BIBLIOGRAPHIC DATA SHEET

1. CONTROL NUMBER: PN-AAJ-963

2. SUBJECT CLASSIFICATION (695): TA00-0000-G750

3. TITLE AND SUBTITLE (240)

A report on initial analysis of the USAID/Mahidol provincial data bank of Thailand

4. PERSONAL AUTHORS (100)

Johnson, T.E.

5. CORPORATE AUTHORS (101)

6. DOCUMENT DATE (110): 1980

7. NUMBER OF PAGES (120): 31p.

8. ARC NUMBER (170)

9. REFERENCE ORGANIZATION (130)

10. SUPPLEMENTARY NOTES (500)

11. ABSTRACT (950)

12. DESCRIPTORS (920)

Thailand
Data
Data Storage
Data Processing
Computer Programs
Statistical analysis

13. PROJECT NUMBER (150)

14. CONTRACT NO.(140): AID-493-0013-T

15. CONTRACT TYPE (140)

16. TYPE OF DOCUMENT (160)

AID 590-7 (10-79)

PN-AAJ-963

A Report on Initial Analysis of the USAID/Mahidol Provincial Data Bank of Thailand

for

U.S. Agency for International Development

Contract No. AID-493-0013-T

by

Thomas Eric Johnson, PhD

August 1980

INTRODUCTION

1. This report covers work done under U.S. Agency for International Development Agency Contract No. AID-493-0013-T and its extensions and amendments. The work performed was the analysis of the data bank file of provinces of Thailand, called the USAID/Mahidol Provincial Data Bank. The object of the work was to review and analyze the data collected in this data bank. Assistance in preparation of the Northeastern Small Scale Irrigation Project mentioned in this contract has already been provided to USAID/Thailand. Although the analysis of the Provincial Data Bank is an important task, the majority of the time allocated in the contract was spent on the analysis of the social survey done for the Northeastern Small Scale Irrigation Project. Therefore, only limited time was available for analysis of the Data Bank.

2. Additional data was added to the Data Bank to increase the physical data on rainfall and mean number of rainy days. Other physical data was not added due to time constraints. However, at this stage in the analysis and utilization of the Provincial Data Bank full comprehensiveness is not necessary since the number of variables already exceeds two hundred. The chief concern at the present time is to analyze this data. This report presents the results of the use of two types of computer - assisted and analytical techniques:

correlation analysis and regression analysis. Both of these techniques are explained below. The data in the analysis existed in an SPSS system file and the Consultant assumes that all formatting and cleaning of the data in the Provincial Data Bank was correct and adequate.

3. This report is divided into three sections. The first section presents the results of the correlation analysis of the data. The second section presents the analysis based upon applications of a step-wise regression program in order to combine noncontiguous provinces into homogeneous areal types. The last section makes suggestions and recommendations in the use of the Data Bank.

CORRELATION ANALYSIS

4. With large sets of data such as the Provincial Data Bank, some method of reducing the number of variables and their variations to dimensions which are more readily understood is necessary. The Consultant was requested by USAID to use correlation analysis as one method of analysis. The SPSS subprogram Partial Correlation was used for this analysis (Statistical Package for the Social Sciences, 1975:301). This subprogram provides the researcher with a measure of association between two variables while controlling for the effects of other variables. (Linear effects are removed statistically). This type of analysis is suitable for data where the number of cases is relatively small as in the case of the Provincial Data Bank. Seventy observations were used. Since this subprogram also provides simple or zero-order correlation coefficients, it is possible to observe the effect of the control variable on the simple correlations of variables in which the researcher is interested. Partial Correlation analysis may be used in three situations: 1) when the researcher thinks a relationship is unlikely; 2) when some causal change or time series exists; and 3) when a relationship which should exist does not exist. The first situation is an example of a spurious correlation. If the original association between two variables is reduced by the introduction of the control

variable, then the relationship is spurious and due to the association of the two variables with the control rather than with each other. In the second situation, the introduction of the control demonstrated that the control is intervening between the relationship of the original variables. For example, a researcher might find that children of working mothers are more likely to be delinquent than those children whose mothers do not work. By introducing a measure of supervision as a control, the researcher might find that the delinquency rates are unaffected by whether the child's mother is working or not. In this case, "supervision" intervenes between working or not working and delinquency rates. In the third situation, the researcher finds that the relationship between two variables he thought were highly related is low. In this case, the relationship might be masked by some other variable, such as income. For example, it might be assumed that the need for a second car determines the purchase of a second car. However, this relationship might be masked by the level of income of the family. If the researcher controlled for income, than the relationship between need and purchase of a second car might appear.

5. The analysis of this section is twofold: 1) To investigate the relationships between variables considered important social and economic variables determining regional and subregional differences in Thailand; and 2) To inspect the

effects of regions and given groupings of provinces (changwats) on the correlations of the important variables. The consultant was given a specific list of variables in which USAID/Thailand was interested and an apriori grouping of changwats which formed six regions or types of changwats which were not necessarily contiguous. To make presentation of the material easier Table 1 lists the variables considered in the analysis, Table 2 presents the more highly correlated variables from the list in Table 1 and presents the effects of the controls, Region, Changwat (the 6 a priori groupings), on the highly correlated variables of Table 2. The original computer output was given to USAID before the Consultant left Thailand.

6. Table 1 lists the variables included in the analysis of this report. Sixteen variables relate to physical and agricultural characteristics of the provinces. Ten variables indicate social and economic characteristics generally considered to be important indicators of development and levels of service and welfare. The regional divisions are the traditional regional administrative designations of Thailand. The groupings of changwats are choices by USAID/Thailand, reflect various land form, rainfall, cropping, and geological differences.

7. Table 2 shows the various combinations of variables which had correlations higher than 0.50. In the analysis 27 variables (not including Region and Changwat) were used. This

resulted in a total of 351 pairwise correlations of the 27 variables. Out of the 351 correlations, sixty were 0.50 or higher (Table 2). Twenty-three correlations were 0.80 or higher. It should be remembered at this point that we are not talking about statistical significance since there is no sampling involved. We have analyzed data on the entire population of provinces in Thailand and all the correlations are significant since there is no sampling error involved in the selection of provinces.

8. Table 2 also compares the correlations before and after the introduction of the controls of region and changwat using partial correlation techniques. It is obvious from this table that region and changwat do not control the associations among the higher associated pairs of variables included in the analysis. All differences between correlations before and after introduction of the controls, Region and Changwat, are small. It also can be seen from the correlations matrix (original computer printout) that Region (values of the variable are 1,2,3,4(and Changwat (values of the variable are grouped 1,2,3,4,5,6 or ungrouped) are only slightly correlated with the variables selected for analysis.

9. It is worthwhile at this point to consider the correlations among the variables which were selected for analysis (27 variables plus Region and Changwat). Table 3 lists those correlations which are 0.80 or higher and the names of the

pair-wise variables. The correlation between elementary teachers per capita and maize production per capita appears spurious since there is no direct relationship to maize production and elementary teachers being employed or otherwise directly involved in maize production. Other explanations for this correlation must be sought. Areas having large numbers of in-migrants also have large numbers of out-migrants according to the next correlation. Since both out- and in-migration per capita are negatively associated with the number of public wells, it might be assumed that numbers of public wells are indicators of welfare and forces related to migration. Or they may be measures of urbanization. Or large numbers of transients move are impacted in areas of low public services. To sort out the issue the relationship between in-migration and out-migration needs to be specified, then this relationship must be related to levels of public well service. Obviously, areas having upland soils also do not have access to electricity as the correlation between upland soils 1,2,3 per capita and number of electricity users per capita indicate. Except for the last three correlations on land forms in Table 4, all the correlations between the various crops would indicate that these crops are generally farmed together and/or are generally evenly mixed in changwats which grow these crops if so many cases did not have missing values. The land forms C,D, and E are essentially

the same forms in that they occur in the same degree per capita. That is, if land form C occurs, then forms D and E occur in a fixed proportion to population size since the correlation is nearly unity.

10. A difficulty arises in this analysis because ten of twenty-six variables have large number of cases with missing values. The correlations which were computed deleted pairwise all changwats for which missing values of either variable existed. However, since a computation and some recoding of variables was done fewer than 2 cases were dropped when in fact more cases had missing values. This issue is dealt with in the following section.

Conclusions

11. Based upon the correlation analysis reported in this section the following conclusions can be made:

1) Variations among regions are unassociated with the 27 variables included in the correlation analysis.

2) A priori groupings of changwats are also unassociated with the 27 variables included in the correlation analysis.

3) The 27 variables included in the correlation analysis are intercorrelated. These variables divide into three groups: socio-economic, five correlations; cropping, sixteen correlations (missing values present); and land forms, three correlations.

4) Region is slightly more important as a control variable than changwat groupings because regions are correlated with mean rainy days and mean annual rainfall.

5) The low correlations of the 27 variables with regions and changwat groupings indicate that regional divisions need to be disaggregated, changwats need to be grouped differently, and additional variables from the 218 variables in the Provincial Data Bank need to be included in the analysis. Thus, low correlations imply regional planning for economic and social development be more decentralized, at a level lower

than the regional level. Additional types of analysis and additional types of data should also be included in the analysis of physical, social, and economic characteristics of the provinces. Some of these issues are discussed in the following section and the concluding section.

STEP-WISE REGRESSION ANALYSIS

12. The analysis of the previous section has been constrained by the numbers of variables included in the analysis and the groupings of changwats. It is possible to increase the number of variables included in the analysis by using correlation analysis. However, this would require looking at region and changwat as controls on the correlations of some 20,000 correlations. If the problem is conceptualized differently, the analysis becomes more manageable. If we consider that the variables, region and changwat groupings, are dependent variables, then we can find through regression analysis those variables which are most highly correlated with the dependent variables region and changwat. Thus, the analysis of the previous section is turned around. Instead of including those variables in which we are most interested, we now let a statistical procedure select these variables. The burden of the analysis of this section is to be able to make meaningful the relationship between the variables selected by the regression analysis and the dependent variables.

13. Other types of techniques of data reduction such as canonical correlation analysis and factor analysis could also be used. While popular, the use of these techniques often results in difficult to understand complex dimensions and indices. Step-wise regression analysis provides uncomplicated dimensions which are independent combinations of the

original variables. In exploratory studies of large data sets such as the Provincial Data Bank such a technique may result in the selection of variables which are difficult to explain, in the sense of making understandable the relationship of the variables selected to the dependent variable and to one another.

14. Step-wise regression is similar to the multiple regression case except independent variables are entered into the regression equation on the basis of a predetermined order either because of a supposed casual order or for the purposes of finding the best predictors of the dependent variable. Although this technique is objected to by some researchers for the purpose of searching for best predictors, it is a legitimate technique when one is doing exploratory research such as the present study. We are now asking the question what variables out of the 218 variables available in the Data Bank best predict the dependent variable, (the spatial groups) either region or changwat. We have used step-wise regression to find these independent variables and estimate the total amount of variance in the dependent variable (Region or Changwat) accounted for by the independent variables.

15. The independent variables or predictors and the various amounts of variance "explained" by the independent variables are contained in Tables 4, 5, and 6. These final three tables were arrived at by first making five separate

computer runs for separate groupings of the 218 variables. This procedure was necessitated by core storage limitations. Not all the more than 20,000 correlations could be calculated and stored at one time. From these separate runs the variables with the highest correlations with the dependent variables were entered in the final regression run which produced these three tables. In each case the eleven most important variables were included although not necessarily the same eleven variables for each dependent variable. Three dependent variables were used: region, changwats grouped, and changwats ungrouped. Eleven variables were chosen as a matter of convenience and contribution to explained variance.

16. Table 4 shows the Land Form Class D, Number of Mosques Registered, and Cassava Production explain about 60 percent of the variation among regions. The next four variables are also important indicators of regional differences. The last four variables in the list have little importance in explaining regional differences. The first seven variables explain about 78 percent (R Square) of the variation in regions. It should be noted here that different coding of the regions might have produced different results.

17. Table 5 presents a similar step-wise regression analysis for changwats ungrouped. In this case "explained" variance is low (percent) most likely due to the arbitrary coding

of the changwats. Land Form Class D is again the most important variable. None of the independent variables, however, make much individual contribution to total variance.

18. The last table, Table 6, demonstrates the effect of including more variables in the analysis than the previous section since the amount of total variance explained is about 95 percent. That is, about 95 percent of the differences between changwats as grouped into the five groups presented in Table 1 can be explained by the variables listed in Table 6. Note that only three variables are very important: Percent of Economically Active Males over the Age of 11, Median Age of Population, and B73E0146 which appears to be in the industrial classification "Services." This variable was not labeled in the initial computer runs and is not contained on the Codebook of the Administrative Areal Data Bank of Thailand, November 1, 1979.

19. If the possible coding problems with the dependent variable, region, are overlooked, then it is obvious from a comparison of Tables 4 and 6 that the areal groupings designated by the changwat groups (listed in Table 1) more adequately represent types of areal units (not contiguous) than the areal units represented by the regional units of Thailand. Nevertheless, one is faced with making sense out of the variables selected and the differences in dimensions which each set of variables is measuring with respect to region and changwat groups as two different types of areal

units. Note that the variables chosen are reasonable social, economic, and physical dimensions which could differentiate areal units in Thailand. Some complications are introduced by the different time periods which some of the variables represent.

20. A further complication exists. Since each independent variable is a continuous variable, it is a very complicated and time consuming task to identify the value of each independent variable selected which is related to the specific regional and changwat groups. Simple R and B are helpful here. For example looking at B one can see that for every increase in units of Land Form Class D, there is a 0.02 unit increase in moving from the Central Region (value 1) to the Northern Region (value 2) to the Northeastern Region (value 3) to the Southern Region (value 4). Or for the dependent variable, changwat groups, as the percent of economically active males over 11 years of age increases, the dependent variable moves from group 12 to group 6 (these were coded from 1 to 5 in the computer runs). That is to say, Group 6 has the highest percentage of economically active males over 11 years of age and group 12 has the lowest percentage. Similar relationships can be traced with the other variables. Simple R gives a measure of the strength of the relationship as a linear relationship. Land Form Class D does not have a strong linear relationship to region. Percent of Economically Active

Males over the Age of 11 years is linearly related to changwat groups.

Conclusion

21. More "explanatory power" has resulted from including all the variables in the Provincial Data Bank in the analysis of the differences in the areal units, region and changwat groups. The ability to understand the dimensions differentiating the areal units may have suffered because the relationships of the variables may have no theoretical basis.

22. The groupings of the changwats (see Table 1) are more important in distinguishing areal differences in Thailand than are regions. However, additional analysis suggested in the concluding section should be taken before this conclusion is pressed too far. One needs to vary the coding of both the dependent variables region and changwat and vary what variables are included in the regression equation before making this a final conclusion.

23. The variables selected as independent variables in the step-wise regression equations presented in Tables 5, 6 and 7 are straight forward dimensions and can be interpreted relatively easily. In the concluding section the interpretation issue is discussed further.

CONCLUSION

24. In this concluding section the Consultant makes some suggestions and recommendations on the use of the Provincial Data Bank. They are not meant to be comprehensive but suggestive of what further analysis and development of the Data Bank might be undertaken.

25. Reality always exhibits more "messiness" and complexity than the users of information and the analysist would like. In the analysis of large data banks one is always faced with complexity in analysis and difficulties in making sense out of the results of the analysis. This fact results from the large number of variables, the differences and possible biases in the different collecting agencies, the varying units of analysis often used to collect information by different agencies, and the lack of raw data from which the published data come. This same situation is faced in analyzing the information contained in the Provincial Data Bank. Never can any one analysis or approach be completely satisfactory. However, a number of steps can be taken. These are:

    1) Analysis of the Data Bank should be continued using additional analytical techniques and directed to specific problems or field oriented questions.

2) Other groupings of changwats should be found which represent different dimensions than those presented in this report and also reflect different research and field oriented questions.

3) These different groupings and dimensions should then be tested in planning and field work so that new approaches to the analysis can be found.

4) The analysis of the data in the Provincial Data Bank should be combined with the analysis of other sources representing "raw" data with consistent units of analysis such as the census and other social and economic surveys undertaken by the National Statistical office. Since USAID is interested in basic needs of the people of Thailand, the analysis of the original data produced by NSO would assist in identifying target groups as well as indicators of poverty and welfare which then could be related to the data in the Data Bank to identify locations where these specific needs exist.

26. The Consultant recommends that groupings of changwats be selected by the use of a categorical step-wise regression computer program called Automatic Interaction Detection (AID) (Sonquist and Morgan, 1964). Independent variables or predictors are utilized as categorical variables. Because the independent variables are categorical, the AID program is able to subdivide or "split" the dependent

variable into groups based upon the independent variables which contribute most significantly to explaining variation in the dependent variable. This is similar to the procedure used above but, in addition, the categorization of the dependent is also allowed to vary. In order to make this technique provide the most insights some iteration should be done using various combinations of meaningful variables once the initial selection of variables has been done by the AID program. The advantage of using this program is that "trees" or diagrams of the dividing or "splitting" of the dependent variable on the basis of the independent variables is easily traced. Interaction among variables becomes evident and the splitting process may result in assymetric divisions of the dependent variable.

27. It is recommended that provisions be made to incorporate the 1980 Census of Thailand into the Provincial Data Base. It is particularly important that the original data is available since it is then possible to trace the characteristics of people as the unit of analysis. Otherwise one is faced with correlations among variables representing only characteristics of changwats. Such data would expand the Provincial Data Bank as well as allow the analysis of original data as mentioned in paragraph 25 (4) above.

28. The Data Bank might also be expanded to include information on distributions of variables where this information is available. The numbers of males and females at various ages would be one such example.

29. To make data banks a useful tool in planning at the sub-regional level and for field work it is necessary to make a long-term and continuous commitment to their development. Continual analysis and updating is necessary to make them work. USAID has access to the Provincial Data Bank and it is recommended that analysis of these data be continued.

TABLE 1: VARIABLES CONSIDERED

REGION:	Central Northern Northeastern Southern
CHANGWAT:	Group 6 Group 7 Group 8 Group 10 (Groups 2 and 3) Group 12 (Groups 1, 4, and 5) (See listing below for names)
B77E0127	Per Capita GNP, 1972 Prices.
A76E0018	Labour Wage per Day Current Year
Var 168	Mean Rainfall
Var 169	Number of Rainy Days
Var 213	Number of Teachers in elementary School per Capita
Var 214	Number of Medical Doctors per Capita
Var 215	Number of Businesses and Industry per Capita
Var 216	In-Migrants in 1977 per Capita
Var 217	Out-Migrants in 1977 per Capita
Var 218	Number of Public Wells per Capita
Var 219	Number of Health Stations per Capita
Var 220	Number of Electricity Users per Capita
Var 221	Uplands Soils 1, 2, 3 per Capita - Missing Values
Var 222	Ricelands 1, 2, 3 Per Capita - Missing Values
Var 223	Rice Production 77-78 Per Capita
Var 224	Maize Production 77-78 Per Capita - Missing Values
Var 225	Cassava Production 77-78 per Capita - Missing Values
Var 226	Sugar Cane Production 77-78 Per Capita - Missing Values
Var 227	Mung Bean Production 77-78 Per Capita - Missing Values
Var 228	Soy Bean Production 77-78 per Capita - Missing Values
Var 229	Ground Nut Production 77-78 per Capita - Missing Values
Var 230	cotton Production 77-78 per Capita - Missing Values
Var 231	Kenof Production 77-78 per Capita - Missing Values
Var 232	Land Form Class C per Capita
Var 233	Land Form Class D per Capita
Var 234	Land Form Class E per Capita
Var 300	Percent Upland Soils 1, 2, and 3

TABLE 1 (continued)

Changwat Groups

Changwat Code	Group 6	Number in Group
42	Mac Hong Son	1
13	Chiang Mai	2
12	Chiang Rai	3
25	Nan	4
51	Lamphun	5
50	Lampang	6
39	Phrexe	7
69	Uttra dit	8
16	Tak	9
38	Phetchabun	10
70	Uthathani	11
29	Prajinburi	12
37	Phetchaburi	13
28	Prajiiabkirikan	14
52	Loei	15
02	Kanchanaburi	(16)
	Group 7	
08	Cholburi	1
47	Rayong	2
66	Nong Kheii	3
68	Udorn	4
10	Chaiyaphum	5
03	Kalosin	6
19	Nakoin Phonom	7
71	Ubon	8
01	Krabi	9
14	Frang	(10)
	Group 8	
26	Buriram	1
53	Srisaket	2
41	Mahasarakham	3
54	Sakon Nakorm	4
62	Sukhothoi	5
04	Kampaengphet	6
49	Lopburi	7
09	Choinat	8
05	Khonkaen	9
20	Khorat	10
65	Surin	11
36	Phitsonuloke	12
48	Ratchaburi	13
60	Saraburi	(14)

TABLE 1 (continued)

Changwat Groups

Changwat Code	Group 10 (2 + 3)	Number in Group
06	Chantaburi	1
15	Trat	2
46	Ranong	3
40	Phuket	4
44	Yala	5
11	Chiumphol	6
33	Phong Noa	7
64	Surat Thani	8
21	Nakhon Si Thommarat	9
34	Phattalung	10
56	Satun	11
55	Songkhla	12
30	Pattani	13
24	Narathiwat	(14)

	Group 12 (1,4,5)	
59	Sammit Sakkong	1
58	Samut Songkhrom	2
72	Bangkok	3
57	Samut Prakam	4
63	Suphanburi	5
67	Ang Thong	6
31	Aynthoya	7
17	Nakora Nayok	8
07	Chochoensoo	9
18	Nakorn Pathom	10
27	Phathurm Thoni	11
23	Nonthaburi	12
35	Phichit	13
22	Nakorn Sawan	14
61	Singburi	15
45	Roi Et	(16)

TABLE 2: HIGHLY CORRELATED VARIABLES

(Variables Listed in Table 1)

Variables		Region as Control	Zero-Order correlation	Changwat as Control
B77E0127	By Var 216	0.71	0.72	0.72
	By Var 217	0.62	0.63	0.62
VAR 169	By Var 224	0.54	0.55	0.55
	By Var 226	0.49	0.56	0.56
	By Var 227	0.55	0.57	0.57
	By Var 228	0.61	0.62	0.63
	By Var 230	0.57	0.59	0.60
VAR 213	By Var 224	0.71	0.71	0.72
	By Var 225	0.71	0.71	0.71
	By Var 226	0.72	0.72	0.73
	By Var 227	0.67	0.68	0.68
	By Var 228	0.65	0.66	0.67
	By Var 229	0.56	0.56	0.56
	By Var 230	0.66	0.66	0.68
	By Var 231	0.82	0.81	0.81
VAR 214	By Var 216	0.59	0.60	0.60
	By Var 217	0.65	0.66	0.66
VAR 216	By Var 217	0.90	0.90	0.90
	By Var 218	-0.48	-0.51	-0.51
VAR 217	By Var 218	-0.53	-0.55	-0.54
VAR 219	By Var 224	0.59	0.59	0.59
	By Var 225	0.50	0.50	0.50
	By Var 226	0.52	0.52	0.52
	By VAR 228	0.57	0.57	0.57
	By Var 230	0.57	0.57	0.56
	By Var 231	0.66	0.66	0.66
VAR 221	By Var 222	0.99+	0.99+	0.99+
	By Var 300	0.53	0.53	0.54
VAR 222	By Var 300	0.53	0.53	0.54
VAR 224	By Var 225	0.79	0.80	0.80
	By Var 226	0.85	0.85	0.85
	By Var 227	0.91	0.91	0.91
	By Var 228	0.84	0.85	0.84
	By Var 229	0.76	0.75	0.75
	By Var 230	0.87	0.88	0.88
	By Var 231	0.86	0.85	0.85

TABLE 2 (continued)

Variables		Region as Control	Zero-Order Correlation	Changwat as Control
VAR 225	By Var 226	0.78	0.80	0.80
	By Var 227	0.73	0.74	0.74
	By Var 228	0.65	0.67	0.67
	By Var 229	0.72	0.72	0.72
	By Var 230	0.71	0.72	0.73
	By Var 231	0.84	0.82	0.82
VAR 226	By Var 227	0.90	0.90	0.90
	By Var 228	0.71	0.72	0.72
	By Var 229	0.63	0.63	0.62
	By Var 230	0.76	0.77	0.77
	By Var 231	0.86	0.82	0.82
VAR 227	By Var 228	0.76	0.77	0.77
	By Var 229	0.68	0.69	0.68
	By Var 230	0.82	0.82	0.83
	By Var 231	0.82	0.81	0.81
VAR 228	By Var 229	0.76	0.76	0.76
	By Var 230	0.93	0.93	0.93
	By Var 231	0.83	0.82	0.83
VAR 229	By Var 230	0.81	0.81	0.80
	By Var 231	0.69	0.69	0.70
VAR 230	By Var 231	0.82	0.80	0.81
VAR 232	By Var 233	0.99+	0.99+	0.99+
	By Var 234	0.99+	0.99+	0.99+
VAR 233	By Var 234	0.99+	0.99+	0.99+

NOTE: Region By Mean Rainfall has coefficient of 0.48.
Region By Number of Rainy Days has coefficient of 0.42.
Grouped Changwat By Any Other Variable Are not Highly Correlated.
Region by any other variables is not highly correlated.

TABLE 3: SIXTEEN HIGHLY CORRELATED VARIABLES

Arbitrary Selection of Pairs with Correlations Above 0.80)

Variable Pairs	Correlation
Var 213 - Var 224 (Elementary Teachers/capita - Maize Production/capita)	0.81
Var 216 - Var 217 In-Migrants/capita - Out-Migrants/capita	0.90
Var 216 - Var 218 In-Migrants/capita - Number of Public Wells/capita	-0.51
Var 217 - Var 218 Out-Migrants/capita - Number of Public Wells/capita	-0.55
Var 221 - Var 220 Upland Soils 1,2,3/capita - Number of Electricity Users/capita	0.99
Var 224 - Var 225 Maize Production/capita - Cassava Production/capita	0.80
Var 224 - Var 226 Maize Production/capita - Sugar Cane Production/capita	0.85
Var 224 - Var 227 Maize Production/capita - Mung Bean Production/capita	0.91
Var 224 - Var 228 Maize Production/capita - Soy Bean Production/capita	0.85
Var 224 - Var 230 Maize Production/capita - Cotton Production/capita	0.88
Var 224 - Var 231 Maize Production/capita - Kenaf Production/capita	0.85
Var 225 - Var 226 Cassava Production/capita - Sugar Cane Production/capita	0.80
Var 225 - Var 231 Cassava Production/capita - Kenaf Production/capita	0.84
Var 226 - Var 227 Sugar Cane Production/capita - Mung Bean Production/capita	0.90
Var 226 - Var 231 Sugar Cane Production/capita - Kenaf Production/capita	0.82
Var 227 - Var 230 Mung Bean Production/capita - Cotton Production/capita	0.82

TABLE 3 (continued)

Variable Pairs	Correlation
Var 227 - Var 231 Mung Bean Production/capita - Kenaf Production/capita	0.81
Var 228 - Var 230 Soy Bean Production/capita - Cotton Production/capita	0.93
Var 228 - Var 231 Soy Bean Production/capita - Kenaf Production/capita	0.82
Var 229 - Var 230 Ground Nut Production/capita - Cotton Production/capita	0.81
Var 230 - Var 231 Cotton Production/capita - Kenaf Production/capita	0.80
Var 232 - Var 233 Land Form Class C/capita - Land Form Class D/capita	0.99+
Var 232 - Var 234 Land Form Class C/capita - Land Form Class E/capita	0.99+
Var 233 - Var 234 Land Form Class D/capita - Land Form Class E/capita	0.99+

NOTE: All the production figures refer to 1977-78.

BEST AVAILABLE DOCUMENT

TABLE 4

DEPENDENT VARIABLE.. REGION REGION

SUMMARY TABLE

VARIABLE		MULTIPLE R	R SQUARE	RSQ CHANGE	SIMPLE R	B	BETA
VAR173	=LAND FORM CLASS 0	0.52696	0.27769	0.27769	0.52696	0.1509750D-01	0.27725
C7750099	NUMBER OF MOSQUES REGISTERED	0.68439	0.46839	0.19070	0.42132	0.5163179D-02	0.34034
VAR187	CASSAVA PRODUCTION 7374 TON	0.75766	0.57405	0.10566	-0.19274	-0.8949707D-05	-0.16665
A76E0016	NUMBER OF RICEMILLS	0.79552	0.63285	0.05880	0.43054	0.3292874D-03	0.19201
B73E0156	OWNERSHIP OF DWELLINGS	0.82210	0.67585	0.04299	-0.33453	-0.1681803D-01	-1.17139
A77T0012	NUMBER OF TEACHERS IN PUBLIC SECOND SCH	0.85856	0.73713	0.06129	-0.13271	0.1537595D-02	1.36292
A77P0052	BIRTH RATE PER 1,000 POPULATION	0.88287	0.77946	0.04233	0.48546	0.5810198D-01	0.26319
A76S0010	HOUSEHOLDS WITH RADIOS	0.88428	0.78196	0.00250	0.03243	-0.4175255D-05	-0.12567
B73E0155	PUBLIC ADMINISTRATION AND DEFENCE	0.88514	0.78346	0.00151	-0.15203	-0.1162918D-02	-0.36198
A70P0115	PERCENT POPULATION AGE 65 AND OVER	0.88584	0.78471	0.00125	-0.29220	0.8535362D-02	0.06738
C7750398	NUMBER OF CHINESE SHRINES REGISTERED	0.88594	0.78489	0.00018	-0.38438	-0.2359793D-02	-0.02752
(CONSTANT)						0.3244700D-01	

Notes: Multiple R: A measure of multiple linearity between independent variables and the dependent variable. Values from 0 to 1.

R Square: Amount of variance explained. Values from 0 to 1.

RSQ Change: Change in R Square between steps, the contribution of additional variable to total variance explained.

Simple R: A measure of the linearity of the relationship between the independent and dependent variables for each independent variable separately.

B: Regression coefficient.

Beta: Standardized regression coefficient.

BEST AVAILABLE DOCUMENT

Table 5

DEPENDENT VARIABLE.. CHANGWAT CHANGWAT

SUMMARY TABLE

VARIABLE		MULTIPLE R	R SQUARE	RSQ CHANGE	SIMPLE R	B	BETA
A69T0139	NUMBER OF SCHOOLS	0.31275	0.09782	0.09782	-0.31275	-0.1032440	-1.14274
A69A0128	NUMBER OF AMPHOE	0.42192	0.17802	0.08020	0.02990	2.560891	0.47983
A70P0116	POPULATION DENSITY PER 1 KM2	0.49198	0.24204	0.06402	0.24121	0.6099397D-02	0.06856
A69T0141	NUMBER OF TEACHERS	0.52370	0.27426	0.03222	-0.25938	0.12435900-01	0.70727
A76E0017	NUMBER OF BUSINESS AND INDUSTRY	0.54309	0.29494	0.02069	-0.14890	-0.30845630-01	-0.20906
A77T0020	NUMBER OF TEACHERS IN ELEMENTARY SCHOOLS	0.55299	0.30580	0.01086	0.15972	-0.2241851	-3.20426
A77T0030	NUMBER OF PUPILS IN MUNI ELEM SCHOOLS	0.65557	0.42978	0.12397	0.20848	0.2494296	3.19792
A76S0015	NUMBER OF PUBLIC WELLS	0.66688	0.44473	0.01495	0.08233	0.2128444D-02	0.12654
A77H0056	NUMBER OF HOSPITAL BEDS	0.67496	0.45557	0.01085	0.22030	0.8972423D-02	0.27431
A77P0052	BIRTH RATE PER 1,000 POPULATION	0.67931	0.46146	0.00589	-0.12237	-0.3332663	-0.08249
C77F0090	NUMBER OF REGISTERED NOURSES	0.68057	0.46317	0.00171	-0.16120	0.91551050-03	0.05379
(CONSTANT)						45.16674	

Notes: See notes to Table 4.

BEST AVAILABLE DOCUMENT

TABLE 6

DEPENDENT VARIABLE.. CHANGWAT CHANGWAT

SUMMARY TABLE

VARIABLE		MULTIPLE R	R SQUARE	RSQ CHANGE	SIMPLE R	B	BETA
A70E0110	PERCENT ECONO ACTIVE MALES AGE 11 +	0.92472	0.85511	0.85511	-0.92472	-0.3093194D-01	-0.65577
A70P0105	MEDIAN AGE OF POPULATION	0.95812	0.91800	0.06289	-0.80400	-0.8413938D-01	-0.37201
B73E0146		0.96911	0.93918	0.02117	-0.25961	-0.8874872D-02	-0.09140
C79G0040	AREA IN KM2	0.97093	0.94271	0.00353	-0.24747	-0.9256277D-04	-0.09476
A76F0008	NUMBER OF RICE - GROWING	0.97270	0.94615	0.00344	-0.04044	0.1135472D-04	0.06271
B73E0144	LIVESTOCK	0.97314	0.94700	0.00085	-0.17761	-0.6326984D-02	-0.07153
A77T0023	NUMBER OF ELEMENTARY SCHOOLS UNDER PA	0.97365	0.94800	0.00100	-0.09266	0.5033245D-03	0.02785
A77T0027	NUMBER OF MUNICIPAL ELEM SCHOOLS	0.97386	0.94839	0.00040	0.02398	-0.1076101D-01	-0.09917
C67P0074	TOTAL POPULATION IN 1967	0.97391	0.94849	0.00010	-0.11158	0.1026833D-05	0.08828
C79B0050	NUMBER OF BANKS	0.97395	0.94857	0.00008	-0.01767	-0.7242727D-02	-0.01435
A76E0016	NUMBER OF RICEMILLS	0.97396	0.94860	0.00002	-0.02414	-0.8358075D-04	-0.01112
(CONSTANT)						42.76976	

Notes: See notes to Table 4.