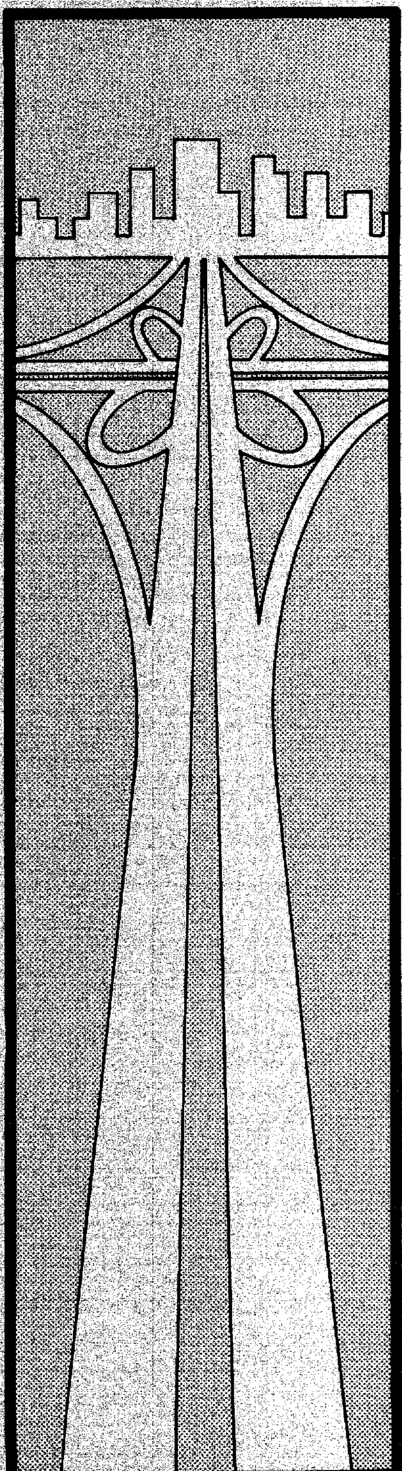
SVERCL

HIGHWAY PLANNING TECHNICAL REPORT



Number 22
March 1971

A Grouping Procedure

U.S. DEPARTMENT OF TRANSPORTATION
Federal Highway Administration

U.S. DEPARTMENT OF TRANSPORTATION
FEDERAL HIGHWAY ADMINISTRATION
WASHINGTON, D.C. 20591

A GROUPING PROCEDURE

By

Nathan Lieder, Survey Statistician
Office of Highway Planning
Program Management Division
Methods Branch

INTRODUCTION

As an aid to reliable long-term planning and to provide information needed for design, State highway departments measure various characteristics of highway usage. The measures include vehicle volume counts, vehicle classification counts, and vehicle axle weights. These do not make up the sum of the measurements. Because of the sizes of the phenomena, measurements must be samples. The values of any of these measurements may be relatively similar at many locations where samples may be taken but differ markedly from the values at other possible sampling points. The points with relatively similar measure of a characteristic form a group. To get maximum benefit from funds and efforts expended, it is important neither to oversample nor to undersample any group. But first the existence of distinct groups must be determined.

Many States use a procedure recommended by the Federal Highway Administration for grouping ATR's on the basis of similarity of monthly factor variation. This procedure cannot be applied to other types of measurement, for example grouping of truck weight stations. Realistic assignment of truck weight stations to groups might indicate that some groups were being oversampled and others undersampled insofar as weight data were concerned.

This report presents two applications of a computerized procedure that can be used for grouping ATR's, truck weight stations, or any other highway related variable for which numerical measurements are available. An earlier application of the procedure appeared in a report 1/ by Professor F. J. Wegmann and Mr. S. K. Pant, both then with West Virginia University.

The report presented a 2-step procedure for assigning truck weight stations to similar groups. The first step allocates n stations to groups by means of (n-1) repetitions of the allocation procedure. A criterion is suggested for deciding which repetition level yields optimum or near optimum groupings. The second step yields for each group a mathematical function entitled a linear discriminant function for that group. The linear discriminant functions serve two purposes. The first is to check whether the first-step groupings are optimum. If they are not, the functions indicate necessary changes in assignment. The second purpose is to provide a means for deciding to which group a new member of the population should be assigned.

In their report, Wegmann and Pant used classification count data for grouping weight stations. This report presents an application of the 2-step grouping procedure to weight data and to truck loading data obtained at 21 stations in Texas during 1969, and also to monthly values calculated by Maryland for 37 ATR stations during 1969.

GENERAL APPROACH

The first step of the procedure determines an initial set of groupings. Briefly, the procedure consists of the following.

1/ "Multivariate Statistical Techniques for Grouping Loadometer Stations and Development of a Computer Algorithm for Routing Trucks in West Virginia."

1. For each station, determine the numerical value of the characteristics which have been selected to be used in the assignment to groups. These characteristics for a truck weight station might be the average 18-kip equivalent for each of several specified vehicle types and the percent loaded of each vehicle type. For ATR's, the characteristics might be the monthly factors at each station.

2. Standardize the values. This consists of calculating the mean and standard deviation of the values for a characteristic, subtracting the mean from each value, and dividing the difference by the standard deviation. The calculation converts each measurement to dimensionless units of about the same order of magnitude. A measurement with a large number, such as ADT for example, is thus prevented from carrying excessive weight in the calculations when compared with a measurement such as the percent of the vehicles of a given type passing a station.

3. For each possible pair of stations, calculate the sum of the squared differences of the standardized values of the station characteristics. For example, assume that the average 18-kip equivalent for 3S2 combinations and the percent loaded for those vehicles were the only two characteristics being used for grouping truck weight stations and that the measurements of the two characteristics for each station had been converted to standard units. The difference between the 18-kip values for station A and station B would be squared. Then the difference for the percent loaded would be calculated and squared. The two squared values would then be added.

4. The pair with the smallest sum forms a 2-station group.

5. Calculate the weighted mean of each characteristic for those two stations and assign those averages to one station in the group. Remove the other station from further consideration but keep a record of its group assignment.

6. Repeat steps 3, 4, and 5 until all stations are assigned to a single group.

Associated with each iteration is a value of the sum of squared differences. This sum increases from iteration to iteration. A convenient point at which to accept the groups already formed and to reject any later groupings may be where the contribution of the sum of squared differences for a pair increases by a relatively large amount when compared with previous pairings. In addition, the contributions of each pairing after that point is as large or larger than the value at the decision point.

The groups determined by the suggested cut-off criterion should approximate an optimum set. However, that criterion need not be adopted. Judgment may be used to determine the point at which to reject any additional pairing.

The second step of the procedure is to compute a set of linear functions, one function for each group, using the actual values not the standardized values of the characteristics of each station in each group. These functions serve as indices for discriminating among items by classifying any item into one of several possible groups. Each linear function may be considered as a plane in multi-dimensional space which maximizes the ratio of between group variation to within group variation.

After the functions have been computed, an item is assigned to that group whose equation yields the largest result when the numerical values of the characteristics for that item are substituted in the equation. This evaluation procedure may show that one or more of the items have been incorrectly grouped in the first step. In that case, the items are reassigned, and the functions are recalculated, and item by item determination of group assignment is made. This iterative procedure is continued until no more reassignments are required. Generally no more than one or two iterations are necessary.

Calculation of the discriminant function has two purposes. The first is to check and amend if necessary assignment of the group components determined in the first step. The second purpose is to assign any new item, as a newly established truck weight station, to an existing group. This, of course, assumes that the groups already formed comprise all the possible groupings into which the population would have been classified had the entire population been under consideration.

The report by Messrs. Wegmann and Pant included a computer program to accomplish the first step of the procedure. This program was modified to print not only the contribution of each pairing to the sum of squared differences but also the components of the items in each pairing. (See tables 2 and 3.) A set of the program cards will be made available to any State desiring to try the procedure.

The multiple linear discriminant functions may be computed with the aid of three programs in IBM's Scientific Subroutine Package. These three are DMATX (Matrix), MINV (Matrix inverse), and DISCR (Discriminant functions). Since some potential users might not have these programs readily available, they have been combined as a single program. A copy of the program cards will be made available to any State desiring to try the grouping procedures. The program output consists of the following:

1. A set of equation coefficients for each group.

2. The value yielded by each equation for each item in all the groups.

3. The value of Mahalanobis D^2 statistic which may be used to test the hypothesis that the mean values are the same in all K groups.

APPLICATION OF PROCEDURE

A. Truck weight stations

Texas collected truck weight data for planning purposes at 21 locations during 1968. Five variables were used for grouping because the sample was sufficiently large at each station to provide reliable values and because the variables are important both for pavement design and for measuring highway service to the economy. These variables were readily available to test the procedure. A State may use those variables it considers most appropriate for its needs.

These were:

The average 18-kip equivalents for 2S2 vehicles at each station.
The average 18-kip equivalents for 3S2 vehicles at each station.
The percent of 2D vehicles that were loaded.
The percent of 2S2 vehicles that were loaded.
The percent of 3S2 vehicles that were loaded.

The 18-kip equivalents and the percent loaded for each vehicle type at each station were determined by running the Texas data through the FHWA weight-frequency program to produce Table W-4. Artificial highway system designations had to be assigned to get station output instead of highway system output.

Table 1 identifies the stations in the sequence they were ordered and the data for each station. The entries in columns 3 through 7 were converted to standardized values. The sum of squared differences of the standardized values were used in determining pairings.

The modified program of Messrs. Wegmann and Pant produced tables 2 and 3. The column headings are not part of the output. Each line of Table 2 is a level indicated in Table 3. Thus the data on the first line of Table 2 yield level 1 in Table 3; the data on the second line yield level 2, etc.

Table 1: - Texas' truck characteristics data classified by station sequence, station number, and highway system

Station sequence	Station number	Average 18-kip equivalent		Proportion loaded			Highway system
		2S2	3S2	2D	2S2	3S2	
1	3	.4926	.8562	.570	.632	.500	Urban
2	4	.3738	.5360	.627	.524	.467	Urban
3	7	.5352	.6403	.806	.667	.656	03
4	16	.4820	.5505	.705	.675	.570	03
5	20	.4328	.7940	.771	.580	.669	03
6	42	.5444	.7730	.745	.600	.649	03
7	72	.4136	.5223	.622	.639	.669	03
8	81	.5933	1.0167	.756	.593	.589	03
9	88	.5712	1.2484	.667	.500	.580	03
10	101	.5676	.8332	.591	.824	.778	31
11	102	.4469	.8547	.569	.694	.674	31
12	145	.5356	.5322	.686	.683	.622	03
13	147	.2230	.7572	.600	.500	.825	03
14	149	.7106	.7302	.656	.688	.718	03
15	201	.7202	.9687	.600	.709	.718	01
16	202	.4339	.9580	.558	.673	.686	31
17	203	.6051	.9011	.576	.506	.777	01
18	301	.6015	.8171	.658	.694	.693	01
19	351	.4984	.9971	.659	.644	.644	31
20	371	.5280	.8309	.520	.537	.599	31
21	452	.6626	1.0746	.716	.740	.753	01

Source: Values in columns 3 through 7 were produced by FHWA weight-frequency program applied to Texas data for 1968 after highway system was modified to make possible output of station values rather than values for a highway system.

Table 2: - Texas truck weight stations paired by sum of least squared differences

Number of stations or station groups considered for pairing	Lowest sequence* number of the two stations or groups paired	Sequence number* of the other station or group paired	Sum of squared differences	Accumulated sum of squared differences
21	11	16	0.4173	0.4173
20	4	12	0.6612	1.0785
19	5	6	1.2202	2.2987
18	14	18	1.4158	3.7145
17	3	5	1.9796	5.6941
16	14	15	2.3525	8.0466
15	11	19	2.4417	10.4883
14	4	7	2.5696	13.0579
13	14	21	2.8887	15.9466
12	1	20	3.0954	19.0420
11	3	4	3.5447	22.5867
10	1	11	3.7542	26.3409
9	8	9	4.1355	30.4764
8	10	14	4.2724	34.7488
7	1	3	5.7284	40.4772
6	1	10	5.2207	45.6979
5	1	8	6.7853	52.4832
4	1	17	6.7163	59.1995
3	1	2	12.0416	71.2411
2	1	13	15.1805	86.4216

*Each group is identified by the station with the lowest sequence number.

Table 3: - Texas truck weight stations grouped at each level (pairing) by sum of least squared differences

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	11	4	5	14	3	14	11	4	14	1	3	1	8	10	1	1	1	1	1	1
	16	12	6	18	5	18	16	12	18	20	5	20	9	14	20	20	20	20	20	20
					6	15	19	7	15		6	11		18	11	11	11	11	11	11
									21		4	16		15	16	16	16	16	16	16
											12	19		21	19	19	19	19	19	19
											7				3	3	3	3	3	3
															5	5	5	5	5	5
															6	6	6	6	6	6
															4	4	4	4	4	4
															12	12	12	12	12	12
															7	7	7	7	7	7
																10	10	10	10	10
																14	14	14	14	14
																18	18	18	18	18
																15	15	15	15	15
																21	21	21	21	21
																	8	8	8	8
																	9	9	9	9
																		17	17	17
																			2	2
																				13

Figure 1 presents a plot of the data in columns 1 and 4 of Table 2. The plot shows a steep jump when seven station groups were considered for pairing. Thereafter the slope grows more steep with minor reversals. The decision was made to reject pairings at that point and at subsequent points. The final acceptable point is shown as level 14 in Table 3.

The cut-off point at the fourteenth level yielded four groups cotaining 18 stations and three ungrouped stations as indicated in Table 4. Data for Table 4 are taken from Table 3.

Table 4: - Grouping of Texas truck weight stations after the fourteenth pairing

Group	Sequence number	Station number	Highway system
1	3	7	03
	5	20	03
	6	42	03
	4	16	03
	12	145	03
	7	72	03
2	1	3	Urban
	20	371	31
	11	102	31
	16	202	31
	19	351	31
3	10	101	31
	14	149	03
	18	301	01
	15	201	01
	21	452	01
4	8	81	03
	9	88	03
Ungrouped	2	4	Urban
	13	147	03
	17	203	01

The appropriate parameters and data for the four groups were then used in conjunction with the consolidated program based on the IBM package to produce the coefficients for the following four equations and the values shown in Table 5 when the data for each of the 18 stations were substituted in each of the four equations.

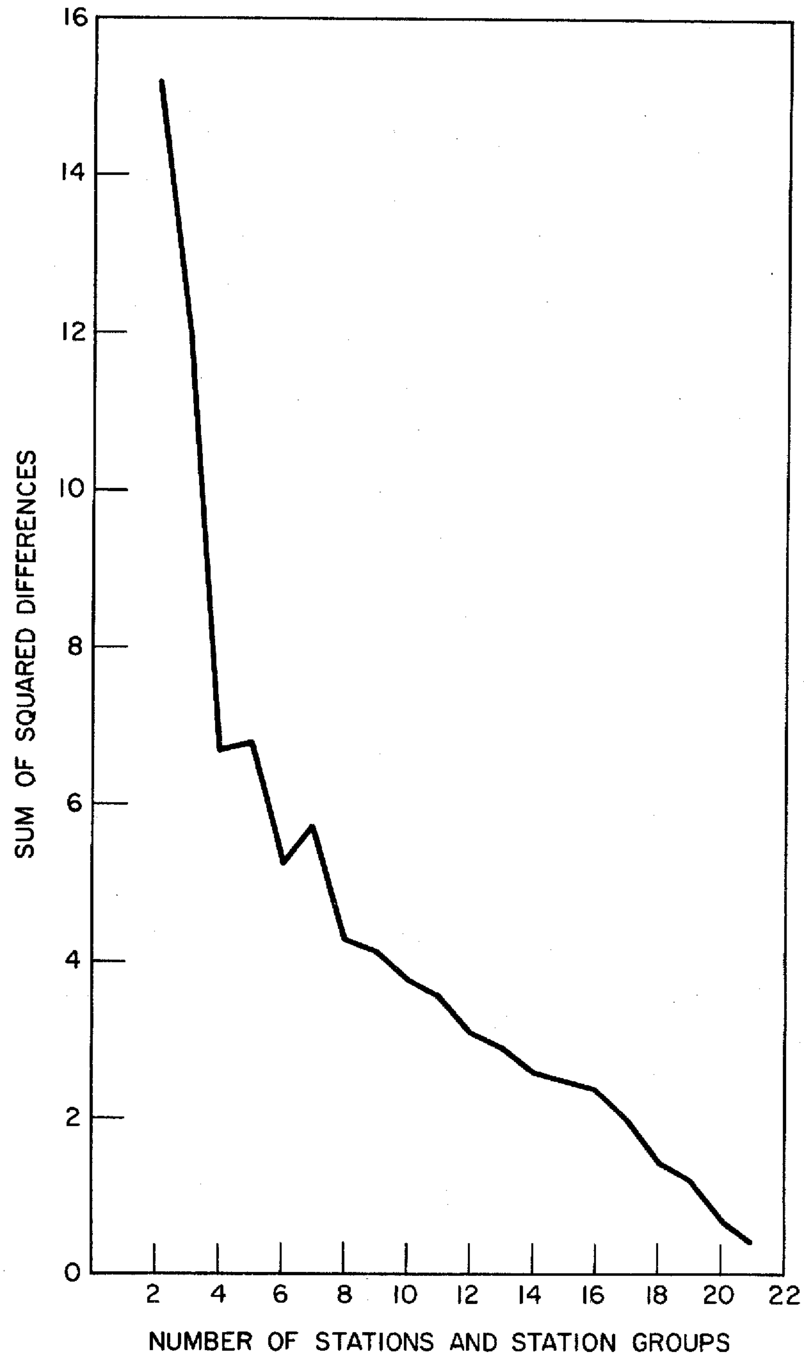


FIGURE 1: - THE SUM OF SQUARED DIFFERENCES CONTRIBUTED BY EACH SUCCESSIVE PAIRING VERSUS THE NUMBER OF STATIONS AND STATION GROUPS CONSIDERED FOR POSSIBLE PAIRING-TEXAS TRUCK WEIGHT STATIONS

(DATA FROM FIRST AND FOURTH COLUMNS OF TABLE 2)

Table 5: - Values yielded by the four discriminant functions applicable to Texas' truck weight stations

Group	Station sequence number	Values from equation I	Values from equation II	Values from equation III	Values from equation IV
1	3	291.825	280.513	278.818	280.404
1	4	247.937	240.959	230.398	237.703
1	5	241.909	235.791	221.629	235.819
1	6	269.813	265.595	259.952	266.154
1	7	229.845	223.759	211.594	217.005
1	12	272.666	265.400	262.418	261.690
2	1	210.958	222.639	203.462	220.988
2	11	248.900	257.684	247.271	253.944
2	16	242.029	254.284	241.701	251.375
2	19	257.370	265.988	255.461	266.805
2	20	211.341	221.756	207.175	218.673
3	10	337.479	344.476	355.114	341.065
3	14	339.146	338.972	353.663	338.176
3	15	341.381	352.641	366.378	353.557
3	18	310.254	311.344	315.624	311.226
3	21	355.888	363.659	376.001	367.276
4	8	272.419	277.055	269.541	281.788
4	9	232.219	248.376	232.189	254.121

$$\text{Group } 1 = -258.998+274.591x_1+7.153x_2+128.416x_3+242.118x_4+204.706x_5 \text{ ---I}$$

$$\text{Group } 2 = -244.469+278.638x_1+40.535x_2+72.340x_3+256.107x_4+184.102x_5 \text{ ---II}$$

$$\text{Group } 3 = -353.355+358.396x_1+31.728x_2+80.262x_3+295.341x_4+241.404x_5 \text{ ---III}$$

$$\text{Group } 4 = -267.963+289.724x_1+52.486x_2+93.399x_3+252.970x_4+176.342x_5 \text{ ---IV}$$

The values of the variables, x_1 through x_5, correspond with the items indicated in columns 3 through 7 of Table 1 and not the standardized values.

Examination of Table 5 shows that the station with sequence number 19 should have been assigned to group 4 rather than group 2 since equation IV yielded a higher value than equation II for that station. Accordingly, the shift was made, the program parameters were changed appropriately and the program was rerun. A new set of coefficients for four equations was produced and a new set of values for each station. The new set of values did not indicate the need for any further station shift between groups. The new equations follow.

$$\text{Group } 1 = -293.958+266.513x_1+20.419x_2+204.278x_3+237.978x_4+255.510x_5$$

$$\text{Group } 2 = -263.571+270.773x_1+51.992x_2+134.320x_3+246.588x_4+198.712x_5$$

$$\text{Group } 3 = -385.075+347.744x_1+45.808x_2+157.087x_3+287.355x_4+260.281x_5$$

$$\text{Group } 4 = -301.871+274.922x_1+60.232x_2+162.373x_3+258.289x_4+204.560x_5$$

In addition to the coefficients for the equations and to the values for each station from each equation, the program produces a statistic known as the Generalized Mahalanobis D^2. This can be used in conjunction with the tabulated values for the Chi Square distribution to test the hypothesis that the mean values are the same in all groups. The degrees of freedom equals the number of variables times one less then the number of groups. In the preceding investigation there were five times three or 15 degrees of freedom for the test. The value of D^2 was highly significant which indicated that the hypothesis should be rejected.

Two additional investigations were carried out with truck weight station data. In the first, the three remaining stations were treated as a fifth group. Five equations were produced as follows:

$$\text{Group } 1 = -242.550+87.102x_1+22.790x_2+283.922x_3+273.229x_4+74.638x_5$$

$$\text{Group } 2 = -203.409+82.028x_1+49.943x_2+210.645x_3+276.239x_4+52.451x_5$$

$$\text{Group } 3 = -279.961+111.533x_1+43.128x_2+250.615x_3+318.249x_4+72.954x_5$$

$$\text{Group } 4 = -237.414+82.831x_1+60.152x_2+252.793x_3+274.973x_4+47.754x_5$$

$$\text{Group } 5 = -177.335+69.327x_1+28.978x_2+233.827x_3+223.178x_4+74.458x_5$$

This shows that a linear discriminant function can be produced with any sets of data; each set must contain two or more components. For the functions to be meaningful, the components of each set must be fairly homogeneous on the basis of recognized criteria or measurements.

In the second investigation, the percentages of eight classes of vehicles passing the 21 loadometer stations were used for grouping. This was undertaken because the report of Messrs. Wegmann and Pant was based upon vehicle classification data rather than truck weight data. The latter data were not available for individual stations in West Virginia. On page 65 of their report, it is stated, "Such average values as shown above could be used as a scale for identifying truck weight characteristics, if a correlation could be made between truck traffic mix and average weight characteristics." The eight classes of vehicles were:

2-axle, 4-tire trucks

2-axle, 6-tire "

3-axle, single unit trucks

2S1 tractor, semi-trailer

2S2 " " "

3S2 " " "

Tractor trains

Truck trailers

Comparison of the entries in tables 6 and 7 with the corresponding entries in tables 2 and 3 shows that classification data do not yield the same groupings as truck weight data. The groupings may be acceptable in terms of classification count data, but should not be used for determining similarity of truck weight characteristics.

B. ATR stations

Maryland's report, "Traffic Trends - 1969," shows 37 ATR stations classified into four groups. Table 8 presents Maryland's grouping and the factor for each month for each station. With some exceptions, the factors for any month within a group fall within the 0.20 range criterion recommended in the "Guide for Traffic Volume Counting Manual." Deviations from that criterion are probably justified by knowledge of unusual conditions and values for previous years. The 2-step grouping procedure was applied to the single year's data shown in Table 8.

Tables 9 and 10 present the output for the ATR data corresponding to the truck weight station output shown in tables 2 and 3. Station sequence numbers 1 through 33 are also Maryland's ATR station numbers. Sequence numbers 34 through 37 represent Maryland's ATR stations 35 through 38. Maryland's report showed no data for station 34.

Figure 2 presents a plot of the data in columns 1 and 4 of Table 9. The plot shows a slope that increases rapidly after seven station groups were considered for pairing. The decision was made to accept groupings determined at that point. That final acceptable point is shown as level 31 in Table 10.

The resulting grouping are shown in Table 11. Table 11 also compares the results of grouping by sum of least squared differences with the groupings in Table 8. The few disagreements may in part be accounted for by discrepancies from the 0.20 range in Maryland's groups.

It was decided to treat stations 17, 21, and 35 as a fourth group for input to the linear discriminant function program. That program yielded the following four equations:

Table 6: - Texas truck weight stations paired by sum of least squared differences based on vehicle classification count data

Number of stations or station groups considered for pairing	Lowest sequence number of the two stations or groups paired	Sequence number of the other station or group paired	Sum of squared differences	Accumulated sum of squared differences
21	8	16	.7500	.7500
20	7	12	1.2283	1.9738
19	1	2	1.3950	3.3734
18	11	20	1.6289	5.0023
17	5	15	2.8637	7.8660
16	8	18	2.9228	10.7888
15	7	13	2.9959	13.7847
14	5	21	3.4569	17.2416
13	3	8	3.7031	20.9447
12	4	7	4.1436	25.0883
11	5	11	4.4385	29.5267
10	3	4	4.5978	34.1246
9	6	9	4.8175	38.9421
8	3	5	5.8510	44.7931
7	1	6	7.2672	52.0602
6	3	19	9.1539	61.2142
5	3	17	12.1324	73.3465
4	3	10	14.1500	87.4965
3	3	14	16.3508	103.8473
2	1	3	18.9162	122.7635

Table 7: - Texas truck weight stations grouped at each level (pairing) using classification count data

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	8	7	1	11	5	8	7	5	3	4	5	3	6	3	1	3	3	3	3	1
	16	12	2	20	15	16	12	15	8	7	15	8	9	8	2	8	8	8	8	2
						18	13	21	16	12	21	16		16	6	16	16	16	16	6
									18	13	11	18		18	9	18	18	18	18	9
											20	4		4		4	4	4	4	3
												7		7		7	7	7	7	8
												12		12		12	12	12	12	16
												13		13		13	13	13	13	18
														5		5	5	5	5	4
														15		15	15	15	15	7
														21		21	21	21	21	12
														11		11	11	11	11	13
														20		20	20	20	20	5
																19	19	19	19	15
																	17	17	17	21
																		10	10	11
																			14	20
																				19
																				17
																				10
																				14

Table 8: - Maryland ATR stations grouped by the criterion of an 0.20 range for the monthly factors

GROUP 1 - Roads with consistent yearly traffic and a moderate seasonal traffic peak

STA. NO.	JAN.	FEB.	MAR.	APR.	MAY	JUNE	JULY	AUG.	SEPT.	OCT.	NOV.	DEC.	AVG.
01	1.11	1.05	1.12	1.15	0.93	0.92	0.95	0.93	0.94	0.88	0.97	0.95	0.99
02	1.12	1.07	1.02	0.99	1.02	0.98	0.97	0.98	1.05	0.97	1.01	1.01	1.02
03	1.05	0.99	0.95	0.96	0.96	0.99	1.10	1.05	1.03	1.06	1.01	0.97	1.01
05	1.14	1.08	1.03	0.96	0.97	0.96	0.96	0.96	0.98	1.05	1.06	1.07	1.02
07	1.13	1.09	1.03	0.97	0.95	0.92	0.91	0.89	0.95	1.04	1.05	1.10	1.00
08	1.15	1.07	1.00	0.96	0.96	0.88	0.94	0.97	0.86	0.92	1.03	1.09	0.99
09	1.08	1.05	1.03	0.95	0.97	0.95	0.89	0.90	0.99	1.00	0.98	0.97	0.98
10	1.12	1.09	1.06	1.00	0.96	0.92	0.91	0.92	0.93	1.00	1.01	1.00	0.99
11	1.11	1.03	0.99	0.95	0.94	0.93	0.96	0.93	0.92	0.97	1.01	1.01	0.98
12	1.08	1.00	0.98	0.97	0.95	0.93	0.92	0.89	0.95	0.95	0.93	1.07	0.97
14	1.06	1.19	1.12	1.02	0.94	0.97	0.92	0.86	0.93	1.02	1.06	1.15	1.04
18	1.11	1.03	0.95	0.96	1.02	1.00	0.95	0.92	0.98	0.99	0.99	0.96	0.99
20	1.12	1.10	1.03	1.00	1.06	1.15	0.95	0.92	0.97	0.93	0.91	1.02	1.01
23	1.14	1.08	1.06	1.03	1.00	1.03	1.04	1.13	1.00	1.06	1.06	1.04	1.06
25	1.10	1.11	1.05	0.96	1.04	0.98	0.96	0.89	0.99	1.21	1.14	0.96	1.03
26	1.17	1.07	1.03	0.99	0.96	0.96	0.92	0.91	0.94	0.96	1.02	1.04	1.00
27	1.15	1.11	1.08	0.99	1.01	0.96	0.92	0.85	0.96	1.11	1.10	1.06	1.03
28	1.27	1.19	1.13	1.04	1.01	0.91	0.92	0.86	0.95	1.03	1.08	1.14	1.04
29	1.18	1.11	1.03	0.95	0.94	0.90	0.90	0.86	0.92	1.05	1.09	1.14	1.01
33	1.42	1.45	1.16	1.10	1.10	1.08	0.94	0.87	0.95	1.08	1.02	1.25	1.12
38	1.06	1.01	1.00	0.99	0.96	0.97	0.87	0.84	0.93	1.02	1.03	1.05	0.98
Avg.	1.15	1.09	1.04	0.99	0.98	0.97	0.94	0.92	0.96	1.01	1.03	1.05	1.01

Table 8: - continued

(GROUPS OF STATIONS WITHIN .20 RANGE)

GROUP II - Roads with consistent yearly traffic

STA. No.	JAN.	FEB.	MAR.	APR.	MAY	JUNE	JULY	AUG.	SEPT.	OCT.	NOV.	DEC.	AVG.
13	1.09	0.99	0.97	0.99	0.99	0.96	0.97	0.86	0.89	0.85	0.85	0.88	0.94
31	1.06	1.00	0.96	0.89	0.91	0.91	0.94	0.84	0.89	0.95	1.03	0.99	0.95
32	1.14	1.01	0.96	0.93	0.93	0.87	0.90	0.88	0.87	0.96	1.03	1.04	0.96
Avg.	1.10	1.00	0.96	0.94	0.94	0.91	0.94	0.86	0.88	0.92	0.97	0.97	0.95

GROUP III - Roads with inappreciable consistent yearly traffic and a moderate seasonal traffic peak

STA. NO.	JAN.	FEB.	MAR.	APR.	MAY	JUNE	JULY	AUG.	SEPT.	OCT.	NOV.	DEC.	AVG.
04	1.37	1.31	1.24	1.07	1.14	0.96	0.85	0.78	0.88	1.21	1.16	1.19	1.10
06	1.31	1.20	1.12	0.98	0.92	0.92	0.93	0.92	0.98	1.30	1.36	1.36	1.11
15	1.40	1.33	1.21	1.04	1.00	0.98	0.84	0.85	1.03	1.26	1.22	1.26	1.12
19	1.71	1.30	1.19	1.04	0.99	0.98	0.86	0.94	0.94	1.09	1.19	1.26	1.12
21	2.85	2.72	2.59	2.41	2.18	1.66	1.77	1.72	1.79	0.28	0.36	2.52	1.90
22	1.44	1.42	1.34	1.22	1.09	0.85	0.78	0.74	1.02	1.26	1.22	1.26	1.14
24	1.33	1.26	1.24	1.13	1.03	0.89	0.83	0.81	1.00	1.19	1.20	1.28	1.10
30	1.33	1.21	1.13	0.99	1.00	0.95	0.90	0.84	0.95	1.11	1.17	1.24	1.07
37	1.25	1.25	1.24	1.10	1.07	0.96	0.75	0.83	1.03	1.07	1.08	1.05	1.06
Avg.	1.55	1.44	1.37	1.22	1.16	1.02	0.95	0.94	1.07	1.09	1.11	1.38	1.19

GROUP IV - Roads with inappreciable consistent yearly traffic and a high seasonal traffic peak

STA. NO.	JAN.	FEB.	MAR.	APR.	MAY	JUNE	JULY	AUG.	SEPT.	OCT.	NOV.	DEC.	AVG.
16	1.51	1.50	1.43	1.24	0.96	0.98	0.77	0.60	0.92	1.45	1.47	1.44	1.19
17	2.12	2.03	1.82	1.40	1.03	0.71	0.54	0.51	1.08	1.75	1.88	2.01	1.41
35	1.76	2.13	1.87	1.25	1.15	1.03	0.69	0.53	1.00	1.37	1.52	1.53	1.32
36	1.53	1.45	1.38	1.19	1.04	0.97	0.75	0.68	0.95	1.25	1.33	1.43	1.16
Avg.	1.73	1.78	1.63	1.27	1.05	0.92	0.69	0.58	0.99	1.46	1.55	1.60	1.27

Table 9: - Maryland ATR stations paired by sum of least squared differences

Number of stations or station groups considered for pairing	Lowest sequence number of the two stations or groups paired	Sequence number of the other station or group paired	Sum of squared differences	Accumulated sum of squared differences
37	7	29	0.1799	0.1799
36	10	26	0.1886	0.3685
35	10	11	0.1991	0.5676
34	10	12	0.3138	0.8814
33	9	10	0.3075	1.1889
32	31	32	0.3296	1.5185
31	14	28	0.3617	1.8802
30	9	37	0.4236	2.3038
29	7	9	0.4282	2.7320
28	5	7	0.4445	3.1765
27	2	18	0.5006	3.6771
26	14	27	0.5335	4.2106
25	14	30	0.5410	4.7516
24	5	31	0.6037	5.3553
23	5	8	0.7530	6.1083
22	15	24	0.9071	7.0154
21	3	23	0.9206	7.9360
20	2	5	0.9263	8.8623
19	2	14	0.9606	9.8229
18	1	2	1.2926	11.1155
17	4	15	1.5188	12.6343
16	4	22	1.5711	14.2054
15	1	25	1.6997	15.9051
14	16	35	1.7254	17.6305
13	4	36	1.8767	19.5072
12	1	13	1.9712	21.4784
11	1	3	2.4724	23.9508
10	4	19	2.5280	26.4788
9	4	33	2.7473	29.2261
8	1	20	2.9895	32.2156
7	4	6	3.0947	35.3103
6	1	4	3.7763	39.0866
5	16	34	8.0188	47.1054
4	1	16	17.0750	64.1804
3	1	17	62.1799	126.3603
2	1	21	281.7903	408.1506

Table 10: - Maryland ATR stations grouped at each level (pairing) by sum of least squared differences

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	34	35	36
7	10	10	10	9	31	14	9	7	5	2	14	14	5	5	15	3	2	2	1	4	4	1	16	4	1	1	4	4	1	4	1	16	1	1	1
29	26	26	26	10	32	28	10	29	7	18	28	28	7	7	24	23	18	18	2	15	15	2	35	15	2	2	15	15	2	15	2	35	2	2	2
		11	11	26			26	9	29		27	27	29	29			5	5	18	24	24	18		24	18	18	24	24	18	24	18	34	18	18	18
			12	11			11	10	9			30	9	9			7	7	5		22	5		36	5	5	36	36	5	36	5		5	5	5
				12			12	26	10				10	10			29	29	7			7			7	7	22	22	7	22	7		7	7	7
							37	11	26				26	26			9	9	29			29			29	29	19	19	29	19	29		29	29	29
								12	11				11	11			10	10	9			9			9	9		33	9	33	9		9	9	9
								37	12				12	12			26	26	10			10			10	10			10	6	10		10	10	10
									37				37	37			11	11	26			26			26	26			26		26		26	26	26
													31	31			12	12	11			11			11	11			11		11		11	11	11
													32	32			37	37	12			12			12	12			12		12		12	12	12
														8			31	31	37			37			37	37			37		37		37	37	37
																	32	32	31			31			31	31			31		31		31	31	31
																	8	8	32			32			32	32			32		32		32	32	32
																		14	8			8			8	8			8		8		8	8	8
																		28	14			14			14	14			14		14		14	14	14
																		27	28			28			28	28			28		28		28	28	28
																		30	27			27			27	27			27		27		27	27	27
																			30			30			30	30			30		30		30	30	30
																						25			25	25			25		25		25	25	25
																									13	13			13		13		13	13	13
																										3			3		3		3	3	3
																										23			23		23		23	23	23
																													20		20		20	20	20
																															4		4	4	4
																															15		15	15	15
																															24		24	24	24
																															36		36	36	36
																															22		22	22	22
																															19		19	19	19
																															33		33	33	33
																															6		6	6	6
																																	16	16	16
																																	35	35	35
																																	34	34	34
																																		17	17
																																			21

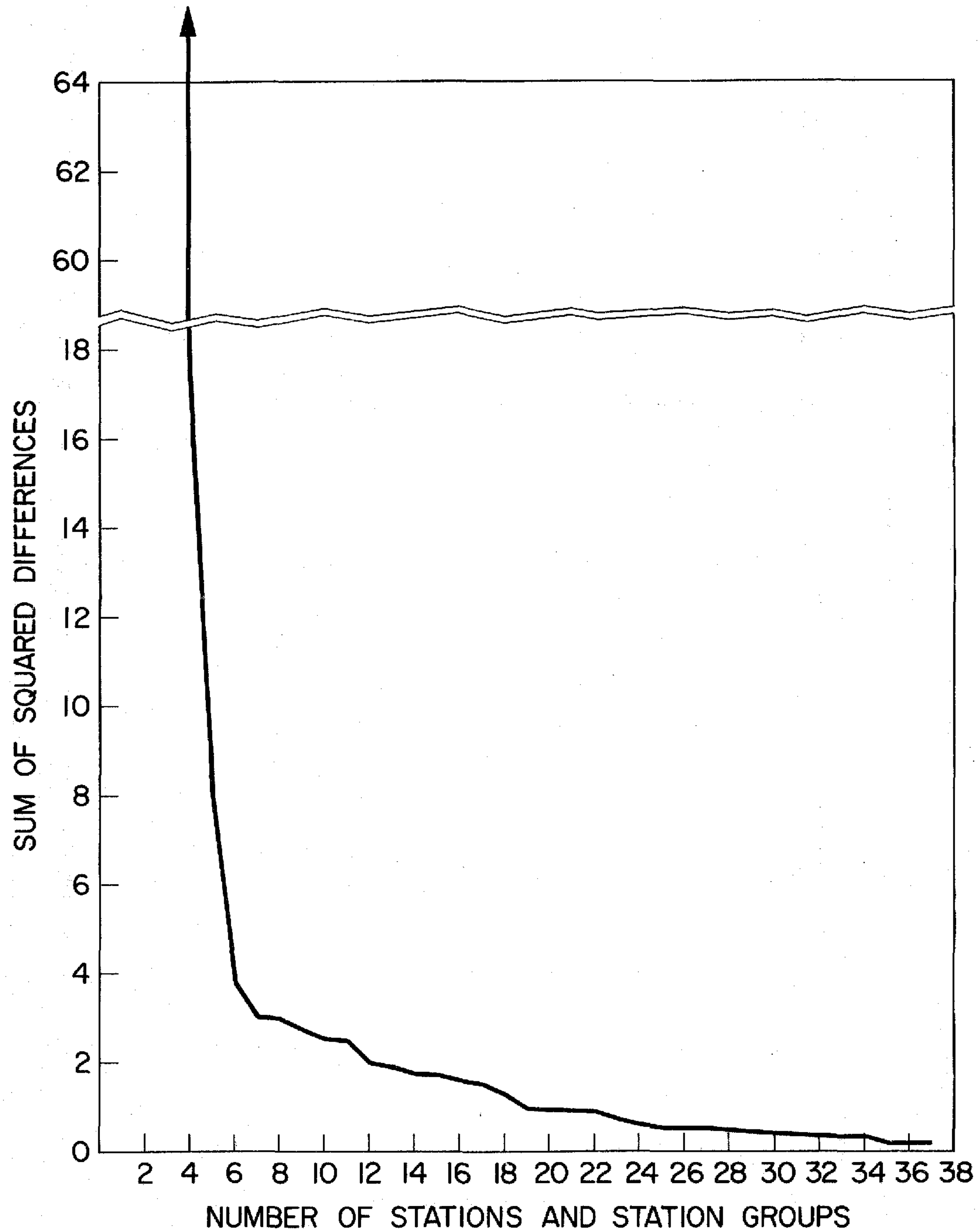


FIGURE 2:-THE SUM OF SQUARED DIFFERENCES CONTRIBUTED BY EACH SUCCESSIVE PAIRING VERSUS THE NUMBER OF STATIONS AND STATION GROUPS CONSIDERED FOR POSSIBLE PAIRING-MARYLAND ATR STATIONS

(DATA FROM FIRST AND FOURTH COLUMNS OF TABLE 9)

Table 11: - Maryland ATR stations grouped by sum of least squared differences and the groupings shown in Table 8 for these stations

Group A		Group B		Group C		Group D 1/	
Step 31	Table 8	Step 30	Table 8	Step 24	Table 8	Step 31 1/	Table 8
4	III	1	I	16	IV	17	IV
6	III	2	I	36 2/	IV	21	III
15	III	3	I			35 2/	IV
19	III	5	I				
22	III	7	I				
24	III	8	I				
33	I	9	I				
37 2/	III	10	I				
		11	I				
		12	I				
		13	II				
		14	I				
		18	I				
		20	I				
		23	I				
		25	I				
		26	I				
		27	I				
		28	I				
		29	I				
		30	III				
		31	II				
		32	II				
		38 2/	I				

1/ Group D contains the three stations ungrouped at the end of step 31

2/ The sequence numbers are increased by one to give station numbers

Table 12: - Values yielded by the four discriminant functions applicable to Maryland's ATR stations

Group	Station number	Values from equation I	Values from equation II	Values from equation III	Values from equation IV
1	4	849.384	839.154	825.541	727.184
1	6	894.686	892.855	877.724	784.275
1	15	834.252	822.647	809.514	733.397
1	19	857.164	845.055	831.096	745.666
1	22	875.415	859.284	855.549	793.482
1	24	849.587	843.435	831.425	742.131
1	33	844.870	833.890	820.498	740.317
1	37	838.217	829.773	810.500	729.502
2	1	799.257	811.277	778.623	652.807
2	2	820.959	831.907	782.628	673.666
2	3	758.236	775.938	713.885	580.191
2	5	768.140	776.772	729.925	616.185
2	7	739.222	747.510	702.547	589.607
2	8	754.778	765.292	713.657	600.718
2	9	706.531	715.863	662.405	547.679
2	10	729.752	737.098	689.910	576.969
2	11	740.407	753.615	701.687	578.065
2	12	661.728	676.274	619.990	485.517
2	13	702.084	720.062	662.819	518.743
2	14	770.753	773.822	747.656	643.598
2	18	813.949	826.231	776.156	646.736
2	20	816.587	824.106	793.084	657.242
2	23	819.468	826.003	782.159	660.276
2	25	848.844	854.277	813.161	699.001
2	26	785.539	794.830	754.431	635.212
2	27	814.662	820.668	786.319	670.441
2	28	803.769	807.116	774.806	677.488
2	29	753.638	760.986	719.395	610.848
2	30	842.515	842.579	820.783	724.465
2	31	748.545	765.572	714.785	591.128
2	32	729.920	743.029	689.276	564.105
2	38	784.025	794.938	758.677	620.383
3	16	1046.188	1023.155	1078.278	1002.383
3	36	991.335	971.717	1004.682	934.322
4	17	1157.649	1097.156	1190.723	1249.642
4	21	1169.875	1112.321	1201.900	1286.290
4	35	1136.653	1081.972	1170.039	1246.108

$$\text{Group A} = -855.442-50.510x_1+137.960x_2-476.302x_3+564.814x_4+449.176x_5 +479.931x_6+88.407x_7-176.958x_8+117.455x_9-597.222x_{10} +1304.132x_{11}-188.124x_{12}$$

$$\text{Group B} = -781.064-60.379x_1+101.094x_2-488.891x_3+563.069x_4+444.960x_5 +466.781x_6+134.975x_7-201.672x_8+129.194x_9-618.301x_{10} +1306.764x_{11}-184.820x_{12}$$

$$\text{Group C} = -1041.474-66.602x_1+111.977x_2-464.678x_3+656.681x_4+420.128x_5 +594.591x_6+117.695x_7-258.085x_8+101.452x_9-666.521x_{10} +1426.918x_{11}-169.226x_{12}$$

$$\text{Group D} = -1260.673-87.528x_1+326.157x_2-431.197x_3+554.365x_4+409.640x_5 +488.461x_6+91.209x_7-224.626x_8+210.119x_9-746.116x_{10} +1495.908x_{11}-196.279x_{12}$$

The values of x_1 through x_{12} are the monthly factors shown in Table 8.

The four equations yielded the values shown in Table 12 for each station. These values indicate that all stations are grouped appropriately. Some of the discriminant values are so close as to indicate possible need for testing stability of grouping over a period of several years.

Once groupings have been firmly established, the equations can be used to assign a new location to a group.

CONCLUSIONS

1. The 2-step grouping procedure provides an objective methodology which yields acceptable results.

2. The first step, grouping by sum of least square differences can yield near optimum groupings provided good judgment is exercised as to the cut-off point.

3. Eighteen of the 21 truck weight stations in Texas formed four fairly homogeneous groups in 1968 as did 34 of the 37 ATR stations in Maryland during 1969.

4. A final determination of groupings should be based on consistency of allocation over several years--possibly three years.

5. Classification count data are not suitable by themselves for grouping truck weight stations.