USE OF THE ELECTRONIC COMPUTER FOR STREAMFLOW ANALYSIS

by

Lynn F. Johnson [1]

The application of statistics to streamflow analysis has been limited by the lack of sufficient data and the number of calculations necessary to make a complete statistical analysis. The restriction caused by lack of data has been removed in many cases for several years. It has been only recently with the wide spread use of electronic digital computers that the time consuming task of calculations could be overcome.

With the installation of computers at most of the state colleges and universities, and many additional data processing centers, it is possible now to make arrangements to do, or have a statistical analysis done, on an electronic computer. This makes it possible for us to eliminate the time consuming task of the mathematical calculations.

Computers have been subjected to publicity which has led to their being called electronic brains, robots, etc. This publicity coupled with their known complicated structure has made potential users hesitate using them. The user needs to know little more of their structure than which button to press to make it go.

The computer is not a "brain", but rather a remarkably fast and phenomenally accurate "moron". It will do exactly what you tell it to do. The computer (computer, as referred to in this paper, is an IBM 650, although most remarks would be applicable to all digital computers) can perform the basic operations of addition and as a result of this capability can perform the remaining arithmetic operations of subtraction, multiplication, and division. In addition to these it can tell if a number is zero or not, negative or positive, and it can also tell if any digit in a ten digit number is an 8 or a 9.

Let us compare a computer and a desk calculator. When using a desk calculator an auxiliary piece of paper is used to store initial data and intermediate results; the paper or the operator's memory is used to keep track of the sequence of steps performed. On a computer, both data and the sequence of instructions (program) are stored in the storage or memory of the machine, although not necessarily the same storage. In a "stored program" machine, both data and instructions are stored in the same storage, and a given memory location can be used to store instructions or data making it possible to have the machine generate its own instructions. The main difference between a computer and desk calculator is that once the computer is started it will follow a list of instructions contained in its memory with no further guidance. Another advantage of a computer is the speed of calculations. As an examplè, the IBM 650, a medium size, medium speed computer, can add two 10 digit numbers in 1/125,000 of a minute. It can multiply two 10 digit numbers obtaining a 20 digit product in 1/6000 of a minute. It can extract the square root of a 10 digit number accurate to 9 or 10 significant digits in about 1/9000 minute. In doing this operation the IBM 650 does about 60 program steps. At Montana State College there is an IBM 650 electronic computer available, therefore, remarks refer specifically to it. However, there are many electronic computers available that will perform the same calculations with equal speed.

Problem

It has been found through previous investigations that a reliable formula for forecasting seasonal volume runoff can be obtained by multiple regression analysis. The factors that have been found to have the greatest effect on the runoff are snow water equivalent (water content), temperature, precipitation, base flow, and soil moisture. There are usually several snow courses on a watershed, or nearby, which may be correlated with expected runoff. The water content at each of these courses is usually measured more than once each season. Weather data are usually available for one or more weather stations within the basin. Thus, the numbers of independent variables to consider for each stream becomes quite large. In most cases the number of variables exceeds the number of observations. Because of the relatively short period of record the number of independent variables finally selected should not exceed three or four if the formula is to have any statistical significance. To investigate all of these independent variables and select the best three or four becomes a sizeable, if not impossible, job on a desk calculator. A project of this size and scope requires the use of modern, high-speed electronic computers if a complete, accurate and economic investigation is to be conducted.

[1] Assistant in Agricultural Engineering, Montana Agricultural Experiment Station, Montana State College, Bozeman, Montana.

Procedure

A "multiple regression analysis" program (IBM file #6.0.001) written for the IBM 650 by Arthur Coehn, formerly of the IBM Washington Data Processing Center, was selected because of flexibility and the number of variables it could accommodate.

This program is divided into four phases:

Phase I. Logarithmic transformation of initial variables and/or creation of new variables as X_i X_j.

Phase II. Calculation of means, standard deviations, and simple correlation coefficients.

Phase III. (a) Inversion of the correlation matrix.
(b) Calculation of partial correlation coefficients (with respect to the dependent variable) and multiple regression coefficients.

Phase IV. Predicting based on the regression equation or calculating the residual between observed and computed dependent variable values.

The dependent variables under investigation, such as April-September runoff, are tabulated with all the independent variables to be examined. These data are prepared for use in the "multiple regression analysis" program which requires that the data contain no more than five digits. The numbers must be in the form XX.XXX and the total number of variables must not exceed 33. These data are then punched on standard data cards in the required format and processed using Phase II of the program. No transformation of data is necessary so analysis starts on Phase II of the program.

The results from Phase II are the mean, standard deviation and all possible simple correlation coefficients according to the following formulae:

$$\bar{X}_i = \frac{\sum x_i}{N} \quad \text{where N is number of observations} \text{ - - - - - - } (1)$$

$$\sigma i = (V_{ij})^{\frac{1}{2}} \quad \text{where } V_{ij} = \frac{1}{N}\left[\sum X_i X_j - \bar{X}_i \sum X_j\right] \text{ - - - - - } (2)$$

$$r_{ij} = \frac{V_{ij}}{\sigma i \quad \sigma j} \text{ - - - - - } (3)$$

The number of simple correlation coefficients using 33 variables is 1089; however, one-half of these are redundant since the program is designed to compute r_{ij} as well as r_{ji} .

It takes approximately 20 minutes to run Phase II with 33 variables and 20 years of records. A run with 20 variables and 20 years of record reduces the time to approximately 15 minutes.

To illustrate the procedure, the Madison River near West Yellowstone, Montana will be used. There were seven dependent variables and twenty-six independent variables chosen on this river. These are tabulated in Plate 1 scaled ready for key punching.

After these data are run in Phase II of the program those independent variables which seem the most pertinent (having the highest correlation coefficient) to each of the dependent variables are retained, the rest are eliminated. The number of the independent variables chosen is generally restricted to nine or less. In no case is more than one measurement of a snow course chosen, and in all cases, only that data which will be available at the time the forecast is made is considered. The correlation coefficients for the Madison River are tabulated in Plate 2. Those marked with an asterisk (*) indicate the variables retained for each dependent variable.

Each dependent variable is retabulated with the independent variables chosen and run through Phase II and Phase III (a) and (b) obtaining the partial correlation coefficients involving the dependent variable, X_1, and the regression coefficients according to the formulae:

MADISON RIVER NEAR WEST YELLOWSTONE, MONTANA

ORIGINAL DATA

YEAR	#	SEASONAL RUNOFF[1]				MAX[2] FLOW	NO. DAYS EXCEED 650 cfs	NO. DAYS RECEED 650-400	HEBGEN DAM W.C.		W. YELLOWSTONE TEMP.			W. YELLOWSTONE W.C.		VALLEY VIEW W.C.		BIG SPRINGS W.C.		ISLAND PARK W.C.		THUMB DIVIDE	LUPINE CREEK W.C.		CREVICE MTN. W.C.		DEVIL'S SLIDE W.C.			CANYON W.C.	LOW FLOW DATA[3]		W. YELLOWSTONE PRECIP.	
		APR-JUL	APR-SEP	MAY-JUL	MAY-SEP				MAR	APR	MAR	APR	MAY	MAR	APR	MAR	APR	MAR	APR	MAR	APR	APR	MAR	APR	MAR	APR	MAR	APR	MAY	APR	LOW DAY	LOW 5 DA	OCT	NOV
1940	1	10.71	14.22	8.34	11.87	9.4	13	3	10.7	11.0	26.5	35.9	47.2	7.5	8.7	7.8	10.1	11.9	14.0	10.9	11.7	14.5	5.3	6.6	7.4	9.6	13.5	18.8	23.8	10.3	30.2	30.5	1.33	0.01
41	2	10.79	14.74	8.59	12.55	9.0	13	4	8.6	7.3	28.2	36.2	47.0	7.8	7.1	5.6	9.3	13.0	14.2	10.3	9.6	11.2	2.2	2.7	5.1	5.6	9.4	12.7	15.2	5.6	29.0	29.2	1.10	2.85
42	3	12.84	16.62	10.07	13.86	11.6	24	7	8.2	8.6	21.8	37.0	40.6	9.0	8.8	7.1	10.9	15.4	16.1	12.5	13.1	14.5	6.0	6.9	5.5	6.7	16.0	21.1	20.9	6.8	30.2	30.8	1.63	2.20
43	4	21.63	27.26	17.92	23.55	20.0	83	25	13.7	14.6	17.7	38.0	41.4	18.4	18.7	18.4	21.1	29.6	30.0	23.3	20.6	30.8	16.0	15.6	14.2	14.6	17.9	20.2	24.9	18.7	31.8	31.9	1.45	4.72
44	5	13.62	18.19	11.02	15.59	10.0	36	13	6.5	7.8	16.6	34.4	44.2	6.4	7.3	6.7	10.5	12.0	15.2	8.3	11.7	18.1	3.6	4.9	4.7	9.2	11.8	16.7	18.7	8.5	42.8	42.5	2.97	0.93
45	6	14.11	18.95	11.75	16.60	10.2	45	5	7.9	8.2	26.0	29.0	43.8	6.9	7.6	6.1	10.0	13.3	16.4	10.5	11.3	15.4	3.9	5.2	5.9	7.3	11.1	15.3	18.6	8.2	36.5	36.5	0.78	2.88
46	7	15.21	19.93	11.64	16.36	10.6	57	10	13.0	15.3	28.0	38.2	41.4	10.9	13.0	12.4	15.5	18.7	23.2	14.4	17.1	24.3	8.3	12.2	7.7	9.4	18.6	24.3	20.2	14.4	37.7	38.0	0.61	1.95
47	8	16.70	21.92	14.04	19.28	13.1	64	12	11.5	11.0	25.4	33.0	46.0	10.6	10.8	13.4	14.3	17.5	20.7	13.9	19.1	24.7	8.1	9.9	8.8	11.2	16.4	22.7	28.0	15.9	39.6	41.1	5.04	2.82
48	9	14.20	18.82	11.48	16.11	12.7	27	17	10.1	11.7	18.4	36.0	45.4	7.2	9.2	8.8	12.8	13.2	18.1	10.2	13.8	20.8	7.5	10.2	11.2	13.8	23.9	26.6	31.6	12.7	39.1	39.2	0.72	4.00
49	10	15.38	19.90	12.17	16.68	12.6	48	10	14.0	14.6	25.9	37.2	46.1	15.0	18.0	17.3	18.8	26.4	27.8	21.8	22.5	28.8	13.3	14.2	11.4	13.0	18.8	21.4	20.4	18.3	35.7	36.3	1.04	1.97
50	11	17.31	22.78	14.48	19.95	14.3	55	14	9.6	14.0	21.9	31.7	39.3	11.8	14.5	15.3	19.5	20.4	25.9	17.3	21.6	31.4	9.2	11.7	7.5	9.4	12.8	19.6	25.4	18.5	37.1	37.9	1.64	0.95
51	12	17.69	23.45	14.77	20.53	15.0	58	30	9.1	10.9	19.3	33.4	45.2	11.8	14.0	10.3	14.9	18.8	22.4	12.9	17.4	27.8	7.1	10.6	7.0	9.5	15.7	20.5	19.7	18.9	43.0	44.2	2.88	1.79
52	13	19.18	24.85	15.73	21.40	13.1	65	25	16.5	19.6	18.9	36.6	44.5	16.5	19.0	23.8	30.0	31.0	34.7	26.4	30.4	29.6	13.7	16.5	11.2	14.5	21.0	28.4	28.7	21.4	42.0	44.4	3.66	1.66
53	14	15.77	20.69	12.99	17.91	15.0	39	12	9.3	12.2	24.7	29.2	38.8	9.6	10.8	12.9	15.6	18.9	21.7	15.1	16.2	21.1	8.7	9.3	7.9	9.2	16.2	19.0	23.3	14.3	39.0	40.6	0.11	0.58
54	15	16.78	21.92	13.82	18.97	13.5	56	18	10.1	12.0	20.0	35.5	45.1	10.3	12.1	12.4	18.4	20.8	24.7	15.5	16.8	28.0	10.5	13.1	7.4	9.8	14.4	20.6	20.0	17.8	37.8	38.4	0.41	1.07
55	16	13.64	18.31	11.28	15.95	10.4	39	10	8.3	9.5	14.3	30.2	42.5	7.7	10.7	8.3	13.2	16.1	21.9	12.4	16.2	20.8	9.2	12.2	5.2	8.7	13.3	17.8	22.3	14.1	37.0	37.7	0.42	0.99
56	17	20.01	25.54	16.80	22.34	19.7	65	22	13.5	13.4	21.5	35.4	47.9	14.8	14.1	17.1	15.9	23.8	22.7	19.8	17.0	35.5	13.4	14.0	7.9	11.9	18.6	24.0	29.2	20.9	32.1	35.9	2.00	4.11
57	18	16.78	21.99	14.09	19.31	13.2	55	10	12.3	14.3	23.8	33.1	46.0	11.8	14.2	13.6	18.1	22.0	26.6	16.4	19.4	25.3	9.2	11.4	5.3	7.9	14.2	19.1	24.6	17.1	40.0	43.0	2.82	0.95
58	19	12.82	17.13	10.21	14.52	10.9	29	11	9.1	11.8	23.2	32.8	50.8	6.5	7.6	9.2	12.5	14.4	18.0	11.8	14.4	17.6	5.8	7.0	5.3	6.1	17.2	21.4	26.6	13.2	39.8	40.4	2.11	1.00

[1] 10,000'S AC FT

[2] 100'S CFS

[3] 10'S CFS

PLATE I

MADISON RIVER NEAR WEST YELLOWSTONE, MONTANA

SIMPLE CORRELATION COEFFICIENTS

	SEASONAL RUNOFF				MAX FLOW	NO. DAYS EXCEED 650 cfs	NO. DAYS RECEED 650-400	HEBGEN DAM W.C.		W. YELLOWSTONE TEMP.			W. YELLOWSTONE W.C.		VALLEY VIEW W.C.		BIG SPRINGS W.C.		ISLAND PARK W.C.		THUMB DIVIDE	LUPINE CREEK W.C.		CREVICE MTN. W.C.		DEVIL'S SLIDE W.C.			CANYON W.C.	LOW FLOW DATA		W. YELLOWSTONE PRECIP	
	APR-JUL	APR-SEP	MAY-JUL	MAY-SEP				MAR	APR	MAR	APR	MAY	MAR	APR	MAR	APR	MAR	APR	MAR	APR	APR	MAR	APR	MAR	APR	MAR	APR	MAY	APR	LOW DAY	LOW 5 DA	OCT	NOV
V	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33
1	1.000																																
2	.997	1.000																															
3	.985	.996	1.000																														
4	.989	.996	.997	1.000																													
5	.885	.866	.886	.863	1.000	*	*																										
6	.951	.953	.938	.940	.746	1.000																											
7	.814	.823	.807	.816	.734	.716	1.000																										
8	.614	.591	.567	.549	.434	.590	.429	1.000																									
9	.664	.652	.616	.610	.463	.649	.494	.903	1.000																								
10	-.381*	-.391*	-.401*	-.407*	-.337*	-.311*	-.602*	.035	-.056	1.000																							
11	.150	.101	.074	.035	.156	.131	.214	.506	.332	.058	1.000																						
12	-.214	-.212	-.200	-.199	-.167	-.235	-.036	.101	-.087	.189	-.153	1.000																					
13	.854	.830	.821	.797	.755*	.813	.636	.820	.769	-.151	.449	-.154	1.000																				
14	.862*	.851*	.831*	.822*	.687	.844*	.687*	.813	.858	-.274	.338	-.234	.949	1.000																			
15	.821*	.806*	.790*	.777*	.655*	.769*	.601	.891	.913	-.152	.292	-.110	.907	.918	1.000																		
16	.756	.752	.727	.726	.513	.721	.610	.840	.914	-.264	.233	-.199	.834	.916	.946	1.000																	
17	.841*	.822*	.808*	.791*	.688*	.814*	.620*	.840	.835	-.215	.349	-.157	.962	.953	.948	.915	1.000																
18	.809	.806	.782	.781	.571	.817	.616	.794	.880	-.293	.211	-.222	.864	.950	.924	.958	.949	1.000															
19	.777*	.752*	.741*	.718*	.645*	.732*	.545	.870	.849	-.160	.355	-.144	.942	.918	.961	.914	.978	.910	1.000														
20	.708	.706	.677	.679	.459	.693	.582	.807	.903	-.275	.198	-.190	.808	.909	.920	.979	.898	.958	.896	1.000													
21	.938*	.928*	.921*	.911*	.861*	.901*	.749*	.656	.726	-.331	.243	-.150	.881	.897	.846	.750	.859	.835	.815	.742	1.000												
22	.851*	.804*	.796*	.771*	.750*	.786*	.600	.793	.809	-.321	.349	-.169	.897	.904	.898	.813	.926	.887	.910	.809	.908	1.000											
23	.813	.800	.780	.770	.661	.801	.663*	.774	.838	-.397	.291	-.176	.830	.910	.873	.836	.880	.912	.850	.856	.898	.960	1.000										
24	.603	.574	.560	.534	.574	.524	.516	.695	.641	-.169	.493	-.135	.720	.681	.700	.632	.680	.622	.706	.582	.661	.737	.680	1.000									
25	.670	.649	.635	.616	.593	.575	.628	.723	.668	-.277	.480	-.052	.709	.692	.741	.664	.671	.625	.686	.616	.894	.734	.712	.923	1.000								
26	.394	.367	.339	.319	.402	.281	.471	.587	.597	-.224	.456	.024	.420	.429	.512	.467	.435	.419	.457	.464	.423	.558	.585	.664	.682	1.000							
27	.436	.422	.390	.383	.342	.344	.515	.641	.680	-.240	.390	.064	.402	.468	.573	.548	.432	.474	.460	.546	.462	.540	.630	.583	.684	.932	1.000						
28	.358	.351	.365	.355	.398	.206	.320	.397	.474	-.282	.009	.127	.213	.281	.427	.377	.236	.295	.282	.339	.382	.405	.426	.474	.528	.670	.725	1.000					
29	.851*	.855*	.841*	.844*	.716*	.813	.743*	.703	.807	-.311	.101	.001	.788	.870	.888	.808	.826	.864	.788	.828	.920	.856	.910	.556	.637	.454	.554	.456	1.000				
30	.144	.172	.212	.229	.141	.010	.304	-.154	-.111	-.474	-.460	.214	-.046	-.003	.066	.136	.044	.092	.015	.144	.059	-.030	-.012	-.017	.098	-.074	-.046	.263	.180	1.000			
31	.200	.228	.270	.285	.197	.056	.336	-.091	-.054	-.465	-.457	.229	.009	.050	.129	.186	.100	.140	.070	.192	.091	.025	.040	-.015	.116	-.045	-.007	.298	.243	.993	1.000		
32	.290*	.303*	.313*	.322*	.110	.282	.305*	.273	.167	-.051	-.060	.318	.230	.205	.290	.282	.182	.177	.158	.253	.158	.038	.068	.026	.204	.080	.232	.299	.254	.386	.419	1.000	
33	.419*	.399*	.410*	.389*	.614*	.344*	.340*	.250	.037	-.119	-.388	.055	.384	.214	.188	.071	.254	.101	.244	-.016	.362	.305	.206	.526	.463	.394	.267	.222	.102	-.149	-.140	.039	1.000

PLATE II

$$r_{1j\cdot}\ (1J)^1 = \frac{-a_{1j}}{(a_{11}\ \ a_{jj})^{\frac{1}{2}}} \quad - - - - - - - - - - - - - - (4)$$

$$b_j = \frac{-(a_{jj})\ (\sigma_1)}{(a_{11})\ \ (\sigma_j)} \quad - - - - - - - - - - - - - - (5)$$

where a_{ij} are the inverse elements.

σ_j are the standard deviations from Phase II.

A test can now be made to measure the significance of the formula developed thus far. The "F" test will provide this measure. This test could be delayed until the final formula is obtained. However, unnecessary work can be eliminated if work is stopped at this point upon finding no significance. Column one in Table 1 gives the results of this significance test for the Madison River near West Yellowstone, Montana. It will be noted that all formulae are significant at the 5 percent level and all except the peak flow are significant at the 2.5 percent level. Because of the nature of the data, the short period of record and large number of variables we can expect this significance test to improve as variables are eliminated.

TABLE 1

Resultsof Significance Tests for Madison River Analysis#

	F* initial	F reduction		F final		R_2
			table		table	
Dependent Variable	calculated	calculated	25% pt.	calculated	0.5%	final
1. April-July Runoff	13.96	1.30	1.62	26.90	6.00	.885
2. April-Sept. Runoff	12.39	1.47	1.62	22.32	6.00	.864
3. May-July Runoff	11.19	1.48	1.62	19.91	6.00	.850
4. May-Sept. Runoff	10.08	1.63	1.62	16.87	6.00	.828
5. Peak Flow	3.64	0.12	1.63	9.87	5.79	.792
6. No. days flow exceed 650 cfs	5.84	0.52	1.63	11.83	5.79	.820
7. No. days flow to receed 650 to 400 cfs	4.98	1.01	1.63	8.15	5.79	.758

*Values of F for comparison in itial test. n = 19 K = 9

$F_{10\%} = 2.44$ $F_{5\%} = 3.18$ $F_{2.5\%} = 4.03$ $F_{0.5\%} = 6.54$

#In this analysis there are 19 observations. Nine variables were included in the initial step. This number was reduced to 4 for dependent variables 1 through 4 and to 5 for dependent variables 5 through 7.

All values of F taken from Snedecor, Statistical Methods, 5th Edition, Collegiate Press, Iowa State College, Ames.

A further reduction in independent variables is accomplished by eliminating those that explain the least amount of the variation of the dependent variable. Because of the interaction that exists between variables these variables must be eliminated after considering the overall effect they have upon each other, and finally the dependent variable. The relative size of the partial correlation coefficients obtained in Phase III is a measure of each variables contribution to the total variation taking all variables into account. The variables that have the largest partial correlation coefficient are retained as the final selection of variables. Negative coefficients are not retained unless there is a physical reason for their being negative. The partial correlation coefficients for the Madison River near West Yellowstone, Montana are listed in Table 2; those marked with an asterisk (*) are the ones retained.

TABLE 2

Partial Correlation Coefficients Madison River near West Yellowstone

Independent Variables	Apr-July Runoff	Apr-Sept. Runoff	May-July Runoff	May-Sept. Runoff	Peak Flow	Days flow exceeds 650 cfs	Days flow receed 650 400 cfs
West Yellowstone - Oct.-Nov. P.#	.034	.031	.034	.031	—	—	.354*
Canyon April W.C.	-.104	.008	-.006	.081	-.007	—	.051
West Yellowstone - April W.C.	-.127	-.111	-.194	-.162	—	.038	.343*
Valley View March W.C.	.108*	.089*	.068*	.056*	-.001	-.016	—
Lupine Creek - Mar. W.C.	-.412	-.491	-.459	-.520	-.001	-.036	—
Thumb Divide - Apr. W. C.	.692*	.671*	.668*	.647*	.407*	.601*	-.191
Big Springs Mar. W.C.	.315*	.319*	.340*	.337*	-.121	.329*	-.017
West Yellowstone - Mar. Temp.	-.040	-.042	-.043	-.040	-.143*	-.120*	-.640*
West Yellowstone Nov. P.	—	—	—	—	.227*	.111*	—
West Yellowstone Mar. W.C.	—	—	—	—	.130*	—	—
Peak Flow	—	—	—	—	—	.198	.329*
Lupine Creek - Apr. W. C.	—	—	—	—	—	—	-.028

#The sum of the Oct. and Nov. precip. was used in this case because each correlated equally with the dependent variable.

The final regression coefficients are now calculated using only those variables retained. Omit cards to be inserted in Phase III (a), (b) make it possible to compute these new coefficients using the results already obtained in Phase II. A test should be performed to see if those variables eliminated in the last step made a significant difference in the results. Column 2, Table 1 gives the results of this test for the Madison River. The May-September runoff is the only case where the reduction had any significance and in this case it was only at the 25% level.

The final step is to compute the multiple correlation coefficient (R^2) and the intercept (A). The multiple correlation coefficient is a measure of the explained variation of the dependent variable, and A gives the remaining value for the regression equation. These can be computed on the computer using the results obtained up to this point and a program written for this purpose. The F test is computed for the final equation to test the significance. The R^2 and F for the final equations for the Madison River are given in Table 1. It will be noted that each equation is highly significant.

Results

This method of analysis and an electronic digital computer makes it possible to examine all of those variables believed to have an effect on the runoff. In the example, seven streamflow characteristics (dependent variables) and twenty-six independent variables were used. The independent variables included water content at several snow courses measured at different times, precipitation, temperature, and low flow data. The number of independent variables was reduced from 26 for each dependent variable to 3 or 4 and an equation developed in each case which was highly significant.

The multiple regression equation derived by this anslysis is of this form:

$$\hat{Y} = b_1X_1 + \ldots + b_n X_n + A \text{ -} (6)$$

where $\hat{Y}$ is the estimated value

b is the regression coefficients computed in the final run of Phase III (b)

X is the values of the independent variables

A is the intercept value

An equation of this type can be derived for each runoff period or streamflow characteristic desired.

Other additional information may be derived from this analysis. It is possible to get a relationship between any two of the variables using the results of the first run of Phase II. The simple regression coefficients can be derived by the following formula:

$$b = \frac{(r_{xy})\sigma y}{\sigma x} \quad \text{- -} \quad (7)$$

where b is the regression coefficient

r_{xy} is the simple correlation coefficient between the two variables

σy is the standard deviation of the unknown variable to be estimated

σx is the standard deviation of the known variable

Discussion

The analysis of the river is only part of the streamflow forecasters problem; periodically forecasts must be made which mean solving all of the equations thus developed. These equations all have the same general form so the computer can be easily used to compute the forecasts. A program has been written for the IBM 650 to solve any equation or series of equations of this form. This program is perfectly general and will handle up to 6 variables.

In addition, this program will compute the percent average the forecast is of a base period. The program is designed to route streamflow as well as compute headwater station forecasts. One card is punched for each equation. Each card contains the regression coefficients, location of the independent variables, the intercept, the average flow for a base period, and identification number. Any equation card may be inserted or withdrawn without affecting the operation of the program. Current data of the independent variables are used as input. The output gives the identification number, forecast and percent average for each period. The average flow for each period is also indicated. The program is designed with internal checks to insure accurate computations. The total computation time in seconds can be computed by this formula, $T = 70 + \frac{N}{2.5}$, where N is number of equations.

Summary

The computer is extremely valuable in an investigation of this type. Many variables can be examined which were not previously considered because of physical limitations. For example, when selecting independent variables snow courses within the drainage basin, as well as those located in adjacent basins, can be examined.

An analysis of a river or stream requires one to two hours computer time, and when completed, one can be assured of accurate results. To attempt this investigation with a desk calculator would require several weeks time, if such a project would even be considered. The time required to compute the monthly forecasts can likewise be reduced to a small portion of the time required using a desk calculator.