#118

VOL. 7, NO. 3 WATER RESOURCES BULLETIN JUNE 1971

DETERMINING SIGNIFICANCE AND PRECISION OF ESTIMATED PARAMETERS FOR RUNOFF FROM SEMIARID WATERSHEDS[1]

H. B. Osborn, L. J. Lane, and R. S. Kagan[2]

ABSTRACT. Significant parameters for predicting thunderstorm runoff from small semiarid watersheds are determined using data from the Walnut Gulch watershed in southern Arizona. Based on these data, thunderstorm rainfall is dominant over watershed parameters for predicting runoff from multiple linear regression equations. In some cases antecedent moisture added significantly to the models. A technique is developed for estimating precision of predicted values from multiple linear regression equations. The technique involves matrix methods in estimating the variance of mean predicted values from a regression equation. The estimated variance of the mean predicted value is then used to estimate the variance of an individual predicted value. A computer program is developed to implement these matrix methods and to form confidence limits on predicted values based on both a normality assumption and the Chebyshev inequality.
(KEY WORDS: regression analysis; hydrology; parameter; estimation; individual prediction; precision)

INTRODUCTION

Hydrologists in water resources investigations constantly look for significant relationships between hydrologic variables. At the same time, engineers look for usable design tools. One reference listing a need for hydrologists and engineers to be involved in water resources investigations is a report published by the USDA-SCS [1970]. Of particular interest in the Southwest is the need for significant relationships to predict runoff on semiarid watersheds. Also, of vital interest is the precision with which estimated parameters can be predicted from significant variables in hydrologic models. In this paper parameters that should influence runoff from semiarid watersheds are investigated for significance. Then the precision of estimates of runoff is determined, using the significant parameters in regression equations. Data from the Walnut Gulch Experimental Watershed in southeastern Arizona are used in the analyses.

EXPERIMENTAL WATERSHED

The Southwest Watershed Research Center of the Agricultural Research Service operates the 58-square-mile Walnut Gulch Experimental Watershed in southeastern Arizona. This gently rolling rangeland watershed is an ephemeral tributary of the north-flowing San Pedro River (Figure 1). The upper one-third of the watershed is primarily grassland; the lower two-thirds of the watershed is primarily brush covered. The watershed is typical of much of the

[1] Paper No. 71043 of the *Water Resources Bulletin* (Journal of the American Water Resources Association). Discussions are open until December 1, 1971. Contribution of the Agricultural Research Service, Soil and Water Conservation Research Division, USDA, in cooperation with the Arizona Agricultural Experiment Station, Tucson, Arizona.

[2] Research Hydraulic Engineer, Hydrologist, and Engineering Technician, respectively, Southwest Watershed Research Center, Tucson, Arizona 85705.

grazed rangelands throughout the Southwest. There are 97 weighing-type recording rain gages and 20 permanent runoff-measuring structures within the 58-square-mile watershed. A more detailed description of the watershed and instrumentation was given by Renard [1970].

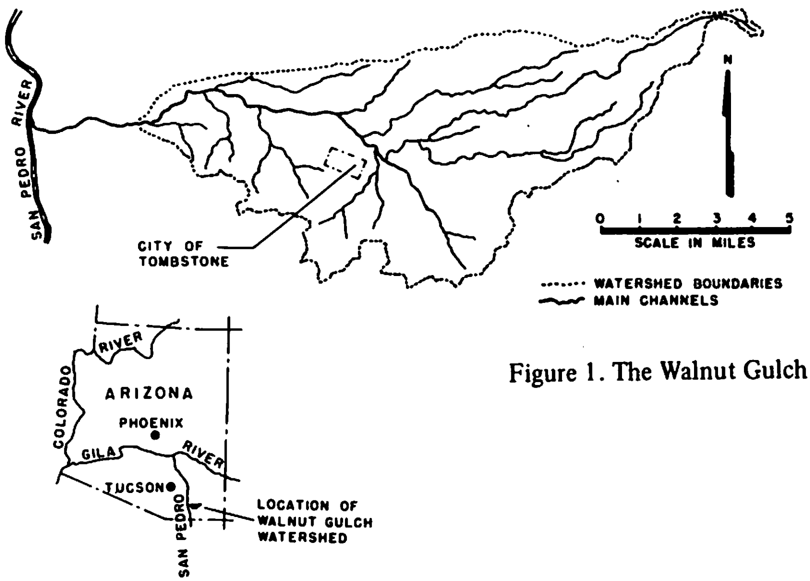


Figure 1. The Walnut Gulch watershed.

PRECIPITATION

Sellers [1960], Osborn and Reynolds [1963] and others have reported on the nature of precipitation in the Southwest. There are two rainy seasons. Summer precipitation is characterized by high-intensity thunderstorm rainfall of limited areal extent. Atmospheric moisture for these storms flows into the region from the Gulf of Mexico. The combination of moist air aloft and convective heating produces air mass thunderstorms that occur more or less randomly within large regions. Almost all rangeland runoff in the Southwest results from summer thunderstorms; thus, these storms are of primary interest to hydrologists involved in rangeland runoff studies in the Southwest.

Winter precipitation, on the other hand, is generally low-intensity rain or snow covering relatively large areas. Winter precipitation results from Pacific storms which have lost most of their moisture before reaching eastern Arizona. More often than not, these storms bring only cloudiness to southeastern Arizona with no measurable precipitation.

SIGNIFICANT VARIABLES

Many precipitation and watershed parameters affect runoff. These include volume and intensity of rainfall, antecedent rainfall and runoff, watershed area, vegetation and soil, channel slope, and the ability of the channel and watershed to abstract water from surface runoff. Many of these characteristics are not entirely independent. They become variables when they differ between and during actual storm events. Some parameters seem to be more important than others and may mask the variation of less important parameters. For example, thunderstorm rainfall for short durations may vary from 0 to 10 inches per hour both in time and space during an individual event on an individual watershed. Since thunderstorms are of relatively short duration, infiltration, on a single soil type, might be expected to vary less than one-tenth as much as rainfall during the storm. Similarly, the actual difference in runoff from differences in vegetation under instances of extreme rainfall variability may be relatively small even though grass- and brush-covered watersheds appear very different to the eye. It is important to understand all the possible variables in watershed runoff, but it is much more important to determine which ones give meaningful results in rainfall-runoff determinations.

Osborn and Lane [1969] used multiple linear regression to investigate runoff from very small watersheds (½ to 11 acres) on Walnut.Gulch for three years of record. Variables in rainfall explained 70 to 80 percent of the variability in peak discharge. Also, of the many parameters that were investigated, only antecedent rainfall improved the basic rainfall-runoff equation significantly, and then only on 1 of the 4 watersheds. Watershed parameters such as slope and area did not add significantly to the relationship. In other words, rainfall completely dominated rainfall-runoff relationships, at least within the accuracy of the data and for those variables that were measured.

In this paper, with 3 added years of record, multiple linear regression (MLR) was again used to investigate which variables might add significantly to rainfall-runoff relationships. The analyses were made on the same group of 4 very small watersheds (½ to 11 acres) and also on a 6-square-mile watershed within the Walnut Gulch watershed. Similar results were obtained for the 4 very small watersheds with rainfall the only significant variable.

When data from the 4 watersheds were lumped together, the equation with the most significant input variable was

$$Q = 2.8\, P_{15} - 0.53 \quad \text{with } R^2 = 0.76 \quad \text{and SEQ} = 0.29 \text{ cfs/acre,} \tag{1}$$

where Q = peak discharge in cfs/acre,
P_{15} = maximum 15-minute depth of rainfall in inches,
SEQ = standard error of estimate of Q,
and R^2 = coefficient of determination.

When total storm rainfall in inches, P_{tot}, was added, the equation became

$$Q = 2.0\, P_{15} + 0.53\, P_{tot} - 0.55 \quad \text{with } R^2 = 0.78 \quad \text{and SEQ} = 0.28 \text{ cfs/acre.} \tag{2}$$

Maximum 15-minute rainfall was the dominant input variable, explaining about 76 percent of the runoff variability. Total storm rainfall was the only other variable that added significantly to the relationship.

The precision of prediction for runoff from very small watersheds (up to 11 acres in size) for Walnut Gulch using Eq. (2) is discussed in a later section of this paper.

For the 6-square-mile watershed, the regression equation with the most significant input variable was

$$Q = 0.0029\, \text{V} - 0.12 \tag{3}$$

with $R^2 = 0.79$

and SEQ = 0.14 cfs/acre,

where Q = peak discharge in cfs/acre,
and V = acre-feet of rainfall above 0.5-inch depth.

Volume of rainfall above 0.5-inch depth explained about 79 percent of the variability in runoff. When the second significant variable was added the equation became

$$Q = 0.0030\, V + 0.19\, R_a - 0.16 \tag{4}$$

with $R^2 = 0.86$

SEQ = 0.11 cfs/ acre,

and R_a = antecedent runoff index.

Antecedent runoff was determined by adding daily estimates of infiltration into the channel alluvium from previous runoff and subtracting estimates of daily loss from deep percolation and evapotranspiration from the channel alluvium. For the 6-square-mile watershed, the antecedent moisture condition of the channel above the station was significant in determining peak discharge. No other variables added significantly to the relationship between rainfall volume and peak discharge.

PRECISION OF PREDICTION

David and Neyman [1938], Draper and Smith [1966], and others have suggested methods for determining variances for predicted values from regression equations of the form

$$Y = C_o + C_1 X_1 + C_2 X_2 + \dots C_n X_n . \tag{5}$$

These methods employ least squares fitting of the sample data. The methods generally indicate the variance of the mean of the predicted value, since the variance of the mean is what is usually wanted by statisticians. Hydrologists and engineers, however, are quite often interested in the variance of an individual predicted value.

A technique for estimating the variance of an individual predicted value from a multiple linear regression equation has been developed at the Southwest Watershed Research Center, Agricultural Research Service, Tucson, Arizona. The variance of the mean predicted value is estimated, using matrix methods described by David and Neyman [1938]. Then the estimated variance about the regression line, calculated as the sum of squares of residuals over the residual degrees of freedom, is used to convert the variance of the mean predicted value to the variance for a single predicted value. Confidence limits are determined assuming normality as the best possible condition and the Chebyshev inequality as the worst.

Using matrix notation from Draper and Smith [1966], the mean predicted value, assuming normality and the desired confidence limits, is

$$Y = \hat{Y} \pm t\,(v, 1 - \frac{a}{2}) \cdot s \cdot \sqrt{1/n + X_o' C X_o} \tag{6}$$

where $\hat{Y}$ = the predicted value
n = number of observations

s = estimated standard deviation about the regression line, calculated as the square root of the sum of squares of residuals over the residual degrees of freedom
t = student's t, with v degrees of freedom
v = n-k-1
k = number of independent variables
X_o = column vector for specific value
C = $(X'X)^{-1}$
X = matrix of all points
X' = transpose of X
a = significance level.

For a single predicted value, within 95% confidence limits, assuming normality,

$$Y = \hat{Y} \pm t(v, 0.975) \cdot s \cdot \sqrt{1 + 1/n + X_o' C X_o} \tag{7}$$

Assuming the Chebyshev inequality, which would hold for any distribution, within the 94% confidence limits,

$$Y = \hat{Y} \pm 4 \cdot s \cdot \sqrt{1 + 1/n + X_o' C X_o} \tag{8}$$

Using the notation of David and Neyman [1938] Eq. (6) becomes

$$Y = \hat{Y} \pm t(v, 1½) \cdot U_F^2 \tag{6a}$$

where U_F^2 is the estimate of the variance of the mean predicted value. Eq. (7) becomes

$$Y = \hat{Y} \pm t(v,1½) \cdot [S^2 + U_F^2]^{½} \tag{7a}$$

and Eq. (8) becomes

$$Y = \hat{Y} \pm 4 \cdot [S^2 + U_F^2]^{½} \tag{8a}$$

where the assumptions and variables are as stated above.

A computer program has been prepared to determine the coefficients in Eq. (5) and the conficence limits as described by Eqs. (7a) and (8a). The program can handle 1,000 observations for 9 independent variables and one dependent variable. Descriptive statistics are derived for the input data and statistics are computed to test for significance of the derived equation. A listing of this program, in Fortran IV, can be obtained from the Southwest Watershed Research Center, Tucson, Arizona. A flow chart of the method is shown in the appendix.

Although statistics to test significance of the regression equation are provided, the program is not designed to determine the proper variables. The program is designed to give a reasonable estimate of the accuracy of predicted values, assuming the form of the equation is specified. Again, both normal and Chebyshev confidence limits are calculated, but the choice of the proper distribution for the residuals is left to the user. The array of residuals is calculated and then printed for the purpose of testing distributional assumptions.

RESULTS

The Walnut Gulch Experimental Watershed contains intensive study areas where on-site runoff and runoff from very small watersheds are measured. Four of these unit source watersheds, from 0.5 to 11.0 acres in size, were selected to relate peak discharge, Q, to total precipitation, P_{tot}, and to the maximum 15-minute depth of precipitation, P_{15}, for events where Q was greater than 0.1 cfs/acre. The multiple regression equation developed from 85 runoff events as shown earlier was

$$Q = 2.0\ P_{15} + 0.53\ P_{tot} - 0.55 \tag{2}$$

with coefficient of determination $R^2 = 0.785$, and standard error of estimate SEQ = 0.28 cfs/acre. Eq. (2) was used to predict the peak discharge resulting from rains of 2.0 inches with 15-minute depths ranging from 0.5 inch (2.0 inches per hour) to 2.0 inches (8.0 inches perhour). These predictions and their confidence limits based on the normal assumption, and the Chebyshev inequality, as described in Eqs. (7a) and (8a), respectively, are shown in Figure 2.

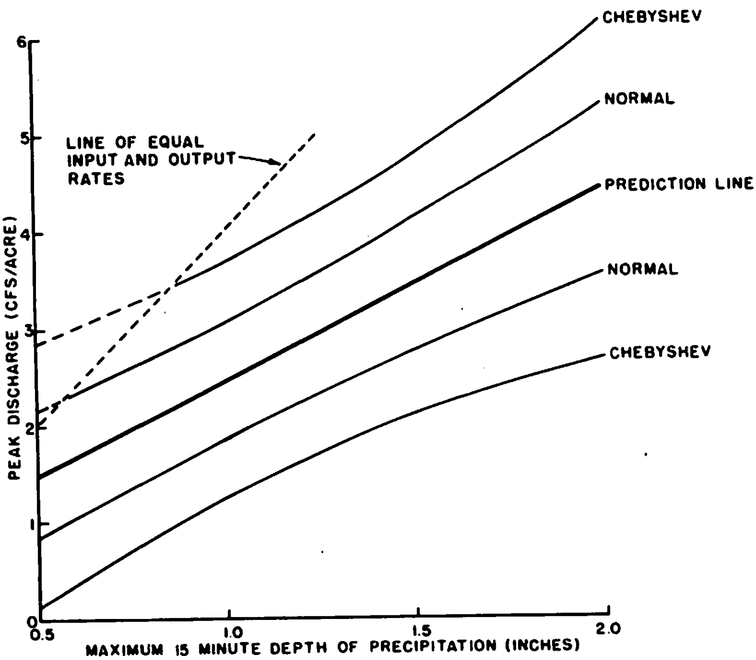


Figure 2. Peak discharge vs maximum 15-minute precipitation for total precipitation of 2.0 inches with confidence limits on predicted values for Lucky Hills watershed, Walnut Gulch.

As shown earlier, the regression equation for the 6-square-mile watershed developed from 25 events was

$$Q = 0.0030\ V + 0.19\ R_a - 0.16. \tag{4}$$

Eq. (4) was used to predict peak discharge resulting from runoff-producing rains (over 0.5-

inch depth) of 4.0, 6.0, 8.0, and 10.0 acre-feet of rain. These predictions and the confidence limits based on the normal and Chebyshev assumption are shown in Figure 3.

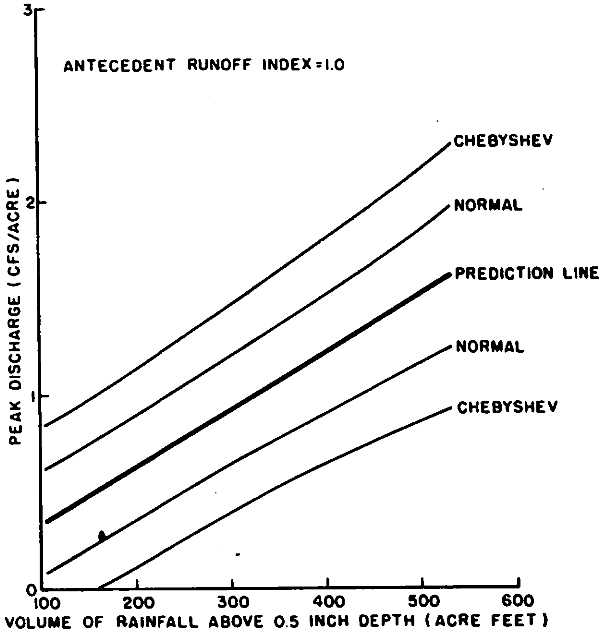


Figure 3. Peak discharge vs volume of rainfall above 0.5 inch depth with confidence limits on predicted values for W-5 watershed, Walnut Gulch.

In both cases, the normal assumptions might be the "best" that one could expect of the prediction, and the Chebyshev inequality represents the "worst." The prediction is probably no better than the normal assumption and no worse than the Chebyshev. Obviously, in many cases the limits of the Chebyshev assumptions will be so great as to render the predictions valueless, unless the normal assumption is valid, or at least nearly valid. In a vew cases, of course, even the normal assumption is not sufficient to justify the prediction.

OTHER USES

Although the program was developed to study rainfall-runoff relationships on the Walnut Gulch Experimental Watershed, other uses are possible. For example, the method might be used to predict annual yield as a function of basin characteristics as shown in Table 1. For nine Geological Survey drainage basins in southeastern Arizona, with corresponding periods of record, mean annual runoff volume, Q_a, in inches, was related to drainage area, A, in square miles, and to gaging station elevation, E, in feet. The multiple regression equation developed was

$$Q_a = -0.605 - 6\text{x}10^{-5}A + 32\text{x}10^{-5}E \quad (9)$$

with coefficient of determination $R^2 = 0.954$, and standard error of estimate, $SEQ_a = 0.073$

inch. Figure 4. shows predicted versus actual mean annual runoff for these nine watersheds. The 95% confidence limits that would be placed on each prediction are also shown. Since the dependent variable was mean annual runoff, confidence limits based on a normal assumption were used, although as stated, confidence limits based on the Chebyshev inequality are included in the printout.

TABLE 1. Characteristics of selected USGS watersheds in southern Arizona.

Location	USGS watershed Id. No.	Drainage area (mi.^{2})	Gage station elevation (feet MSL)	Period of record (yrs.)	Mean annual runoff volume (inches)
Nogales	94805	533.0	3702	28	0.59
Continental	94820	1662.0	2836	24	.17
Lochiel	94800	82.2	4620	20	.90
Patagonia	94815	209.0	3828	28	.55
Tucson	94860	918.0	2284	28	.13
Solomon	94570	2192.0	2960	28	.09
Palomines	94707	741.0	4188	28	.68
Charleston	94710	1219.0	3954	28	.63
Redington	94720	2939.0	2941	21	.23

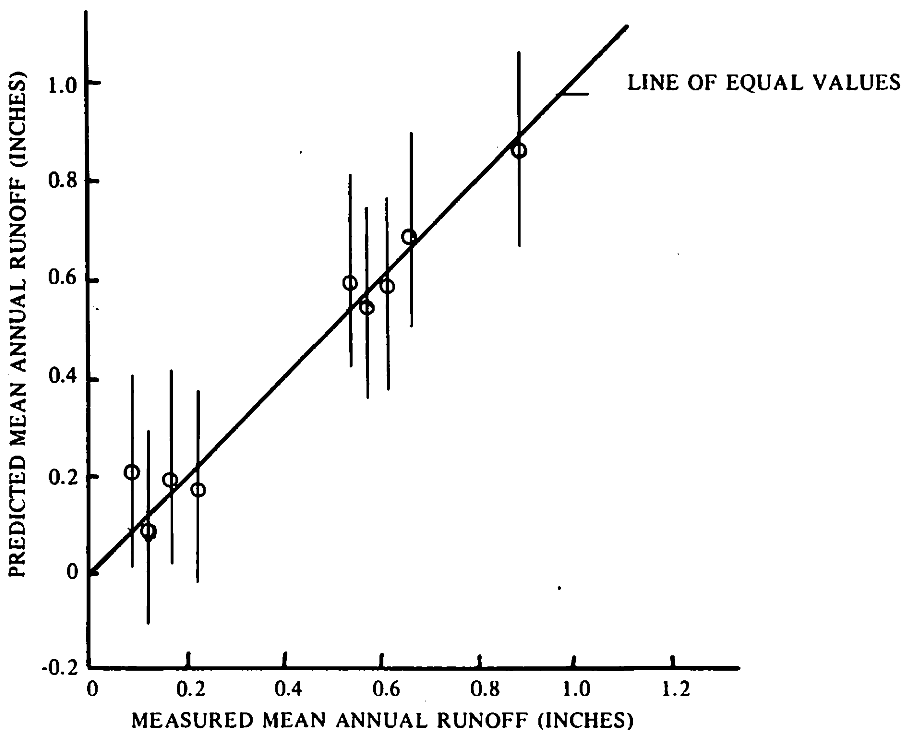


Figure 4. Predicted vs actual mean annual runoff from nine USGS stations in southern Arizona, showing confidence limits on the individual values.

SUMMARY

1. A technique was developed to determine the precision of individual predictions from multiple linear regression equations.
2. In this study precipitation variables were dominant over both antecedent moisture and watershed characteristics for predicting peak discharge from very small semiarid watersheds.
3. The most significant precipitation variables for predicting peak rate of runoff on very small (0.5- to 11-acre) watersheds are the maximum 15-minute depth of precipitation and total depth of rainfall.
4. For larger watersheds, of a few square miles in size, significant variables were rainfall volume above 0.5-inch depth and an antecedent runoff variable.

APPENDIX

Method

Using an extension of the Markov Theorem, David and Neyman [1938] showed that the variance in the predicted mean could be estimated by

$$U_F^2 = \frac{-\Delta_0 \Delta_1}{(n\text{-}s)\Delta^2}$$

and the linear regression equation could be denoted by

$$Y_{pred} = \frac{-\Delta_\theta}{\Delta}$$

where n = number of observations
s = number of independent variables plus one
and Δ, Δ_0, Δ_1 and Δ_θ are determinants involving the sum of variable crossproducts.

Notation (in order of appearance in the flow chart)

n number of observations

r number of independent variables

s r + 1

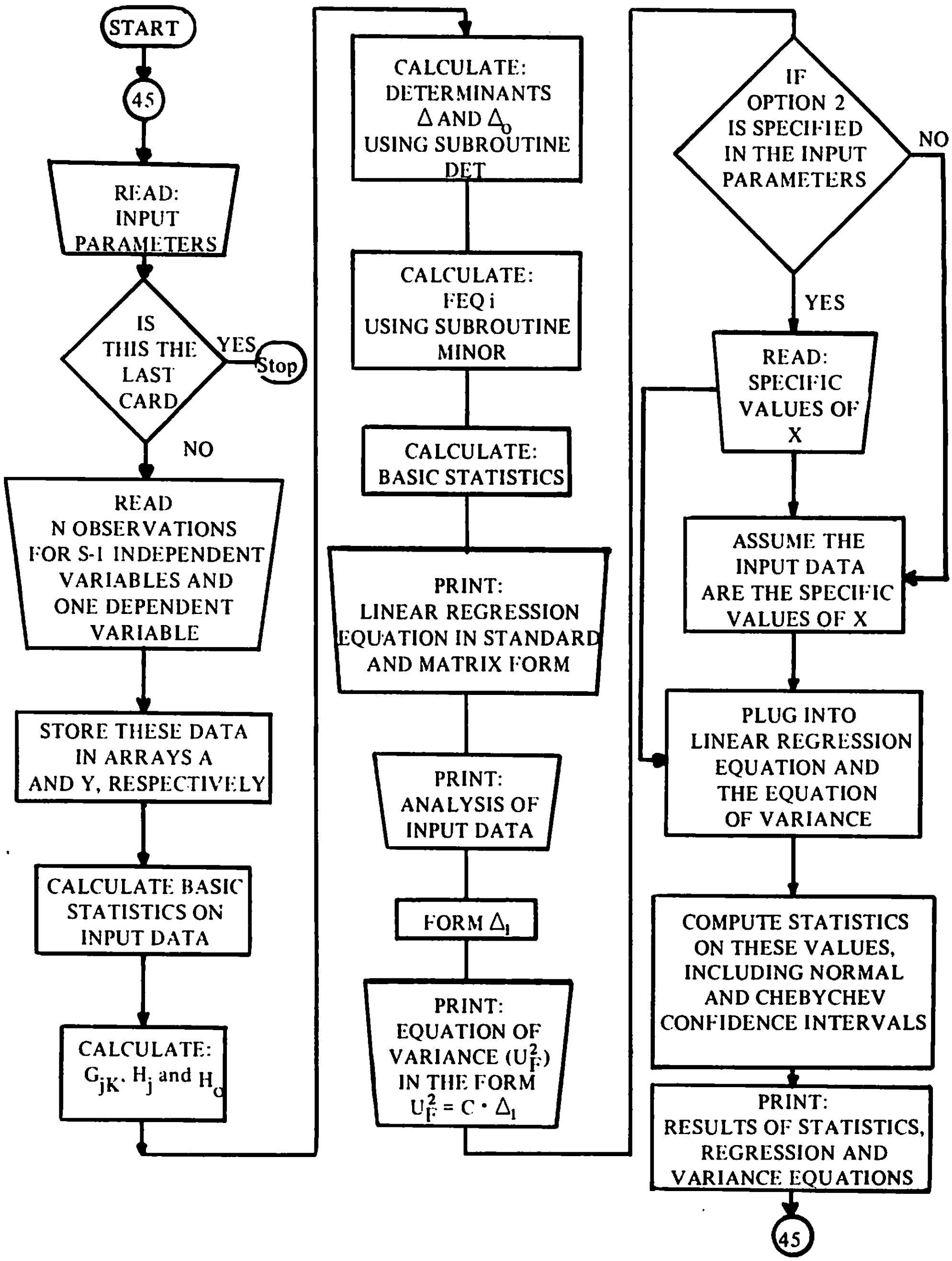
A n by s matrix containing the observations of the independent variables. For every member a_{ij} of matrix A, the i and j indicates observation and variable respectively and j goes from 2 to s.
$a_{i1} = 1$ for all i

Y 1 by n matrix containing the dependent variable observations. For a member y_i of Y, the i indicates the observation.

G s by s matrix containing the sums of the crossproducts of the independent variables. For a member g_{jk} of matrix G,

$$g_{jk} = \sum_{i=1}^{n} a_{ij} \cdot a_{ik}$$

H 1 by s matrix containing the sums of the crossproducts between the dependent and independent variables. For every member h_j of maxtrix H, Flow Chart of Program MARKO-2



$$h_j = \sum_{i=1}^{n} a_{ij} \cdot y_i$$

also, by definition

$$h_o = \sum_{i=1}^{n} y_i^2$$

Δ determinant of $|G|$
i.e., $\Delta = G$

Δ_o s by s determinant:

$$\Delta_o = \begin{vmatrix} h_o & h_1 & \cdots\cdots & h_s \\ h_1 & g_{11} & g_{12} \cdots & g_{1s} \\ h_2 & g_{21} & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ \cdot & & & \cdot \\ h_s & g_{s1} & \cdots\cdots & g_{ss} \end{vmatrix}$$

FEQ(K) coefficients of the linear regression equation where FEQ(K) corresponds to the formal variable name b_K and

$$\mathrm{FEQ(K)} = \frac{1}{\Delta} \sum_{j=1}^{s} \Delta_{jK} H_j$$

and $b_1 = 1$

Δ_{ij} minor of the determinant Δ, i^{th} column and j^{th} row with the correct sign.

Δ_1 s by s determinant:

$$\Delta_1 = \begin{vmatrix} 0 & b_1 & b_2 & b_3 \cdots & b_s \\ b_1 & g_{11} & g_{12} & g_{13} \cdots & g_{s1} \\ b_2 & g_{21} & & & \\ b_3 & g_{31} & & & \\ & & & & \\ b_s & g_{s1} & \cdots & \cdots\cdots & g_{ss} \end{vmatrix}$$

Δ_θ s by s determinant:

$$\Delta_\theta = \begin{vmatrix} 0 & b_1 & b_2 & \cdots & b_s \\ b_1 & g_{11} & g_{12} & & g_{1s} \\ b_2 & g_{21} & & & \\ & & & & \\ b_s & g_{s1} & \cdots & \cdots & g_{ss} \end{vmatrix}$$

REFERENCE CITED

David, F. N. and J. Neyman. 1938. Extension of the Markov Theorem on least squares. Statistical Research Memoirs 2:105-116.
Draper, N. R. and H. Smith. 1966. Applied regression analysis. John Wiley & Son, 407 pp.
Osborn, H. B. and W. N. Reynolds. 1963. Convective storm patterns in the southwestern United States. Bull. IASH 8(3):71-83.
Osborn, H. B. and L. Lane. 1969. Precipitation-runoff relationships for very small semiarid rangeland watersheds. Water Resources Research 5(2):419-425.
Renard, K. G. 1970. The hydrology of semiarid rangeland watersheds. USDA Pub. ARS 41-162.
Sellers, W. D. ed. 1960. Arizona climate. Tucson, Ariz.: University of Arizona Press.
USDA-SCS. 1970. Report of research needs for soil and water conservation. U.S. Dept. of Agr., SCS.