AN APPLICATION OF DISCRIMINANT ANALYSIS TO THE DIVISION OF TRAFFIC BETWEEN TRANSPORT MODES

IWR

INSTITUTE
FOR
WATER RESOURCES

DEPARTMENT OF THE ARMY
CORPS OF ENGINEERS

MAY 1971

IWR REPORT 71-2

INSTITUTE FOR WATER RESOURCES

The Institute for Water Resources is the focal point of efforts to improve U. S. Army Corps of Engineers' planning for the development and management of the nation's water and related land resources, particularly as national objectives are affected by changing public values. Through in-house and contract studies, the Institute identifies emerging planning issues and requirements; evaluates planning efficiency and effectiveness; develops applications to Corps' planning of analytical and projectional techniques and planning management systems; and formulates the bases for improved Corps' planning policies and procedures. In concert with the Office, Chief of Engineers, it advises and assists Corps' agencies in utilizing such techniques and carrying out such procedures. Finally, it evaluates and recommends improvements in the professional development program for Corps' planners; develops training programs; and conducts training.

The Institute welcomes your comments and suggestions on its program and publications.

R. H. GROVES
Brigadier General, USA
Director

An Application of Discriminant Analysis to the Division of Traffic between Transport Modes

A Report by the
Center for Economic Studies
U.S. Army Engineer Institute for Water Resources
206 North Washington Street
Alexandria, Virginia 22314

by
Lloyd G. Antle
Richard W. Haynes

Copies may be purchased from:
The National Technical Information Service
Department of Commerce
Springfield, Virginia 22151

May 1971

IWR Report 71-2

FOREWARD

A. Purpose

This research was undertaken to test the operational application of a methodology suggested by Moses (1970) for applying discriminant analysis to estimate the division of traffic between transport modes. A task force in the Ohio River Division gathered data for the model and staff members of the Institute for Water Resources analyzed the data and prepared this report.

B. Findings

As used in transportation studies, discriminant analysis, statistically relates observed choice of mode to the characteristics of both the shipment and the chosen mode. The technique offers one method for estimating the demand schedule at a stated time for transportation relevant to a given mode.

A test of the concept of discriminant analysis using coal shipments in the Upper Ohio River Region was conducted and the results are presented below. A demand function for waterway transportation for coal was developed under competitive conditions observed in 1969.

C. Assessment

It is recommended that this work be extended to cover a wider geographic area, to include other commodities, and to explore other characteristics which might significantly influence modal choice. This wider coverage should include the choice of unit trains which may constitute a distinct statistical population. Development of procedures for application of the modal split model to projected futures (economies and technologies) should be undertaken. The level of effort was quite limited (both in quantity and regional extent of data gathered) so that the learning process could be expedited. Therefore the conclusions of the study should not be generalized to other regions or commodities.

D. Status

This research represents the findings, conclusions and independent judgment of the team of researchers. It is therefore not to be construed to represent the view of the Corps of Engineers. Policy and procedural changes which may result from this research will be implemented by directives and guidelines provided by the Chief of Engineers through command channels.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

I. INTRODUCTION

Every investment decision in transportation requires some explicit analysis of the quantity of traffic expected to use a facility and the benefits relevant to that quantity of traffic affected by the investment plan.

The utilization of one of the tools of multivariate analysis, discriminant analysis, and a theoretical basis for estimating the demand for waterway transportation was advanced by Leon Moses et al, in research recently concluded for the Corps of Engineers.*/

It should be noted that an operational modal split model is required if systematic analysis of the inland waterway system is to be approached in the context of a multimode National transportation system.

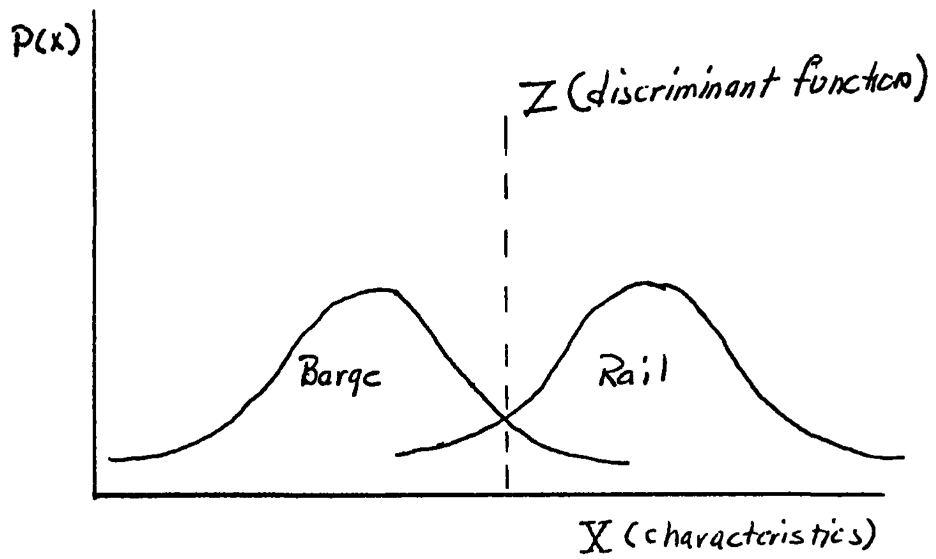
The subject research is directed towards the operational testing of discriminant analysis as a potential procedure for improved estimates of both quantity and benefits of traffic expected to use a waterway. As will be indicated later, a considerable measure of success was achieved in developing a modal split model on the Upper Ohio River System for coal movements.

---

*/ See Leon Moses, et al, Cost-Benefit Analysis for Inland Navigation Improvements, (1970), IWR 70-4, in 3 Volumes. The first volume contains a summary report, the second includes several papers dealing with cost and demand functions and the third volume deals with an economic criticism of regulation effects on intermodal behavior and includes a model of regulatory behavior.

A. Use of Discriminant Analysis to Estimate Demand

Discriminant analysis is a means to statistically weigh transportation modal characteristics and thus determine the characteristics influence on choice of mode. A function is developed dividing mode populations (assuming each to be distributed normally) in such a way as to minimize errors in classification amongst modes. Thus, the function Z reflects the characteristics which define each modal population as shown below.



The discriminant function developed in this investigation is in the form $Z = f(X_1, X_2, \ldots\ldots X_n)$ where the characteristics (X_i) determined to influence modal choice include the following:

X_1 Annual Tonnage of the Commodity Movement

X_2 Distance Hauled (miles)

X_3 Average Travel Time

X_4 Average Shipment Tonnage

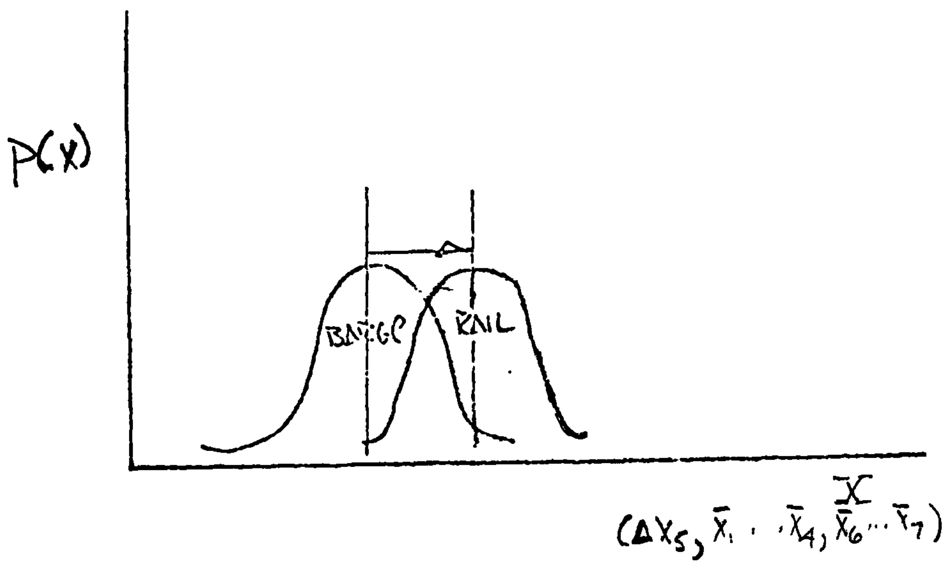
X_5 Rate of Selected Mode

X_6 Rate of Alternative Mode

X_7 Handling Cost Using Selected Mode

When the discriminant function (Z) is estimated, characteristics of each movement can be introduced and a $\hat{Z}$ test value calculated. Test values (in the two mode case) indicate to which mode a particular movement should be assigned. A more complicated classification procedure is required in a three mode case.

Given the discriminant function, the demand for transportation by each mode can be estimated by holding all chracteristics except rate constant, increasing the rate and calculating the shift from the selected mode into other modes as demonstrated below.



II. DEVELOPMENT OF A DATA BASE

Data for the analysis were gathered during the summer of 1970.*/ The project was begun with the idea of surveying firms receiving shipments of coal, chemicals and petroleum (located in the Upper Ohio River Valley). The first field investigation was localized in the Huntington-Pittsburgh reach of the river. About 30 firms were interviewed, but only half of the interviews resulted in complete observations over the several characteristics. The resulting acceptable observations were heavily weighted towards the exclusive choice of the waterway mode. Because of this apparent imbalance and to enlarge the sample, firms were surveyed as far as 100 miles from the main stem of the Ohio River as well as along the main stem. Some 87 complete observations were ultimately gathered representing three commodity groups and covering seven characteristics which were considered to be important in explaining modal choice. The following sections deal with these characteristics.

A. Selection of Characteristics

The initial selection of characteristics descriptive of rail and barge movement was based on four studies--Allen**/, Beuthe**/, Stucker**/,

---

*/ The data were gathered by a Task Force from the Ohio River Division. Copies of the Task Force report which includes their comments with respect to sample design, survey procedures and recommendations with respect to further research are available from IWR on request.

**/ See Volume II of IWR Report 70-4, Cost-Benefit Analysis for Inland Navigation Improvements, (1970).

and Herendeen*/. Both Allen and Beuthe essentially selected the same characteristics including:

1. Quantity the firm produces and ships.

2. Market price of the product.

3. Time required to ship the goods from point of production to the point of delivery.

4. Transport charge per unit of product.

5. Interest rate.

6. Damage, pilferage, loss or perishability rate.

Stucker deviated in his initial variable selection somewhat. He suggests that the following be used:

1. Time required.

2. Schedules and convenience of shipping times.

3. Reliability of schedules.

4. Breakage, spoilage, and deterioration of product enroute.

5. Packaging or special handling costs.

6. Interface costs or joint hauls.

Herendeen used the following variables in his study of the theoretical development and his preliminary testing of a mathematical model for predicting freight modal split.

1. Reliability of mode k

2. Relative cost = $\frac{\text{lowest rate by all available modes}}{\text{rate by mode k}}$

---

*/ Herendeen, J. H., Jr. 1969, "Theoretical Development and Preliminary Testing of a Mathematical Model for Predicting Freight Modal Split," Pennsylvania Transportation and Traffic Safety Center and Department of Civil Engineering, Pennsylvania State University.

3. Relative transit time by mode =

$$\frac{\text{lowest time by all available modes}}{\text{time by mode k}}$$

4. Relative frequency of service = $\frac{\text{frequency of service by mode k}}{\text{best frequency of service}}$

The previous work indicates that the selection of variables can be divided into four general areas:

1. Quantity shipped.

2. Time required (both major and alternative).

3. Rates (both major and alternative).

4. Quality of transport service.

After the data collection started it became evident that acceptable measurements could not be obtained during the survey for certain characteristics*/ and that certain modification of characteristics would be desirable. The characteristics finally used include:

X_1 Annual Tonnage of the Commodity Movement

X_2 Distance (miles)

X_3 Average Travel Time

X_4 Average Shipment

X_5 Rate

X_6 Alternative Rate

X_7 Handling Cost

---

*/ Indicating the quality of service (losses, pilferage, deterioration of quality of goods) is probably not significant between modes for the commodities under study.

There are three obvious statistical problems with these chosen characteristics. The first of these is collinearity between the distance and travel time and between annual and average tonnage shipped. The second is that many firms do not know the exact rate by alternative modes, requiring reconstruction of alternative rates. The final problem was that of determining handling costs in a consistent manner. Most respondants included only the direct (out-of-pocket) cost involved and ignored sunk costs, an acceptable procedure.

B. Data Summary

Data were collected from 63 firms covering 92 movements of which 87 were complete observations. These observations are:

Number of Observations		Commodity and SIC Code
Barge:	28	Coal (121)
	6	Petroleum (Crude) (131)
	9	Industrial Chemicals (281)
	3	Refined Petroleum (291)
	4	Coke (331)
Rail:	23	Coal (121)
	1	Refined Petroleum (291)
	5	Coke (331)
Joint:	8	Coal (121)
	87	

Table II-1 shows the mean of the observations collected for each commodity listed above, while Table II-2 presents the observations utilized in subsequent analysis.

TABLE II-1
Group Means of Observations by SIC Code and Selected Mode

Mode	Commodity	Annual Tonnage 1 (tons)	Distance of Shipment 2 (miles)	Hours Shipment Time 3	Average Shipment (Tons) 4	Rate for Selected Mode ($) 5	Rate of Alternative Mode ($) 6	Handling Costs ($) 7
Barge	Coal	867,016.607	98.929	49.571	13,416.071	.532	2.768	.136
	Petroleum	319,062.500	820.000	197.000	786.167	5.245	13.490	.007
	Ind. Chem.	122,168.889	933.556	244.667	3,081.667	2.686	11.462	.015
	Refined Petroleum	1,042,153.333	181.667	80.000	4,025.000	.760	11.740	.000
	Coke	44,750.000	103.750	66.000	850.000	1.025	2.590	.500
Rail	Coal	146,022.043	142.261	93.739	876.522	4.115	3.947	.670
	Refined Petroleum	185,000.000	60.000	24.000	350.000	9.000	.120	3.000
	Coke	30,769.600	428.000	127.200	2,480.000	6.176	7.820	.382
Joint	Coal	436,256.125	203.000	67.125	3,150.000	2.627	4.655	.222

TABLE II-2 SUMMARY OF OBSERVED DATA

SELECTED MODE	SIC CODE	ANNUAL TONNAGE	TRANSIT TIME	TRANSIT DISTANCE	AVG. SHIPMENT	RATE	ALT. RATE	HANDLING COSTS	ALT. MODE
1	121	781000	7.	1.	4050.	.15	.65	.180	2
1	121	307000.	2.	1.	3000.	.10	.65	.180	2
1	121	193000.	20.	1.	3000.	.20	.65	.180	2
1	121	347000.	29.	4.	3750.	.21	1.50	.090	2
1	121	135000.	50.	7.	4500.	.30	1.56	.090	2
1	121	732000.	56.	7.	3600.	.35	1.56	.090	2
1	121	1207329.	92.	48.	9000.	.57	3.49	.000	2
1	121	609677.	127.	48.	9000.	.74	3.52	.000	2
1	121	204000.	155.	48.	3600.	.75	4.96	.110	2
1	121	2893654.	154.	36.	6000.	.60	4.82	.060	2
1	121	3070000.	6.	1.	9400.	.09	.35	.005	2
1	121	4500000.	19.	2.	0000.	.16	.55	.005	2
1	121	589000.	70.	24.	1400.	.42	3.00	.005	2
1	121	200000.	2.	1.	1800.	.03	.85	.005	2
1	121	300000.	260.	60.	5400.	1.60	1.56	.005	2
1	121	500000.	117.	48.	9000.	.76	3.57	.005	2
1	121	46600.	95.	36.	900.	.75	3.35	.750	2
1	121	900000.	83.	168.	6400.	.60	3.49	.000	2
1	121	1100000.	245.	168.	6400.	1.30	3.95	.000	2
1	121	365000.	45.	120.	2600.	.50	1.56	.000	2
1	121	43500.	340.	168.	2600.	1.45	5.58	.000	2
1	121	1369500.	41.	24.	3600.	.69	3.57	.250	2
1	121	1369500.	38.	24.	3600.	.70	3.57	.250	2
1	121	1369500.	60.	24.	3600.	.41	3.57	.250	2
1	121	424319.	147.	19.	2600.	.56	4.48	.300	2
1	121	22500.	90.	24.	0000.	.36	3.57	.250	2
1	121	27500.	300.	240.	3000.	.36	5.22	.250	2
1	121	469686.	120.	36.	1650.	.15	1.56	.500	2
2	121	278125.	118.	48.	500.	3.21	2.69	.220	3
2	121	260900.	100.	48.	500.	3.13	2.69	.220	3
2	121	25591.	245.	6.	180.	5.34	1.32	.100	1
2	121	33349.	260.	6.	180.	5.34	1.32	.100	1
2	121	63421.	200.	5.	180.	5.34	2.82	.100	3
2	121	27000.	10.	8.	150.	1.66	10.00	9.740	4
2	121	6700.	90.	84.	100.	6.38	.80	.000	1
2	121	336065.	103.	48.	500.	3.21	2.04	.220	3
2	121	78000.	125.	72.	500.	1.47	1.40	.180	1
2	121	12600.	210.	156.	60.	5.14	10.00	.000	4
2	121	1346.	70.	144.	150.	3.43	2.56	.005	1
2	121	365000.	200.	168.	1400.	4.41	3.80	.030	3
2	121	120000.	75.	96.	600.	1.50	3.39	.640	3
2	121	70000.	65.	72.	250.	1.83	3.39	.640	3
2	121	12600.	250.	156.	60.	5.34	10.00	.000	4
2	121	1073600.	220.	120.	3050.	4.76	5.21	.000	3
2	121	4810.	140.	144.	150.	3.15	.60	.005	1

SELECTED MODE	SIC CODE	ANNUAL TONNAGE	TRANSIT TIME	TRANSIT DISTANCE	AVG. SHIPMENT	RATE	ALT. RATE	HANDLING COSTS	ALT. MODE
2	121	12500.	91.	240.	3100.	4.00	.45	.000	1
2	121	12500.	315.	240.	3100.	5.50	1.12	.000	1
2	121	38400.	50.	2.	3200.	9.50	10.00	.500	4
2	121	500000.	130.	5.	1750.	2.50	10.00	1.250	4
2	121	10400.	125.	120.	200.	3.37	2.60	1.450	3
2	121	15600.	60.	168.	300.	5.13	2.57	.000	3
3	121	1161241.	140.	5.	2500.	2.16	3.50	.220	2
3	121	413955.	90.	4.	2500.	2.02	3.57	.220	2
3	121	86650.	145.	84.	2600.	2.69	3.49	.220	2
3	121	404000.	400.	120.	5400.	3.20	5.58	.000	2
3	121	777400.	130.	84.	2800.	2.04	3.49	.220	2
3	121	238691.	147.	48.	2800.	2.45	6.56	.300	2
3	121	330912.	160.	84.	2600.	2.45	6.56	.300	2
3	121	75000.	412.	108.	3600.	4.01	4.49	.300	2
1	131	40000.	835.	420.	165.	4.25	8.55	.020	2
1	131	4375.	240.	120.	365.	20.50	6.21	.000	2
1	131	40000.	955.	540.	165.	4.25	10.05	.020	2
1	131	1710000.	300.	30.	3692.	1.71	9.74	.000	2
1	131	90000.	440.	48.	327.	.36	13.72	.000	2
1	131	30000.	150.	24.	3.	.36	12.97	.000	2
1	281	22000.	180.	120.	1135.	1.11	6.60	.000	2
1	281	75000.	025.	696.	7000.	6.60	18.05	.001	2
1	281	49000.	586.	336.	1500.	5.00	15.00	.130	2
1	281	7200.	150.	48.	600.	1.35	6.20	.000	2
1	281	300000.	109.	456.	9500.	4.50	24.62	.000	2
1	281	34000.	32.	18.	1000.	.03	3.10	.000	2
1	281	36320.	000.	480.	3000.	5.00	22.71	.000	2
1	281	230400.	200.	30.	2000.	.40	3.37	.000	2
1	281	345600.	120.	18.	2000.	.18	5.31	.000	2
1	291	7772305.	115.	48.	6000.	.67	9.24	.000	2
1	291	1181536.	287.	96.	6000.	.81	12.99	.000	2
1	291	172619.	143.	96.	75.	.80	12.99	.000	2
2	291	185000.	60.	24.	350.	9.00	.12	3.000	1
1	331	38000.	25.	72.	800.	1.00	1.50	.350	2
1	331	50400.	195.	60.	700.	1.25	3.95	.650	2
1	331	33600.	150.	60.	1100.	.85	3.35	.650	2
1	331	57000.	45.	72.	600.	1.00	1.56	.350	2
2	331	50000.	200.	156.	100.	5.00	10.00	.650	4
2	331	3500.	375.	96.	60.	5.07	7.04	.500	3
2	331	28348.	230.	96.	240.	5.30	7.04	.500	3
2	331	43200.	500.	144.	6000.	6.56	6.81	.130	3
2	331	28800.	835.	144.	6000.	6.95	8.21	.130	3

MODE
1-BARGE
2-RAIL
3-JOINT
4- NO ALT MODE

SIC CODE
121 COAL
131 PETROLEUM (CRUDE)
281 INDUSTRIAL CHEMICALS
291 PETROLEUM (REFINED)
331 COKE

C. Sample Size

A methodology for determining an adequate sample size, based on Kendall*/, was developed and applied to this project.

a. The variance**/ of a single value was determined and defined by Kendall.

Z_{diff}***/ = .0975

$$VAR\ X = \frac{Z_{diff}}{df}$$

where df is degrees of freedom (in this case 51-2 = 49)

$$VAR\ X = \frac{.0975}{49} = .00199 = (\bar{S}_x)^2$$

where $(\bar{S}_x)^2$ is the variance of the $\hat{Z}$ values of the 51 movement sample.

b. The population variance can be estimated using a method from Ezekiel and Fox.****/

$$S_m = \frac{\bar{S}_x}{n}$$

---

*/ Kendall, M. G., A Course in Multivariate Analysis, Number 2 of Griffin's Statistical Monographs & Courses. Charles Griffin and Company Limited, London, 185 pp. (1957).

**/ Variance is a measure of variation around the average value of a variable, say X, then variance =

$$\frac{\sum_{i=1}^{n} \bar{X} - X_i}{n}$$

where $\bar{X} = \frac{\Sigma X}{n}$

***/ Z_{diff} is the difference in $\hat{Z}$ values based on group means of the respective populations.

****/ Ezekiel, M. and Fox, K. A., Methods of Correlation and Regression Analysis; John Wiley & Sons, New York 548 (1959).

where

S_m = standard error of the population

$\overline{S}_x$ = sample variance

n = sample size

$$S_m = \frac{.00199}{51}$$

S_m = .00628

S_m can be interpreted as the estimate of the standard deviation of the group of averages if similar samples were repeated. It is the standard error of the population mean.

c. A chi-square test can be used to determine the number of observations required based on the ratio of the population variance to sample variance. The ratio is:

$$\frac{.047}{.00628} = 7.5$$

Using this number and going back into a chi-square table (in particular Table V.3 in Beyer's Handbook of tables for Probability and Statistics), the required degrees of freedom can be estimated. In this case 10 degrees of freedom are required, at a 5 percent level of significance along with seven additional degrees of freedom to allow for the characteristics and selected mode, indicating that 17 observations are sufficient to satisfy adequate sample size.

III. RESULTS

A. Commodity Approach

The best results were obtained from the initial runs for 51 coal movements, 28 by barge and 23 by rail.

The group means were as follows:

	Barge	Rail	Diff
Annual Shipment (X1) (100,000 tons)	8.67	1.46	7.21
Distance (miles) (X2)	98.93	142.26	-43.33
Hours travel time (X3)	49.57	93.74	-44.17
Average shipment (X4) (100,000 tons)	.1342	.0088	.125
Rate (X5)	.5321	4.175	- 3.58
Rate advantage over alternative (X6)	- 2.24	.1683	- 2.40
Handling cost (X7)	.136	.6696	- .534

From the information summarized above a linear discriminant function was established as follows: $\hat{Z} = .000226X1 + .000037X2 - .000074X3 + .010233X4 - .012778X5 - .001895X6 - .007627X7$. Using this equation, a test value ($\hat{Z}$) for each movement was calculated. If the value exceeded -.028856, the movement was classified as barge, for lower test values the movement was classified as rail. The coefficient of determination (R^2) for this equation was .745, indicating that 74.5 percent of the variation is explained by the discriminant function. An F-test was made to see if there was a statistically significant difference between the $\hat{Z}$ values calculated at the mean values of each mode. The calculated value was 17.932, which is much greater than the test value of 3.508 (with 7 and 43 degrees of freedom). Therefore, the hypothesis that there is no significant difference (the null hypothesis) between the populations can be rejected.

B. Sensitivity Analysis

For sensitivity analysis, each characteristic was eliminated singly to determine its relative influence. As indicated in Table III-1, only deletion of X_5 (Rate of Selected Mode) appears to significantly affect statistical properties of the estimated discriminant functions. Based on this information the number of variables were reduced to four and a new discriminant function was calculated. The function was:

$$\hat{Z} = -.000053X1 + .014388X2 - .013872X3 - .006522X4$$

where

X_1 = time in hours for the movement
X_2 = average shipment size (scaled by a factor of 100,000)
X_3 = rate of selected mode
X_4 = handling cost

A test value of - .015625 or more indicates barge, a lower test value indicates rail. The R^2 is .724 and the F-test value is 30.12 -- somewhat higher than for the case where 6 or 7 characteristics are included (as indicated in Table III-1).

TABLE III-1
Summary of Sensitivity Analysis--Commodity Approach

Variable Omitted	R^2	F Value	Number Movements Misclassified
0	.745	17.932	4
X_1 Annual Shipment	.743	21.25	4
X_2 Distance	.741	20.95	4
X_3 Travel Time	.733	20.10	4
X_4 Average Shipment	.743	21.22	4
X_5 Rate of Selected Mode	.471	6.54	8
X_6 Rate Advantage over Alternative Mode	.734	20.25	3
X_7 Handling Cost	.696	16.79	4

This suggests, at least for the given data in our sample, that information about rates of movement by alternative mode, annual tonnage and distance of shipment are not statistically significant for the efficiency of the discriminant function. Whether this can be safely generalized is a subject for further testing. However, the collinearity between distance and time is generally recognized, and a similar relationship between annual and average shipments would appear reasonable with respect to modal choice. The estimation of rates for alternative modes is not only time consuming and difficult but the quality and relevance of resulting estimates are often subject to criticism. Therefore, omission of estimated alternative rate is appropriate to obtain effective and efficient procedures for estimating the waterway transportation demand function. Since the discriminant function is based on observed choice of mode, the presence of movements by both rail and waterway modes reflects sufficient information about comparative advantages of each mode to each movement.

Another point of interest is the movements which are misclassified (that is, actual rail movements classified r waterway movements and vice versa). The number of movements misclassified is shown in Table III-1 (35 movements were misclassified in 8 runs where various combinations of characteristics were used). However, of the total misclassified movements the same four are misclassified in each run except when X_6 (rate advantage over alternative mode) is deleted. These four were actual rail shipments but the model classified them as barge shipments. These four movements were unusual in that they moved directly on the east-west rail connections as shown in Figure III-1, are evidently very efficient, and possess a much lower rate than is typical for rail movements in the sample. Rates for these movements are: \$1.47, \$1.50, \$1.83, and \$2.50, compared to the average rail rate of \$4.11. This indicates that the first three movements have rates much closer to the average rate for barge movement (\$.65) than that for rail movement, thereby exhibiting barge characteristis to the discriminant function. The last movement has a rate higher than the first three and its misclassification is due to a combination of factors. This movement exhibits barge characteristics to the discriminant function because it is a large movement and moves very quickly (5 hours as opposed to the rail average of 93.7 hours). In the four variable model this movement is properly classified since annual shipments are omitted in that model.

C. Aggregate Approach

The seven variable model was rerun with all commodities aggregated. Here the observations were sorted only for mode while commodity designation was ignored. The results from this approach were poor, apparently because sample variance tended to increase with aggregation. (See Table III-2).

There were five movements which were consistently misclassified, four of these five movements were the same as those for the commodity approach. The other misclassified movement is a small crude petroleum barge movement (4375 tons per year). This movement exhibited transport times and rates even greater than the average for the rail population.

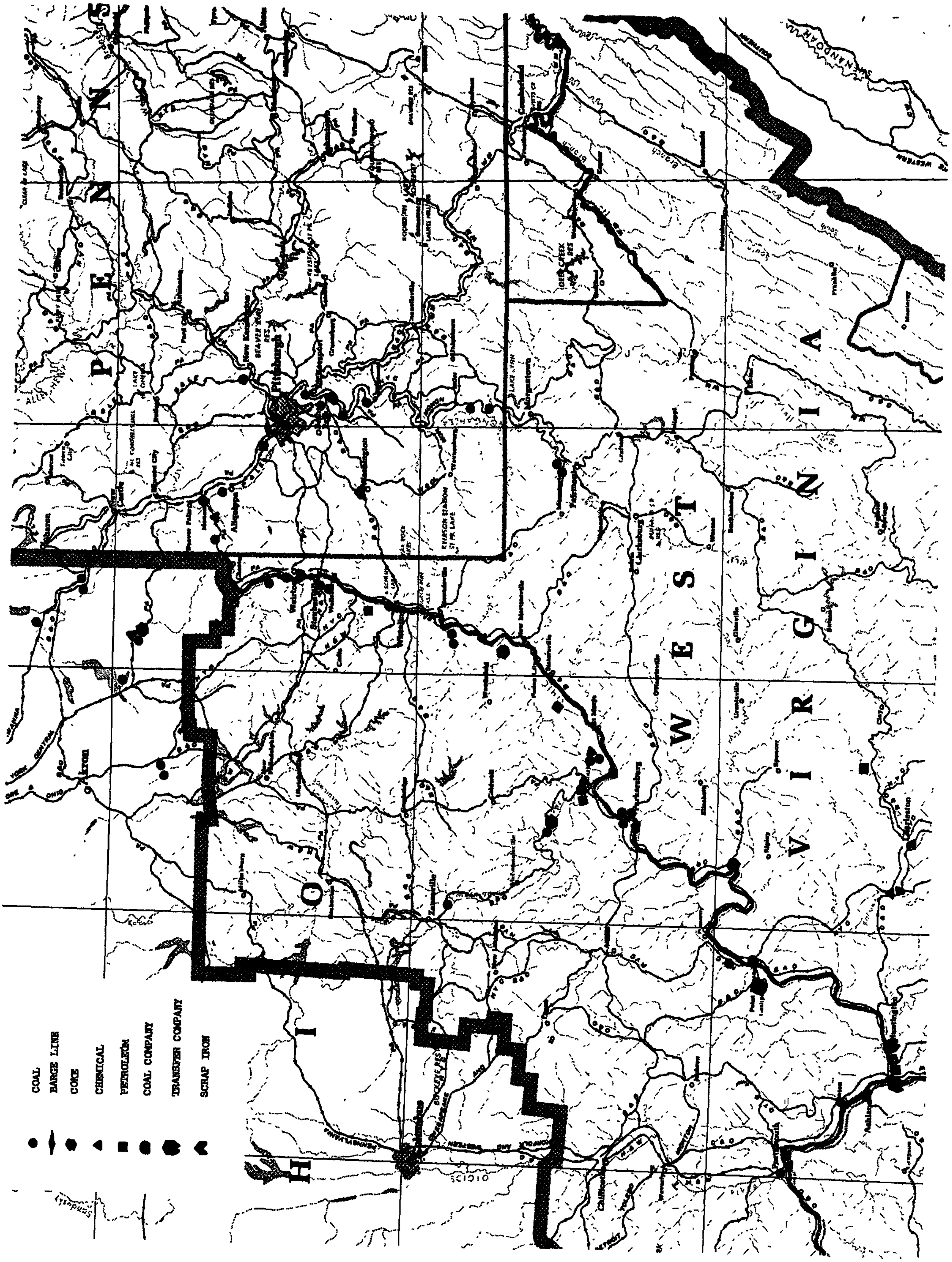


FIGURE III-1 Map of Study Area

TABLE III-2
Summary of Sensitivity Analysis--Aggregate Approach

Variable Omitted	R^2	F Value	Number Movements Misclassified
0	.394	6.60	5
X_1 Annual Shipment	.387	7.59	5
X_2 Distance	.372	7.11	5
X_3 Travel Time	.379	7.34	5
X_4 Average Shipment	.386	7.59	5
X_5 Rate of Selected Mode	.321	5.68	13
X_6 Rate Advantage over Alternative Mode	.387	7.58	5
X_7 Handling Cost	.362	6.80	4

D. Tri-Mode Case

A multi-mode model for barge, rail and joint rail-barge movements was developed largely to demonstrate the feasibility of this type of approach. A certain degree of ambiguity cannot be avoided since the joint mode represents aggregation of extreme values.*/ However, the joint mode has characteristics which are significantly different from the others and for that reason the model was estimated. While only eight observations were obtained for this population, considerably less than the 17 required, it is thought they may still yield significant results because of the homogeneity of the sample data.

The means of the population were as follows:

---

*/ As an alternative procedure, joint movements (which most movements really are if strictly defined, since there are normally some truck hauls from strip or deep mines to the tipple) might be identified by the single mode which dominates the movement and rates and handling costs between modes entered as handling costs.

	Barge	Rail	Joint
X_1 Annual Shipment	8.67	1.46	4.36
X_2 Distance	98.93	142.26	203.0
X_3 Travel Time	49.57	93.74	67.1
X_4 Average Shipment	.134	.009	.031
X_5 Rate	.532	4.115	2.63
X_6 Rate Advantage over Alternative Mode	- 2.24	.168	- 2.03
X_7 Handling Cost	.136	.670	.222

Mahalanobis D^2 = 172.4899*/

The equations for each mode**/ :

	Barge	Rail	Joint
Constant	- 2.093	- 8.448	.00056
X_1 Annual Shipment	.177	.1333	1.7462
X_2 Distance	.0098	- .0011	- 5.084
X_3 Travel Time	.0085	.0249	.1673
X_4 Average Shipment	1.37	- 1.257	.01699
X_5 Rate	.388	3.247	.0058
X_6 Rate Advantage over Alternative Mode	- .381	.00056	- 1.144
X_7 Handling Cost	.127	1.7462	1.7795

---

*/ A test of the significance of the distance between groups, see P. C. Mahalanobis: "On Generalized Distance in Statistics," Proceedings of the National Institute of Science of India, Vol. 12 (1936), pp. 49 ff..

**/ Instead of directly estimating the discriminant function as in the two mode model, equations are developed for each mode and in a second step each movement is classified to a mode by the following definition:

The regions defined by ($\cap$)

$$M_1 \cap r_1f_1 \geqslant r_2f_2, \quad r_1f_1 \geqslant r_3f_3$$

$$M_2 \cap r_2f_2 \geqslant r_3f_3, \quad r_2f_2 \geqslant r_1f_1$$

$$M_3 \cap r_3f_3 \geqslant r_1f_1, \quad r_3f_3 \geqslant r_2f_2$$

There were six movements which were misclassified -- 2 barge and 2 rail movements were classified as joint movements and 2 rail were classified as barge movements.

The first shipment was a small shipment whose alternative rate is less than the selected mode rate. Therefore, this movement would be classified into another mode which exhibits the same characteristics (joint). The second was a small shipment moving a large distance (300 miles) and requiring a great deal of time (168 hours). The alternative rates for each of these were based on rail movement which points out a discrepancy of this approach of using arbitrary assignment.

Two misclassifications have been explained in earlier sections. It may also be noted that the movements lie along one of the main east-west rail lines and available rates are considerably lower than the average for the group.

Two movements have been consistently misclassified through the study. For one movement alternative mode is actually a joint movement but the other movement has no available alternative. This movement will be consistently misclassified because the time required for movement is extremely low. (5 hours actual opposed to an average of 142 hours for rail movements in the sample). your kidding!

E. <u>Demand Inferences</u>

To test the effect of various rate changes on the demand for barge service, the original observed barge rates were increased by a factor ranging from 1.5 to 10.0 times the observed rate values and the adjusted rates were substituted for the observed rates in the discriminant analysis model. New test values were calculated for each movement to determine whether the movement would be classified as a barge movement or not.

The result can be used to construct an aggregate demand curve showing effective rates and corresponding tonnages.

TABLE III-3
Effect of Barge Rate Increases on Shipment Mode

Multiple of Observed Barge Rate	R^2	F Test	Number of Shipments Removed from Barge Group
1.5	.701	14.82	0
1.75	.684	13.33	0
2.00	.660	11.90	1
2.50	.603	9.33	3
4.25	.379	3.75	6
10.00	.389	3.91	25

Critical F Value is 3.508

Table III-4
Points for Demand Curve

Rate	Quantity Shipped Millions of Tons
.65	24.2
1.62	22.8
2.26	21.4
2.66	21.2*
3.19	21.2*
5.32	.53

* Estimates (F-test values were less than the critical 3.508 value, therefore, the statistical confidence tests are not met at the 5% level).

This curve can be interpreted as a demand curve for barge transportation of coal relevant to movements in the Upper Ohio River during 1969. As was expected, the curve is truncated with a portion exhibiting relatively elastic properties and a portion exhibiting relatively inelastic properties. The arc elasticity from point A to B of Figure III-2 is approximately -.1025.*/

---

*/ arc $e = \frac{Q_1 - Q_o}{P_1 - P_o} \times \frac{P_1 + P_o}{Q_1 + Q_o}$

where

$$Q_o = 24{,}200{,}000$$
$$Q_1 = 21{,}445{,}000$$
$$P_1 = \$2.263$$
$$P_o = \$\ .65$$

$$e = \left[\frac{2{,}755{,}000}{1.613} \times \frac{2.913}{48{,}440{,}000}\right]$$

$$e = [1.708 \times .06]$$
$$e = -\ .1025$$

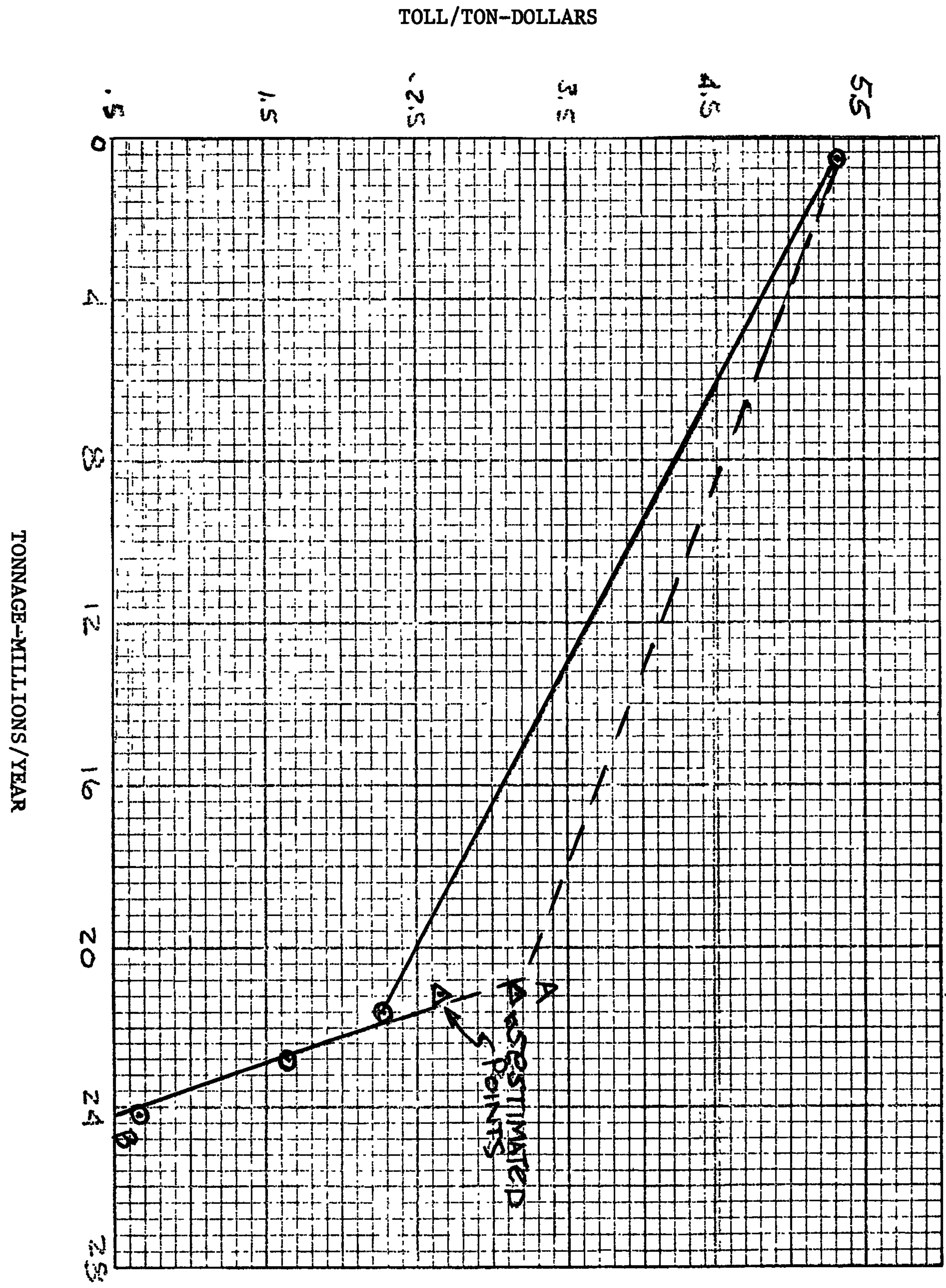


FIGURE III-2
Estimated Aggregate Demand Curve for Barge
Transportation--Upper Ohio Region

IV. EVALUATION OF CONCLUSIONS

The following sections are for the purpose of discussing the validity of and application of the results of this study.

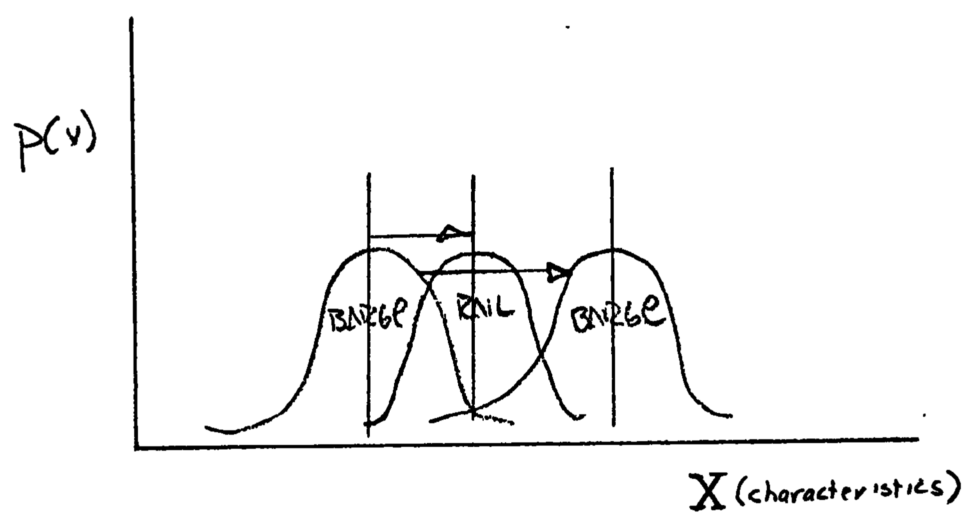
A. Demand Model

The usefulness of the application of discriminant analysis for the derivation of transportation demand has been reasonably well demonstrated for certain commodities in the Upper Ohio Region. The results on coal appear to be reasonable, consistent and sufficient for application to planning studies. It should be noted that coal movements are a sizeable component of the waterway traffic in the Upper Ohio.

Since the analysis is static there is a need for developing procedures for use in projecting future traffic and future demand function. Improved estimates of future traffic division between modes could be expected if independent projections of shifts in rates, time of travel, and technology are provided by mode. The projected values could be analyzed by this discriminant function to determine modal split under projected characteristics of various modes. Much lower statistical confidence could be attributed to such an application as opposed to the use of the discriminant function on current movements, yet in comparison with alternative methods it appears to retain substantial advantages.

Accompanying the application is the legitimate question of whether the method could accommodate the introduction of new modes (such as energy by wire). There is little possibility of using discriminant analysis for this application since the method relies upon observed choice of mode and the characteristics of the mode relative to that choice. Therefore, introduction of a new mode and related characteristics without observed choice would provide no basis for statistical analysis.

The points designated as "estimated" on Figure III-2 are not statistically significant, that is, the F test is below the critical value of 3.508. This phenomena is pictured below. Therefore, we have chosen to draw the demand function, as indicated by the solid line on Figure III-2.



If the area under the demand curve is calculated, resulting benefits reflecting "willingness to pay" for barge transportation for coal in the Upper Ohio River are estimated to be $86.5 million annually.

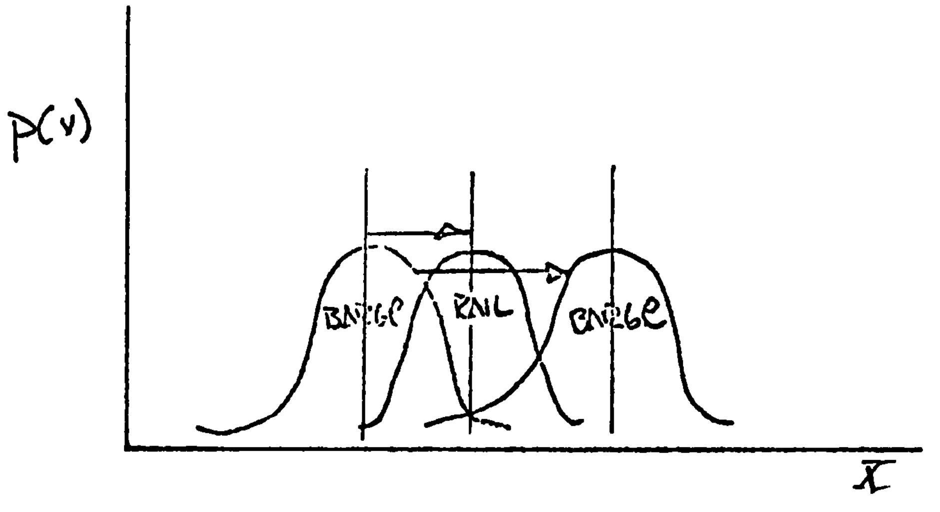
This figure, with some adjustments, can be compared with the conventional method for calculating benefits specified by the

Transportation Act of 1965. The Act specified benefits as the savings to shippers, calculated by the difference in rates between the waterway mode and the next best alternative mode times the quantity of traffic using the waterway. In this case benefits by this criteria would equal $54.2 million annually (24.2 million tons X $2.24--the rate differential), since the method specified by the Act estimates differential savings (net of barge costs). The area under the demand schedule should be reduced by barge costs of $12.9 million (24.2 million tons X $0.5321) to make the estimates comparable. Thus, the comparable estimate is $73.6 million ($86.5 - $12.9) for the area under the demand schedule as opposed to $54.2 million, the savings in shipper costs.

The primary reason for the difference between the net estimates is that the parameters of the demand schedule are influenced by non-rate characteristics which affect modal choice. It should be re-emphasized that the procedure for estimating the price quantity axis of the demand schedule requires that characteristics other than rate are held constant. Therefore, the demand schedule represents the various quantities demanded at all possible prices, other things equal.

B. Statistical Tests

All of the results presented in Section III were statistically significant at the five percent level, with the exception of two points on the demand schedule labeled "estimated." The test for significant is a test of the (null) hypothesis, that the populations (the modes) are not significantly different. The reason was that the barge population shifted from the position on the left in Figure III-3 to a position on the rail population, which produced the statistical effect that the discriminant function could not "discriminate."



The Mahalanbois D^2 test*/ can be used as chi-square (under the assumption of normality) with fourteen (m(g-1)) degrees of freedom, for testing the hypothesis that the mean values are the same in all the g groups for these m variables. The test value for the five percent level of significance is 23.68 and the critical chi-square value is 1.6918. Therefore, the null hypothesis, that the means of each population are not significantly different, can be rejected.

C. Possible Uses

Some of the more obvious uses of discriminant analysis for constructing a model of modal split have been mentioned:

---

*/ A test of the statistical significance of the distance between the means of the modal populations.

(1) To classify movements to modes on the basis of their characteristics.

(2) To estimate the demand function for a given mode of transportation.

On the basis of the data and analysis report in this document, a reasonable case for the utility of discriminant analysis to coal shipments in the Upper Ohio has been developed. There is insufficient information to suggest that the analysis is transferable to other geographical regions or commodities.

Other advantages of the method appear to be in the inclusion in the analysis of characteristics other than rate, such as handling costs, time of shipment and influence of average shipment size, which play a significant role in choice of mode. Discriminant analysis provides a basis for assessing the relative weights given to each characteristic by purchases of transportation services and provides a means for statistically testing the sensitivity of modal choice to one or more of the group of characteristics.

D. Future Work

The need for further effort in the application of discriminant analysis in studies of modal choice has been indicated. A wider geographical sample area and analysis of more commodity groups should be considered.

The conceptual and operational validity of application of discriminant analysis to mixed mode shipments has not been determined.

Whether such movements can be defined in an ambiguous matter is at issue.

When the discriminant function is utilized to estimate the demand function (quantity demanded at various prices) a statistical penalty is suffered since alteration of the rate pushes the model means closer together thereby decreasing the F test ratio.

It is obvious that at some point we could no longer assert that the means of the respective modes are significantly different.

Development of procedures by which the modal split model can be utilized to evaluate alternative futures, which reflect improvements in technology, in pricing practices and shifts in production patterns is an obvious need.

UNCLASSIFIED
Security Classification

DOCUMENT CONTROL DATA - R & D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author) U.S. Army Engineer Institute for Water Resources	2a. REPORT SECURITY CLASSIFICATION Unclassified
	2b. GROUP N/A

3. REPORT TITLE

An Application of Discriminant Analysis to the Division of Traffic Between Transport Modes

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)

Final Report

5. AUTHOR(S) (First name, middle initial, last name)

Lloyd G. Antle
Richard W. Haynes

6. REPORT DATE	7a. TOTAL NO. OF PAGES	7b. NO. OF REFS
May 1971	35	5

8a. CONTRACT OR GRANT NO. N/A	9a. ORIGINATOR'S REPORT NUMBER(S) IWR Report 71-2
b. PROJECT NO. N/A	
c.	9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report) None
d.	

10. DISTRIBUTION STATEMENT

This report has been approved for public release and sale; its distribution is unlimited.

11. SUPPLEMENTARY NOTES None	12. SPONSORING MILITARY ACTIVITY U.S. Army Engineer Institute for Water Resources, 206 N. Washington St., Alexandria, Virginia 22314

13. ABSTRACT

This report analyzes the division of coal traffic between rail and waterway modes in the Upper Ohio Region by means of discriminant analysis. Characteristics which explain choice of mode include annual and average shipment size, transport time and distance, transport rate, handling costs and transport cost by alternative mode. A demand schedule for waterway transportation of coal was estimated on the basis of competitive conditions observed in 1969.

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Demand for waterway transportation Modal split Cost-benefit analysis Discriminant analysis						