.70447D 02 | 3.04467D 01 |
| 77 | -9.30000D 02 | -9.39202D 02 | 9.20202D 00 |
| 78 | -9.10000D 02 | -9.04458D 02 | -5.54213D 00 |
| 79 | -8.65000D 02 | -8.65732D 02 | 7.32400D-01 |
| 80 | -8.20000D 02 | -8.22738D 02 | 2.73795D 00 |
| 81 | -7.90000D 02 | -7.75289D 02 | -1.47110D 01 |

# OUTPUT FROM SAMPLE RUN NO. 2

```
       T I T L E                                          N     M
LOST SPRINGS AREA, KANSAS                                 81    6

MEANS
-9.23185D 02  5.00000D 00  5.00000D 00  3.16667D 01  2.50000D 01  3.16667D 01

STANDARD DEVIATIONS
 8.42142D 01  2.59808D 00  2.59808D 00  2.66393D 01  1.95576D 01  2.66393D 01

VARIABLE    4 ADDED    F =   0.143767D 04 FOR 1 AND    79  DEGREES OF FREEDOM      Q COMPUTED =  -0.226934D-07    V4 = X^2

VARIABLE    2 ADDED    F =   0.212844D 02 FOR 1 AND    78  DEGREES OF FREEDOM      Q COMPUTED =   0.147259D-04    V2 = X

VARIABLE    6 ADDED    F =   0.204599D 01 FOR 1 AND    77  DEGREES OF FREEDOM      Q COMPUTED =   0.156601D 00    V6 = Y^2

VARIABLE    5 ADDED    F =   0.171842D 02 FOR 1 AND    76  DEGREES OF FREEDOM      Q COMPUTED =   0.860719D-04    V5 = XY

VARIABLE    3 ADDED    F =   0.135464D 01 FOR 1 AND    75  DEGREES OF FREEDOM      Q COMPUTED =   0.248110D 00    V3 = Y

INDEPENDENT VARIABLES IN THE REGRESSION EQUATION
  2   3   4   5   6

C(I) ARRAY
***   1   1   1   1   1

INVERSE MATRIX. (UPPER TRIANGLE AND DIAGONAL ONLY)
  2.423052D 01    3.750000D 00   -1.997426D 01   -5.645795D 00   -9.925394D-14
  0.0             2.423052D 01   -9.825233D-14   -5.645795D 00   -1.997426D 01
  0.0             0.0             2.048052D 01    1.479233D-13    2.587464D-27
  0.0             0.0             0.0             8.500000D 00    1.485478D-13
  0.0             0.0             0.0             0.0             2.048052D 01

REGRESSION CONSTANT AND COEFFICIENTS (CONSTANT FIRST)

-1.02384D 03  9.96305D 00 -3.84034D 00  1.60101D 00  1.11056D 00 -2.65873D-01

MULTIPLE CORRELATION COEFFICIENT =   0.9836151   F VALUE FOR   5 AND    75 DEG OF FR =     446.519

WTD SUM OF RESIDUALS SQUARED =  0.1844007D 05

REGRESSION WEIGHTS

 3.07368D-01 -1.18478D-01  5.06443D-01  2.57912D-01 -8.41029D-02

STANDARD ERROR OF REGRESSION WEIGHTS

 1.02471D-01  1.02471D-01  9.42087D-02  6.06918D-02  9.42087D-02

STANDARD ERROR OF REGRESSION CONSTANT AND REGRESSION COEFFICIENTS

 1.13186D 01  3.32151D 00  3.32151D 00  2.97820D-01  2.61336D-01  2.97820D-01

PERCENT OF TOTAL SUMS OF SQUARES OF DEPENDENT VARIABLE EXPLAINED =  96.750
```

| ROW | OBSERVED | COMPUTED | RESIDUAL |
|---|---|---|---|
| 1 | -1.01000D 03 | -1.01527D 03 | 5.27354D 00 |
| 2 | -9.95000D 02 | -9.99397D 02 | 4.39690D 00 |
| 3 | -9.80000D 02 | -9.80318D 02 | 3.18249D-01 |
| 4 | -9.45000D 02 | -9.58038D 02 | 1.30376D 01 |
| 5 | -9.30000D 02 | -9.32555D 02 | 2.55488D 00 |
| 6 | -9.05000D 02 | -9.03870D 02 | -1.12983D 00 |
| 7 | -8.90000D 02 | -8.71983D 02 | -1.80166D 01 |
| 8 | -8.40000D 02 | -8.36895D 02 | -3.10532D 00 |
| 9 | -7.80000D 02 | -7.98604D 02 | 1.86039D 01 |
| 10 | -1.01000D 03 | -1.01880D 03 | 8.80094D 00 |
| 11 | -9.98000D 02 | -1.00181D 03 | 3.81375D 00 |
| 12 | -9.80000D 02 | -9.81625D 02 | 1.62455D 00 |
| 13 | -9.60000D 02 | -9.58233D 02 | -1.76668D 00 |
| 14 | -9.70000D 02 | -9.31640D 02 | -3.83599D 01 |
| 15 | -9.00000D 02 | -9.01845D 02 | 1.84480D 00 |
| 16 | -8.80000D 02 | -8.68848D 02 | -1.11525D 01 |
| 17 | -8.35000D 02 | -8.32648D 02 | -2.35180D 00 |
| 18 | -7.90000D 02 | -7.93247D 02 | 3.24687D 00 |
| 19 | -1.01000D 03 | -1.02286D 03 | 1.28601D 01 |
| 20 | -9.95000D 02 | -1.00476D 03 | 9.76235D 00 |
| 21 | -9.85000D 02 | -9.83463D 02 | -1.53741D 00 |
| 22 | -9.65000D 02 | -9.58961D 02 | -6.03920D 00 |
| 23 | -1.00000D 03 | -9.31257D 02 | -6.87430D 01 |
| 24 | -8.80000D 02 | -9.00351D 02 | 2.03512D 01 |
| 25 | -8.40000D 02 | -8.66243D 02 | 2.62433D 01 |
| 26 | -8.30000D 02 | -8.28933D 02 | -1.06654D 00 |
| 27 | -7.90000D 02 | -7.88422D 02 | -1.57842D 00 |
| 28 | -1.02000D 03 | -1.02745D 03 | 7.45100D 00 |
| 29 | -9.95000D 02 | -1.00824D 03 | 1.32427D 01 |
| 30 | -9.80000D 02 | -9.85832D 02 | 5.83238D 00 |
| 31 | -9.65000D 02 | -9.60220D 02 | -4.77996D 00 |
| 32 | -9.40000D 02 | -9.31406D 02 | -8.59432D 00 |
| 33 | -8.80000D 02 | -8.99389D 02 | 1.93893D 01 |
| 34 | -8.65000D 02 | -8.64171D 02 | -8.29105D-01 |
| 35 | -8.20000D 02 | -8.25750D 02 | 5.75047D 00 |
| 36 | -7.80000D 02 | -7.84128D 02 | 4.12803D 00 |
| 37 | -1.04500D 03 | -1.03257D 03 | -1.24264D 01 |
| 38 | -1.02000D 03 | -1.01225D 03 | -7.74521D 00 |
| 39 | -9.95000D 02 | -9.88734D 02 | -6.26609D 00 |
| 40 | -9.70000D 02 | -9.62011D 02 | -7.98899D 00 |
| 41 | -9.10000D 02 | -9.32086D 02 | 2.20861D 01 |
| 42 | -9.10000D 02 | -8.98959D 02 | -1.10408D 01 |
| 43 | -8.65000D 02 | -8.62630D 02 | -2.36979D 00 |
| 44 | -8.20000D 02 | -8.23099D 02 | 3.09923D 00 |
| 45 | -7.70000D 02 | -7.80366D 02 | 1.03662D 01 |
| 46 | -1.04000D 03 | -1.03823D 03 | -1.77197D 00 |
| 47 | -1.03500D 03 | -1.01680D 03 | -1.82014D 01 |
| 48 | -1.02000D 03 | -9.92167D 02 | -2.78328D 01 |
| 49 | -9.90000D 02 | -9.64334D 02 | -2.56663D 01 |
| 50 | -9.40000D 02 | -9.33298D 02 | -6.70173D 00 |
| 51 | -8.80000D 02 | -8.99061D 02 | 1.90608D 01 |
| 52 | -8.60000D 02 | -8.61621D 02 | 1.62127D 00 |
| 53 | -8.10000D 02 | -8.20980D 02 | 1.09797D 01 |
| 54 | -7.80000D 02 | -7.77136D 02 | -2.86382D 00 |
| 55 | -1.03000D 03 | -1.04441D 03 | 1.44142D 01 |
| 56 | -1.01500D 03 | -1.02187D 03 | 6.87420D 00 |
| 57 | -9.90000D 02 | -9.96132D 02 | 6.13222D 00 |
| 58 | -9.60000D 02 | -9.67188D 02 | 7.18821D 00 |
| 59 | -8.90000D 02 | -9.35042D 02 | 4.50422D 01 |
| 60 | -9.00000D 02 | -8.99694D 02 | -3.05863D-01 |
| 61 | -8.60000D 02 | -8.61144D 02 | 1.14407D 00 |
| 62 | -8.30000D 02 | -8.19392D 02 | -1.06080D 01 |
| 63 | -7.80000D 02 | -7.74438D 02 | -5.56213D 00 |
| 64 | -1.06000D 03 | -1.05113D 03 | -8.86795D 00 |
| 65 | -1.04000D 03 | -1.02748D 03 | -1.25185D 01 |
| 66 | -1.00000D 03 | -1.00063D 03 | 6.28990D-01 |
| 67 | -9.50000D 02 | -9.70574D 02 | 2.05744D 01 |
| 68 | -9.30000D 02 | -9.37318D 02 | 7.31785D 00 |
| 69 | -9.00000D 02 | -9.00859D 02 | 8.59242D-01 |
| 70 | -8.70000D 02 | -8.61199D 02 | -8.80138D 00 |
| 71 | -8.20000D 02 | -8.18336D 02 | -1.66402D 00 |
| 72 | -7.80000D 02 | -7.72271D 02 | -7.72869D 00 |
| 73 | -1.07000D 03 | -1.05838D 03 | -1.16183D 01 |
| 74 | -1.05000D 03 | -1.03362D 03 | -1.63794D 01 |
| 75 | -1.00000D 03 | -1.00566D 03 | 5.65751D 00 |
| 76 | -9.40000D 02 | -9.74492D 02 | 3.44924D 01 |
| 77 | -9.30000D 02 | -9.40125D 02 | 1.01253D 01 |
| 78 | -9.10000D 02 | -9.02556D 02 | -7.44391D 00 |
| 79 | -8.65000D 02 | -8.61785D 02 | -3.21508D 00 |
| 80 | -8.20000D 02 | -8.17812D 02 | -2.18828D 00 |
| 81 | -7.90000D 02 | -7.70636D 02 | -1.93635D 01 |

KANSAS GEOLOGICAL SURVEY COMPUTER PROGRAM
THE UNIVERSITY OF KANSAS, LAWRENCE

PROGRAM ABSTRACT

Title (If subroutine state in title):

Trend analysis using stepwise regression and polynomial and/or nonpolynomial models

Date: August 7, 1968

Author, organization: Robert Terrazas, U.S. Geological Survey

Direct inquiries to: Author or A.T. Miesch, U.S. Geological Survey

Name: Address: Federal Center

Denver, Colorado 80225

Purpose/description: To search and examine map data by fitting least-square surfaces.

Mathematical method: Stepwise regression.

Restrictions, range: Number of data points cannot exceed 500 unless dimension statement for Y is adjusted. See comments in program.

Computer manufacturer: IBM Model: 360/65

Programming language: FORTRAN IV

Memory required: 220 K Approximate running time: For CN=81 and CM=39, Go Time = 15 sec.

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)

Basic technique after Efroymson (1960). This program is a modification of previous programs by D.S. Handwerker and G.I. Selner, U.S. Geological Survey.

COMPUTER CONTRIBUTIONS

Kansas Geological Survey
University of Kansas
Lawrence, Kansas

Computer Contribution