University of New Mexico
UNM Digital Repository

Economics ETDs | Electronic Theses and Dissertations

6-1-1954

A Regression Analysis of Airline Cost.

Jesse W. Proctor

Follow this and additional works at: https://digitalrepository.unm.edu/econ_etds

Part of the Economics Commons

Recommended Citation
Proctor, Jesse W.. "A Regression Analysis of Airline Cost.." (1954). https://digitalrepository.unm.edu/econ_etds/80

This Thesis is brought to you for free and open access by the Electronic Theses and Dissertations at UNM Digital Repository. It has been accepted for inclusion in Economics ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact disc@unm.edu.

UNIVERSITY OF NEW MEXICO-UNIVERSITY LIBRARIES
A14429 088133

378.789
Un 3 Op
1954
cop. 2

PROCTOR

A REGRESSION ANALYSIS OF AIRLINE COST

THE LIBRARY
UNIVERSITY OF NEW MEXICO

Call No.
378.789
Un3Op
1954
cop.2

Accession
Number
196844

A14407 614879

DATE DUE

OCT 30 '86

UNM OCT 28 '86

DEMCO 38-297

UNIVERSITY OF NEW MEXICO LIBRARY

MANUSCRIPT THESES

Unpublished theses submitted for the Master's and Doctor's degrees and deposited in the University of New Mexico Library are open for inspection, but are to be used only with due regard to the rights of the authors. Bibliographical references may be noted, but passages may be copied only with the permission of the authors, and proper credit must be given in subsequent written or published work. Extensive copying or publication of the thesis in whole or in part requires also the consent of the Dean of the Graduate School of the University of New Mexico.

This thesis by Jesse W. Proctor
has been used by the following persons, whose signatures attest their acceptance of the above restrictions.

A Library which borrows this thesis for use by its patrons is expected to secure the signature of each user.

NAME AND ADDRESS	DATE
Roxanne Nai 322 Hokona Zuni-UNM	Sept 30, 1986

A REGRESSION ANALYSIS OF AIRLINE COST

UNIVERSITY OF NEW MEXICO ALBUQUERQUE

By

Jesse W. Proctor

A Thesis

In Partial fulfillment of the

Requirements for the Degree of

Master of Arts in Economics

The University of New Mexico
1954

This thesis, directed and approved by the candidate's committee, has been accepted by the Graduate Committee of the University of New Mexico in partial fulfillment of the requirements for the degree of

MASTER OF ARTS

E. H. Castetter
DEAN

6/1/1954
DATE

Thesis committee

Julian S. Duncan
CHAIRMAN

David Hamilton

Mervyn Cadwallader

378.789
Un3Op
1954
cop. 2

CONTENTS

CHAPTER		PAGE
I.	FACTORS WHICH DETERMINE COST	1
	The controversy	1
	Issues	2
	Seven important variables	6
	Objectives	20
II.	DATA AND METHOD	23
	General nature of data	23
	Introduction to specific data	27
	Statistical theory	36
	Regression technique	40
III.	COMPUTATIONS	46
	Seven independent variables	46
	Three independent variables	61
IV.	LENGTH OF THE FLIGHT	71
	Accounting	71
	Conditions of the industry	79
	The cost study by American Airlines	89
V.	CONCLUSIONS	96
	Further research	96
	Theory	102
	Public policy	113

196844

I. FACTORS WHICH DETERMINE COST ---------- 1

The controversy ---------- 1

Issues ---------- 2

Seven important variables ---------- 6

Objectives ---------- 20

II. DATA AND METHOD ---------- 29

General nature of data ---------- 29

Introduction to specific data ---------- 37

Statistical theory ---------- 36

Regression technique ---------- 40

III. CONJUNCTIONS ---------- 46

Seven independent variables ---------- 46

Three independent variables ---------- 61

IV. LENGTH OF RUN FACTOR ---------- 71

Accounting ---------- 71

Conditions of the industry ---------- 79

The cost study by American Airlines ---------- 89

V. CONCLUSIONS ---------- 96

Further research ---------- 96

Theory ---------- 102

Public policy ---------- 113

BIBLIOGRAPHY - 120

MATHEMATICAL APPENDIX - - - - - - - - - - - - - - - - - - - 123

Vocabulary - 124

Symbols - 126

Fundamental Concepts - - - - - - - - - - - - - - - - - 133

Multiple correlation - - - - - - - - - - - - - - - - - 136

Residual deviance - - - - - - - - - - - - - - - - - - - 139

Omission of a variable - - - - - - - - - - - - - - - - 141

Partial correlation - - - - - - - - - - - - - - - - - - 142

The Q ratio - 145

The F ratio - 147

A confidence test for regression coefficients - - - 149

Significance of the t test - - - - - - - - - - - - - 153

Interpretation - 156

Calculation - 158

SCATTER DIAGRAMS

DIAGRAM		PAGE
I.	Operating Expense and Load Factor	48
II.	Operating Expense and Capacity of Planes	49
III.	Operating Expense and Daily Flight Time per Aircraft	50
IV.	Cost per Station and Length of Flight	85

DIAGRAM PAGE

I. Operating Expense and Load Factor - - - - - - - - - - - - 49

II. Operating Expense and Capacity of Plane - - - - - - - 49

III. Operating Expense and Daily Flight Time per Aircraft - 50

IV. Cost per Station and Length of Flight - - - - - - - - - 85

TABLES

TABLE		PAGE
I.	Distribution of Passenger Expenses and Profit by Length of Segment and Type of Service, April 1952 Expenses at 65 per cent Load Factor	17
II.	Operating and Cost Data	28
III.	Ton-mile Load Factor and Available Tons per Aircraft Mile	29
IV.	Assets employed	30
V.	Calculation of Estimated Linear Equation	55
VI.	Statistics on Three Variables	62
VII.	Selected Comparisons of Estimated and Observed Costs per Revenue Ton-mile	67
VIII.	Estimated Costs at High Load Factors	69
IX.	Cost of Tons Lifted for Feeders and Trunks	88
X.	Relationship of Operating Cost per Ton-mile to Length of Flight	93

TABLES

I. Distribution of Passenger Revenue and Profit
by Length of Segment and Type of Service,
April 1952 Expenses at 65 per cent Load
Factor - - -

CHAPTER I

FACTORS WHICH DETERMINE COST

The controversy. This thesis is an attempt to determine what factors are responsible for airline cost.[1] The question is important both in the formation of airline route structure and in the determination of fares. Some authorities maintain that the load factor is almost the sole determinant of airline costs. Others hold that the length of the flight[2] is extremely important. The capacity of equipment, speed of the plane, and extent to which equipment is utilized have also been

---

1 The word "cost" as used here refers to unit cost. This can mean cost per plane-mile, cost per available mile, or cost per revenue ton-mile.

---

2 "Length of the flight" refers to the average non-stop flight that is made by each plane. This is almost, but not quite, synonymous with the distance between stations. It is generally true that as the length of the interstation flight is increased, the average non-stop flight of each plane will increase. The term "length of the haul" refers to the distance cargo moves. This distance will usually be two or three times the length of the flight. There is some tendency to use these terms interchangeably in spite of the fact that they refer to different distances. This tendency probably rises from the fact that the length of the flight is highly correlated with the length of the haul. An increase in the length of the flight will usually mean a corresponding increase in the length of the haul. The term "flight stage" is, for practical purposes, synonymous with length of the flight. Flight stage is obtained by dividing the average number of miles a carrier flies during a given year by the average number of flight departures. The term "trip distance" is also used instead of length of the flight.

---

considered important. Each of these factors affects airline route structure and some of them affect rate structure. This chapter will briefly review this controversy and show its significance.

I. ISSUES

Outline of issues. Koontz[3] has indicated that the

---

3 Harold D. Koontz, "Domestic Air Line Self-sufficiency", The American Economic Review, Volume LIX, No. 1, March, 1952, p. 122

---

following issues are paramount in a discussion of airline economics: (1) Should the present practice of subsidizing airlines be continued? (2) Should high density routes be opened to weaker carriers?[4] (3) Should the domestic airline

---

4 No person may engage in air transportation unless he obtains from the Civil Aeronautics Board a "certificate of convenience and necessity". For a discussion of this certificate, see Joseph L. Nicholson, Air Transportation Management (New York: John Wiley & Sons, 1951), p. 30. This certificate must be obtained to operate either scheduled or non-scheduled flights. Such certificates are often difficult to obtain. The applicant must appear before the Civil Aeronautics Board and show reason why he should be granted a certificate. Usually companies which are already operating in the area will be permitted to show why no more air transportation should be permitted in the area. As an example of this, see Civil Aeronautics Board, General Passenger Fare Investigation, Exhibits of American Airlines, Inc., Docket No. 5509 (Washington, D. C., April 1, 1953.) These certificates specify the routes over which a company is permitted to operate. The question of whether or not a

a carrier will be certificated for a given area depends on the public policy which the Civil Aeronautics Board is following at the time the application is made.

---

structure be radically and rapidly redrawn? (4) Should a merger of certain carriers be effected so as to slowly redraw the airline route structure. To these may be added another important issue. Is the tapering fare justified?

Subsidy. It will be noted that some airlines are subsidized more heavily than others. This results from the fact that the Civil Aeronautics Board attempts to provide such subsidies as are necessary to keep an airline in operation.[5]

---

5 Nicholson, op. cit., pp. 180 and 226

To keep a line in operations means principally to pay all operating costs and provide a fair return to investors. Hence the principal determinants of subsidy are the factors which determine cost.

Competition. Consider the question of opening high-traffic density routes to non-scheduled carriers.[6] American

---

6 The term "non-scheduled carrier" means a carrier who has no regularly scheduled time for arrival and departure. For a discussion of the non-scheduled carrier, see Nicholson, op. cit., p. 32

Airlines uses the significance of the length of the haul to indicate the inadvisability of permitting irregular carriers to usurp the markets of the presently certificated carriers. The length of the flight is one of the principal reasons the length of the haul is significant, in the analysis of American Airlines. That is, American has argued that the longer hauls mean longer flights, and that a substantial reduction in unit cost results from these longer flights. American has argued that non-certificated carriers are attempting to secure the long-haul, low-cost routes. This question will be discussed further after the principal issues have been outlined.

Reformation or merger. In either the reformation of airline structure or in the merger of existing structures, it seems likely that cost-determining factors will play a large part. When either of these changes is contemplated, it will be desirable to analyse the costs of all segments of the structure that is being formed in order that the outcome of the reorganization may be evaluated. As an example, suppose a proposal to merge two lines and eliminate low traffic density stops is under consideration. The elimination of low density stops will often mean that the length of the flight will be increased. The question is, will a substantial reduction in costs result from such a move? This question can be answered only if the effects of an increase in the length of the flight can be precisely evaluated.

Tapering charges. The principle of tapering transportation charge is a method of charging under which the unit cost per mile to the shipper decreases as the length of the haul increases. The principle of tapering fare is an application of this to the cost of passenger tickets. American Airlines has advocated this method of charging and it has received some recognition by the Civil Aeronautics Board. Nicholson[7] cites a table which shows the manner in which

---

7 Ibid., p. 404

---

air carrier freight rates decrease as the length of the cargo haul increases. This table shows that at 400 miles the freight rate (per unit mile) is only 91 per cent of the rate for 200 miles. At 600 miles, the freight rate is only 84 per cent of the rate for 200 miles. Nicholson argues that this cost differential is inadequate. The airline freight diffenential is one-third less than a comparable differential charged by the railroads. This differential is usually justified on the ground that it reflects operating costs.[8] Reduction in unit

---

8 Ibid., p. 190

---

costs due to an increase in the length of the flight is

thought to be associated with reduction in unit cost due to the length of the haul. A significant part of the reduction in cost that purportedly comes from an increase in the length of the haul is assumed to be the result of an increase in the length of the flight. This relationship will be explained shortly. At this point it is sufficient to note that an investigation of the length of the flight is partially an investigation of the principles underlying the tapering charge.

II. SEVEN IMPORTANT VARIABLES

Importance of cost factors. In each of the issues previously discussed, cost plays an important part. The purpose of this thesis is not to resolve each of these issues, but rather to determine the importance of certain cost factors. The purpose of this study is to determine the facts. A brief attempt will be made to show the meaning of these cost factors in terms of the issues involved, but this attempt is not intended to be in any manner definitive. The factors which will be considered here are the following: load factor,[9] size of the firm, capacity of the plane, length of the flight, speed

---

9 The "load factor" is the average percentage of available load that is carried. The revenue passenger load factor is obtained by dividing the total available passenger miles by the number of revenue passenger miles carried. The quotient is multiplied by 100 to avoid use of the decimal. The revenue ton-mile load factor is

obtained by dividing the total number of available ton miles, for a given period, by the number of ton miles carried for revenue producing purposes. A load factor of 100 is the maximum load possible with either of these measures.

---

of the plane, plane utilization, and metropolitan population served. Each of these variables has at one time or another been considered important as an instrument for explaining the economics of airline cost.

Load factor. In one of the most brilliant and authoritative statements available on Airline Economics, Harold D. Koontz argues that the route structure is the most important element in cost determination.

> The differentiating factor appears not to be competition, but rather the size and density of the traffic generating routes which they may serve. This is a matter of route structure, which in turn is a matter of government policy as exercised by the Civil Aeronautics Board. Given a route structure like that enjoyed by American or United, with a good number of high traffic density pairs of cities on it, and with little more than half of the traffic from among the lower density route segments, it appears that an air line may become financially self-sufficient if the air line has a fair degree of cost control and a reasonably effective management.[10]

---

10 Koontz, op. cit., p. 121

The phrase "size and density of traffic generating routes"

denotes routes which produce high load factors. The load factor is the determining element in the Koontz analysis.[11]

---

11 Julian S. Duncan, Introduction to Transport Economics, (Albuquerque: Department of Economics, University of New Mexico, 1953), p. XXIV-13.

Capacity of plane. Along with load factor, Koontz admits that the capacity or size of the equipment plays an important part in determining unit cost. Attempts have been made to construct efficient, low-cost planes of 15-to 20-passenger capacity. It has not been possible to construct such planes that can compete with larger planes on a unit cost basis. The reason for this is that it costs almost as much to build a 20-passenger plane as it does a 40-passenger plane. Also, the crew cost for a small plane is practically as great as the crew cost for a large plane.[12]

---

12 Harold D. Koontz, "Air Line Self-Sufficiency: Rejoinder", The American Economic Review, Volume XLIII, Number 9, June, 1953, p. 375

Size of the firm. Economic theory predicts the way in which unit costs should behave as the size of the firm increases. In the initial stages, costs are expected to decline as the size of the firm increases. Eventually a minimum cost is reached. After this, the unit cost increases

as the size of the firm increases.[13] This raises two important

---

13 John F. Due, Intermediate Economic Analysis (Chicago: Richard D. Irwin, 1950), p. 202

---

questions with respect to the airline industry. (1) Does the size of the firm play an important role in determining airline cost? (2) At what size of the firm is the minimum cost obtained? Koontz gives an answer of "no" to the first question. That is, he finds that the size of the firm does not play an important part in cost determination.

> Thus for most of the air lines, the cost experience of 1949 does not support a strong case for the existence of important economies of size over a wide range of plant sizes. Only the four very small air lines appear clearly to be at a serious cost disadvantage on account of their size. . . . Records in other recent years would not show essentially different results.[14]

---

14 Harold D. Koontz, "Domestic Air Line Self-Sufficiency", The American Economic Review, Volume XLII, Number 1, March, 1952, p. 114

As indicated by the quoted paragraph, Koontz has answered the second question. Since the four largest firms have no cost advantage, it follows that none of them have yet increased in size to where they have passed the point of minimum cost.

The point that Koontz makes here may have more significance than Koontz has indicated. In the first place,

the denial that the size of the firm plays an important part in the airline industry is essentially a refutation of a long-established principle of classical economics. Koontz appears to have hinted that the size of the firm is unimportant only because the diseconomies of large size have not yet been attained. This avoids the implication that classical economic theory is directly challenged. However, the assertation that load factor is the important determinant in airline cost could have been made only where it was shown that the size of the firm and competition were ruled out. This is equivalent to saying that the airline industry is an important exception to the well-founded principles of classical economics. The reader can carry this assertion one step further. He can well reason that if the airline industry is one important exception, perhaps there are others, and if there are others in sufficient number, perhaps the principle which defined the behavior of long-run average cost is in fact little more than a myth.

Length of the flight. Koontz explains why the length of the flight is significant in airline economics.

> Heavy travel brings favorable load factors, particularly when aircraft can be scheduled non-stop between the high traffic generating pairs of cities. For if the volume of traffic (and the route structure provided by the Civil Aeronautics Board) permits adequate schedules on a non-stop basis, a carrier's load factor is not held down by the inevitable loss of revenue space when passengers must be accomo-

> dated en route. It is a difficult enough job to schedule seats to meet, at reasonably good load factors, the variations in passenger demands by days of the week and times of the day. But when, in addition, the variations occasioned by trying to piece passenger trips together on a multi-stop schedule are added to the variations of time, the task of giving service and of maintaining a profitable load factor is difficult indeed. Just as many of the 50 to 100 pairs of cities lend themselves to high volume, non-stop scheduling, most of the 15,000 pairs of cities require many-stop scheduling between points of low traffic density.[15]

---

15 Ibid., p. 120

The important feature of this explanation is that the length of the flight is significant because it affects the load factor.

In addition, Koontz attributed a certain significance to the length of the flight independent of load factor. This aspect of the length of the flight did not receive strong emphasis. It was employed incidentally to explain why the negative correlation between cost and the size of the firm is significant.

> In the first place, the high negative rank correlation may be traced to a few expense classifications. In the case of total direct operating expenses, the correlation is due to the significant inverse relationship in flying operations expense (the expense involved in actual airplane operations). The relationship is largely coincidental, since the larger air lines generally have routes which make it possible to schedule a higher average trip distance, and this factor is not necessarily a characteristic of size. Most flying expenses tend to vary with the hours of airplane operation. Since the number of miles an airplane travels per hour varies sharply with trip distance, especially for distances under 250 miles,

> one would expect direct flying costs per available ton mile to be lower the longer the trip distance.[16]

---

16 Ibid., p. 111

The flying operations expense mentioned above constitutes approximately one-third of the total operating expenses of an airplane. The change in flying operations expense associated with an increase in the size of the firm does not appear to be great in magnitude. The length of the flight is treated almost casually here. Koontz apparently does not attribute great significance to it.

American Airlines has a somewhat different conception of the length of the flight. The following quotation presents this position.

> Operating costs per unit of traffic generally decrease as length of the haul becomes greater. The cost per passenger mile for carrying a passenger from New York to Boston is much greater than the per passenger mile cost of carrying a passenger from New York to Los Angeles.
>
> This tapering downward of costs as the length of the haul becomes longer results from the basic characteristics of the individual items making up the overall cost. Fundamentally, the overall costs can be divided into three basic categories.
>
> The first group of costs is the direct flight cost and related items. Included in this grouping are the pay of pilots, cost of fuel, maintenance costs, and depreciation of the aircraft. Our studies indicate that these costs are approximately the same for each ramp-to-ramp hour. In other words, the

cost per ramp-to-ramp hour for a flight from New York to Chicago is about the same as on a flight from New York to Boston. Since a smaller percentage of the ramp-to-ramp time is consumed in taking off and landing on the New York to Chicago flight, the average speed is greater. The cost per plane mile of operating on the longer flights are lower, therefore, since approximately the same costs per hour are spread over a greater number of miles. This downward trend in flight costs as the length of the flight increases is conservative in our opinion. It is possible that the aircraft is subjected to more wear and tear on take-off and landing than it is during cruising. If this were true, of course, the differential between the costs for short and long flights would be greater.

Apart from indirect maintenance, which is a function of direct maintenance and therefore has the same basic characteristics, ground and indirect costs divide into two general groupings.

One group of these costs is constant for flights of all distances. Included in this grouping are costs like reservations and ticketing and cargo handling that are the same per passenger or ton of cargo loaded no matter what the distance of carriage. For example, the salaries of reservations and ticketing personnel in April 1952 on American's system amounted to $1.16 per passenger. This cost when divided by the mileage for a trip from New York to Boston results in a cost per passenger mile of .63 cents, while for the longer trip from New York to Chicago the cost per passenger mile is .16 cents, or 75% less. When these costs are plotted according to the length of the haul, the resulting tapering curve is quite similar to that resulting from plotting the direct flight costs.

Other ground and indirect costs do not taper downward as the length of the haul is increased. They are about the same per revenue passenger mile, regardless of the length of the flight. For example, the passenger service costs and passenger liability insurance are principally a function of passenger miles. In April 1952 the passenger costs of meals and other supplies were .194 cents per passenger mile no matter what the length of the flight. Since these costs do not taper and when plotted according to distance result in a straight line, they modify to some extent the taper resulting from plotting all of the other costs.

> If all three of these groupings of costs are plotted, the operating cost per unit of traffic for any given length of flight can be calculated by adding together the costs shown on the various bases listed above. Because roughly 70% to 80% of the costs taper downward as the length of the flight increases, this operating cost per unit will have a definite downward trend. The taper will be especially sharp for flights under 400 miles and will then gradually level-off.[17]

---

17 Civil Aeronautics Board, op. cit., p. 82

This quotation does not seem to clearly distinguish between the terms "length of the flight" and "length of the haul". The two terms seem to be used synonymously in this quotation. The quotation indicates that the first group of costs, direct flight costs and related items, depends upon the length of the flight, not the length of the haul. The second group of costs are what may be described as terminal costs. They apparently depend upon the length of the haul, not upon the length of the flight. The above quotation seems to suggest that both the length of the haul and the length of the flight are important determinants of cost. Unfortunately the quotation did not indicate the relative importance of each variable. It seems obvious that correlation between length of the haul and the length of the flight is high. It is generally true that the longer the length of haul, the longer will be the length of the flight. For this reason, it may be assumed that the length of the haul and the length of the flight are roughly the same variable.

In the Transcontinental Coach Type Service Case, American argued that the length of the haul is a most significant factor in determining cost. Here non-certificated airlines were applying to the Civil Aeronautics Board for permission to operate coach services in competition with American Airlines and other certificated carriers. American Airlines stated:

> . . . the whole pattern of service developed by the applicants and others in their so-called irregular services has been preoccupied with these long-haul services. . . . The applicants are quite frank as to their intentions. It is the long haul traffic that they want, and only that traffic. . . . The amount of this important long-haul traffic which would be taken from the present carriers by certification of even one of the applicants is impressive.[18]

---

18 Civil Aeronautics Board, Transcontinental Coach-Type Service Case, Brief of American Airlines, Inc., Docket Nos. 3397 et al. (Washington, D. C., April 20, 1950, p. 14

The above argument made the length of the haul an issue in airline competition. The length of the haul had previously been accepted as the factor responsible for the tapering cost principle. American's argument that the length of the flight and speed of the plane are important in connection with the length of the haul has important implications. This argument implies that both the length of the flight and speed of the plane are important determinants of airline cost. If the length of the flight and speed of the plane were not important determinants of airline cost, there would be no reason for considering them in connection

with length of the haul. Since length of the haul is considered important in connection with the tapering cost principle as well as in airline competition, it follows that length of the flight and speed of the plane are also important for these reasons.

Table I was extracted from American's exhibit in the General passenger fare investigation. This table shows the way in which expenses and profit per passenger mile decrease as the length of the haul increases. This table is based on a load factor of 65%. Higher load factors would decrease each of the costs shown here and lower load factors would increase these costs. According to this table, both the DC-6 and the Convair must travel a distance of approximately 300 miles before costs begin to level off. If these planes are used at less than that distance, the cost per passenger mile increases rapidly as the distance decreases. Although there appears to be significant decreases in cost beyond the 300 mile segment, large increases in distance produce only small decreases in cost.

Speed of the plane. The discussion by American Airlines of the speed of the plane aids in the definition of this variable. In most airline statistics, the speed of the plane will be obtained by dividing the total distance traveled by the travel time. A measure obtained in this manner will depend upon the time required for take-off and landing. It will also depend upon the speed with which the plane travels after it is in the air. This means that any discussion of the speed of the plane should encompass take-off and landing time as well as the mechanical

TABLE I

DISTRIBUTION OF PASSENGER EXPENSES AND PROFIT
BY LENGTH OF SEGMENT AND TYPE OF SERVICE
APRIL 1952 EXPENSES AT A 65 PERCENT LOAD FACTOR

DC-6 Tourist		Convair	
Length of Segment	Expenses and Profit per Passenger Mile	Length of Segment	Expenses and Profit per Passenger Mile
		49	22.5
		81	15.8
109	10.2	99	14.1
120	9.6	120	13.0
		143	12.0
184	7.7	168	11.0
250	6.4	215	10.1
303	5.6	303	8.8
423	5.2	436	7.9
511	4.9	511	8.1
724	4.5	730	7.9
1245	4.0		
1751	3.8		
2475*	3.8		

* New York-Chicago-Los Angeles

Source: Civil Aeronautics Board, General Passenger Fare Investigation, Exhibits of American Airlines, Inc., Docket No. 5509, April 1, 1953, pp. 95-96

performance of the plane.

As previously mentioned, American's argument implies that the speed of the plane has an important influence on cost. It is particularly interesting to note that Koontz did not emphasize the speed of the plane.

Plane utilization. Another writer on airline economics develops his theory without reference to either speed of the plane or length of the flight. An article by Dr. Charles R. Cherington appears in a senate hearing. Cherington emphasizes load factor and plane utilization as the two determining factors in airline economics.

> Thirdly, while carrier prosperity is in the long run dependent on high load factors and the fullest possible utilization of available aircraft, there has not been nor is there likely to be, in the future a direct correlation between the growth curve and carrier prosperity.[19]

---

19 Charles R. Cherington, "The Essential Role of Large Irregular Air Carriers", Future of Irregular Airlines in United States Air Transportation Industry, Hearings Before a Subcommittee of the Select Committee on Small Business, U. S. Senate, March 31 thru May 8, 1953 (Washington, Government Printing Office, 1953), p. 578

In the article from which the above quotation was taken, the author argues that unscheduled carriers maintain a supply of planes which they are free to devote to irregular movements of persons or property, whereas the regular carriers have only a limited amount of equipment available for irregular movements.

performance of the plane.

As previously mentioned, Doerfler's argument implies that the speed of the plane has an important influence on cost. It is particularly interesting to note that Doerfler did not emphasize the speed of the plane.

Plane utilization. Another writer on airline economics develops his theory without reference to either speed of the plane or length of the flight. An article by Chamberlin in Chamberlin appears in a series heading. Chamberlin emphasizes load factor and plane utilization as the two determining factors in airline economics.

Thirdly, while carrier prosperity is in the long run dependent on high load factors and the fullest possible utilization of available aircraft, there has not been any thing likely to be, in the future a direct correlation between the growth curve and carrier prosperity.[13]

---

[13] Charles K. Chamberlin, "The Financial Role of Large Irregular Air Carriers," Future of Irregular Airlines in United States Air Transportation (Hearings, Hearings Before a Subcommittee of the Select Committee on Small Business, U. S. Senate, Seventy-Eighth Congress, 1st Session, pursuant to S. Res. 55, May 3, 1943 (Washington, Government Printing Office, 1943), p. 270.

In the article from which the above quotation was taken, the author argues that unscheduled operators maintain a number of planes which they are free to devote to transport movements of persons or property, whereas the irregular operators have only a limited amount of equipment available for irregular movement.

A significant part of Cherington's argument is his emphasis on aircraft utilization. It is clear from the context of Cherington's argument that aircraft utilization is not the same as load factor. The fact that he mentions both load factor and plane utilization in the above quoted paragraph is also evidence that load factor is separate from plane utilization. In fact, it is clear from Cherington's discussion that plane utilization refers to the extent to which planes are in continuous use. This might suggest, as a convenient measure of plane utilization, the average number of hours per day that a plane is in use. Cherington finds that there are periods during which the regular carriers are loaded to capacity and unable to carry the traffic offered. During such periods it seems likely that companies which have routes which permit the maximum use of their equipment would be at a distinct disadvantage. In any event, if carrier prosperity is dependent on the fullest possible utilization of available aircraft, it is to be expected that companies which utilize their planes most effectively will show lower unit cost.

Metropolitan population served. Koontz[20] finds that a few

---

20 Koontz, op. cit., pp. 118-120

---

cities are responsible for a large percentage of the air traffic. There is a distinct tendency for the larger cities to show the

largest amount of air traffic. It seems plausible that a decrease in population served would produce a decrease in traffic, and that this would result in lower load factors and higher operating costs. Since 200 cities are responsible for the bulk of air traffic, an ideal situation would be a statistical measurement based directly on these cities. Such a measurement is cumbersome, however, and it seems worth while to investigate the possibility that a measurement of metropolitan population served by each airline will give the same results.

III OBJECTIVES

A choice of factors is necessary. A number of different theories on airline economics have been discussed. Each theory seems to be predicated upon the assumption that one or more factors is responsible for airline cost (or prosperity). Yet, there is considerable difficulty in forming an synthesis of these theories, due to inconsistencies or apparent contradictions. It appears that no writer has presented a complete and satisfactory version of the factors which determine airline cost. Each writer seems to have developed his theory about one or two elements and to have ignored the remaining factors. It may be suggested that the reason for this situation is the difficulty of determining the most important factors. This thesis describes an attempt to determine the most important factors.

A statistical approach. In the case of airline economics,

it cannot be said that the chief difficulty with theorists is that they are dominantly a priori. The study by American Airlines, as an example, is almost as empirical as a study can be. The conclusions of Koontz appear to be based on empirical considerations. The principal criticism that can be made of the American study is methodological. The study is based on one firm and it fails to weigh the significance of one variable against that of another. A dominantly statistical approach to a cross section of the airline industry should provide a valuable check on the study by American Airlines. Koontz has used the cross sectional approach and come up with an answer somewhat different from that of American Airlines. It is difficult to be sure that Koontz investigated whether or not the length of the flight has a significance independent of load factor.

A cost formula. Koontz did not attempt to construct a formula for predicting cost based on what he found to be the most important cost factors. It is believed that a formula based on cost factors would serve the following purposes: (1) it would enable the statistician to state in precise form the relative importance of each cost factor; (2) it would enable the statistician to determine confidence limits for each cost factor; and (3) it would facilitate accurate prediction of cost based on each of the factors selected. It is easy to see that the last-named purpose would make the meaning of each cost-determining factor considerably clearer than it would otherwise be.

When the question of airline merger was previously introduced, the need for a method of evaluating the effect of a reduced non-stop flight was indicated. The need for evaluating the effect of a reduced load factor was also indicated. It is believed that a formula of the nature described above would serve these purposes.

Scope of study. The problem connected with airline cost has been briefly stated. The reasons for a statistical study have been outlined. The theoretical value of a cost formula has been suggested. In order to carry out this program, data must be collected, the data must be manipulated so as to obtain sound statistical results, and these results must be evaluated. In pursuance of this program, Chapter II will outline the data selected and the methods used to extract meaning from that data. Subsequent chapters will present the computations in detail and an attempt to interpret these computations.

CHAPTER II

DATA AND METHOD

Chapter I outlined the problem of determining what factors are responsible for airline cost. It suggested that a statistical approach would help to clarify the problem and it indicated that the following factors are responsible for airline cost: Load factor; capacity of the plane; size of the firm; length of the flight; speed of the plane; plane utilization; and population served. This chapter will present data on each of the above factors. A short sketch of the theory of regression analysis will follow. Finally, a general description will be given of some of the techniques which were employed in this study.

I. GENERAL NATURE OF DATA

Choice of variables. As indicated in Chapter I, the overall purpose of this study is to determine the relationship between cost and other factors. This end may best be accomplished by collecting a series of unit cost observations and analysing the relationship between unit cost and the various factors which are believed to be responsible for changes in cost. Unit cost will become the independent variable and each of the seven cost determinants will become a dependent variable. It appears from the evidence gathered in this

study that these seven variables explain cost reasonably well. It is possible, however, that some unknown factor not considered here is primarily responsible for changes in cost and that the correlation between these seven variables and cost is merely the result of changes in that unknown factor. The choice of variables is the most critical part of a study such as this and it is the aspect about which the least is known.

Cross sectional data. As previously mentioned, this study was based on a cross section of the airline industry. This means that a number of firms are investigated during a given period. In general, this is equivalent to holding time constant. Klein[1] calls this the investigation of inter-

---

1 Lawrence R. Klein, Econometrics (Evanston, Illinois: Row, Peterson and Company, 1953), p. 185

---

firm differences. He classifies the unit observations as microeconomic and the variables as macroeconomic. A cross sectional study is primarily a study of the differences in conditions throughout an industry. It is a study of how conditions in Firm A differ from conditions in Firm B, how Firm B differs from C, etc. Each firm represents an experiment in operating under certain conditions. The object of this thesis is to formulate generalizations about these conditions.

It is always possible that special conditions prevail in one firm that are not applicable to the entire industry. As an example, one firm might use one type of plane only. An investigation of this firm would not show the advantage of using a large plane instead of small planes. The cross sectional approach is the only approach by which generalizations can be made about the size of the firm. The cross sectional method has the advantage of eliminating variations in cost due to differences in cost-accounting methods or managerial policies. Such variations will tend to cancel each other and leave only variations due to the variables under investigation. Another advantage of this method is that it avoids errors due to the simultaneous occurrence of certain conditions in the industry. For example, an investigation of one firm might not reveal that the cost advantage of that firm were due to a certain combination of length of the flight and size of the firm.

Number of observations. The larger the number of observations, the less is the likelihood of error in making predictions. From this standpoint, it would have been desirable to use at least fifty observations. However, in a study of this nature it is necessary to weight the value of the information against the labor of obtaining it. For reasons which will be shown, considerable clerical work was necessary to get this data in usable form. For that reason, the number of observations was kept as small as possible with the consideration that a certain

number of observations were absolutely necessary if useful results were to be obtained. It developed that there were 31 airlines for which all the needed information was available. While twice this number could have been obtained by using observations taken in two consecutive years, it was believed that the work necessary to process this additional data would be excessive in view of the limited additional information it would yield.

Time of observation. All the data used in this study applies to the year 1951. At the time of writing, this was the latest year for which data was available. It may be assumed that if an economic law applies, it will apply in one year just as effectively as in any other year. There is much to be said for using observations coming from more than one year, but the advantage of this is chiefly that fluctuations around a norm will tend to approach a limit as the number of observations increases. There is also a distinct disadvantage to using observations that come from more than one year. Costs may fluctuate due to general economic conditions so that a less stable norm, rather than a more stable one, is obtained by using additional samples.

Independence of observations. It is believed that each observation or sample used here is as independent of any other observation as is practical in economic data. It is true that in economic time series observations may not be independent.[2]

---

2 Herman Wold, Demand Analysis (New York: John Wiley & Sons, 1953), p. 44

The Civil Aeronautics Board supervises the manner in which most accounts and records are maintained by the airlines. Presumably each company makes an independent interpretation of the way in which accounts or records should be maintained, and if there is a serious question of procedure such a question is referred to the Civil Aeronautics Board.[3] For this reason, errors in

---

3 Civil Aeromautics Board, Uniform System of Accounts for Air Carriers, Form 41 Manual (Washington, D. C.: CAB, 1951, effective January 1, 1947), p. 1

---

observation due to variations in reporting may be considered independent. It would also appear that operating conditions in one company are relatively independent of operating conditions in another company.

II. INTRODUCTION TO SPECIFIC DATA

Source. Tables II, III, and IV contain data on cost and the factors which are believed to be responsible for variations in cost. As indicated, these tables were taken from Roadcap's World Airline Record.[4] They were compiled for this study from

---

4 Roy R. Roadcap & Associates, compiler, World Airline Record, (Chicago: published by compiler, 1952), 396 pp.

---

data contained throughout that publication. The introduction to

TABLE II

OPERATING AND COST DATA

	(1) Length of the Flight (Flight stage, in Miles)	(2) Speed of the Plane (Average Miles per Aircraft Hour)	(3) Plane Utilization (Daily Flight Time per Aircraft)	(4) Metropolitan Population Served (000)	(5) Cost (Total Operating Expense per Revenue Ton Mile, in cents)
All American	57	133	6:06	20,200	116.3
American	270	216	6:56	56,928	43.0
Bonanza	100	140	4:27	183	141.5
Braniff	176	182	6:36	11,869	50.6
Capital	142	167	7:28	41,097	51.0
Central	51	134	4:40	1,757	318.5
C & S	175	175	8:36	18,000	59.2
Colonial	112	150	6:52	13,500	77.0
Continental	131	179	6:30	3,831	62.3
Delta	174	191	7:36	13,119	45.3
Eastern	182	187	9:30	44,000	42.6
Empire	59	143	4:50	451	112.4
Frontier	81	141	7:28	2,500	125.2
Lake Central	73	142	3:51	5,405	169.3
Mid-Continent	144	167	6:12	6,725	64.8
Mohawk	79	137	5:53	9,250	100.5
National	199	207	8:17	23,431	42.9
Northeast	94	150	6:46	16,000	81.1
Northwest	271	202	6:53	27,000	56.7
Piedmont	90	153	8:08	3,362	75.4
Pioneer	89	150	7:03	2,050	71.3
Southern	67	147	6:09	2,337	150.1
Southwest	54	127	5:51	5,143	78.9
Trans-Texas	78	147	6:05	2,600	130.3
TWA	293	193	6:24	51,500	46.2
United	270	207	6:38	48,913	42.3
West Coast	58	135	4:30	1,850	103.3
Western	172	181	6:40	6,250	44.2
Inland	147	158	9:23	1,000	64.7
Wiggins	45	116	2:21	2,500	820.9
Wisconsin	69	142	5:43	6,500	130.9

Source: World Airline Record (Chicago, Roadcap and Associates, 1952), data applicable to year 1951

TABLE III

TON-MILE LOAD FACTOR AND AVAILABLE TONS PER AIRCRAFT MILE

Company	(1) Revenue Tons per Aircraft Mile	(2) Ton-Mile Load Factor*/	(3) Available Tons Per Aircraft Mile, Column (1) Divided by Column (2)
All American	.96	.400	2.40
American	3.98	.689	5.77
Bonanza	.79	.358	2.21
Braniff	2.57	.557	4.61
Capital	2.68	.510	5.25
Central	.35	.167	2.10
C & S	2.17	.558	3.89
Colonial	1.68	.505	3.32
Continental	1.70	.537	3.17
Delta	2.74	.598	4.58
Eastern	3.07	.528	5.81
Empire	.69	.313	2.20
Frontier	.72	.398	1.81
Lake Central	.48	.212	2.26
Mid-Continent	1.53	.565	2.71
Mohawk	1.02	.476	2.14
National	3.12	.567	5.50
Northeast	1.82	.596	3.05
Northwest	4.30	.570	7.54
Piedmont	1.08	.449	2.41
Pioneer	1.16	.483	2.40
Southern	.60	.317	1.89
Southwest	1.18	.491	2.40
Trans-Texas	.65	.372	1.75
TWA	3.62	.670	5.40
United	3.75	.630	5.95
West Coast	.82	.421	1.95
Western	2.68	.656	4.09
Inland	1.54	.575	2.68
Wiggins	.07	.166	.42
Wisconsin	.81	.430	1.88

*/ Converted to decimal form

Source of first two columns: World Airline Record (Chicago, Roadcap and Associates, 1952), data applicable to year 1951

TABLE IV

ASSETS EMPLOYED

	Total Assets (00000's)	Investments and Special Funds (00000's)	Adjusted Assets (Total Assets less inv. and special funds) (00000's)
All American	21.13	3.21	18
American	1436.53	165.22	1271
Bonanza	6.65	.01	6.6
Braniff	160.30	5.81	154
Capital	195.02	6.06	189
Central	14.02	.01	14
C & S	114.16	3.11	111
Colonial	35.34	.79	35
Continental	49.74	3.49	46
Delta	174.32	3.87	170
Eastern	1042.58	187.84	855
Empire	4.71	.02	4.7
Frontier	17.66	.12	18
Lake Central	6.32	.01	6.3
Mid-Continent	76.12	2.44	74
Mohawk	14.54	.09	14
National	181.27	16.14	165
Northeast	49.38	2.11	47
Northwest	471.50	7.98	464
Piedmont	21.60	3.10	19
Pioneer	19.11	.01	19
Southern	12.44	.01	12
Southwest	16.06	.08	16
Trans-Texas	11.10	.42	11
TWA	1217.98	120.07	1098
United	1127.25	79.98	1047
West Coast	10.98	.01	11
Western	137.39	17.02	120
Inland	16.61	.03	17
Wiggins	2.03	.62	1.4
Wisconsin	10.80	.01	11

Source of first two columns: World Airline Record (Chicago, Roadcap and Associates, 1952), data applicable to year 1951

this publication[5] indicates that it is gathered from published sources. Presumably this would be from Civil Aeronautics Board

---

5 Ibid., p. 4

---

reports in the case of U. S. firms. Roadcap does not indicate the source of the data.

Cost per revenue ton mile. Cost per revenue ton mile for the year 1951 is given in Table II, Column (5). This cost is derived from the audited statements of the 31 firms indicated in the table. The following items are included under the caption "Total Operating Expenses" and are therefore part of the cost measured here: total aircraft operating expenses; ground operations expense; ground and indirect maintenance; passenger service expense; traffic and sales expense; advertising and publicity expense; general and administrative expense; depreciation expense on both flight and ground equipment.[6] The above costs include practically

---

6 Civil Aeronautics Board, Recurrent Report of Financial Data. (Washington: Civil Aeronautics Board, for the years 1950 and 1951)

---

all costs of furnishing air service except return to stockholders.

There might be considerable debate on whether or not some other definition of unit cost would serve the purpose of this study more effectively than cost per revenue ton mile. Cost per passenger mile, for example, might have been more useful. The

use of cost per passenger mile would have been justified chiefly on the basis that the cost of transporting passengers is different from the cost of transporting cargo. It seemed a reasonable assumption that for purposes of this study the cost of transporting passengers and the cost of transporting other cargo are relatively the same. It is true that passengers demand extra service once they are on the plane, but it is also true that special crews and equipment are not needed to place passengers on the plane. The assumption made here is consistent with the findings of the Civil Aeronautics Board that: "The relative volumes of mail, passengers, and freight, comprising a carrier's total tonnage, seem to make no difference in its level of expenses."[7]

---

7 Civil Aeronautics Board, Cost Standards - Domestic Scheduled Air Carriers (Washington, D. C.: Civil Aeronautics Board, July 1950), p. 13

Cost per available ton-mile might have been used instead of cost per revenue ton-mile. The use of this measure would certainly be justified if the investigation were directed toward factors other than load factor. If load factor is to be investigated, however, it is essential that cost per revenue ton-mile should also be used. This enables the statistician to make a comparison between load factor and all the other factors which are responsible for variations in unit cost. The previously cited study by Koontz was based principally on cost per available ton-mile. It is believed that a study based on cost per revenue

ton-mile may serve as a valuable check on the conclusions reached by Koontz. There are cases in which it would not be expected from a priori considerations that cost per revenue ton-mile and cost per available ton-mile would give the same results. Consider, for example, the size of the firm. When the size of the firm is correlated with cost per available ton-mile, a firm with a capacity load should have slightly higher costs than a firm with a load of less than capacity. The reason for this is that it requires more personnel, gasoline, and maintenance to handle a greater load. Suppose that a larger firm has greater load factors and hence greater costs. If cost per available ton-mile is used, the larger firm with the larger costs will be at a disadvantage. In this case, there is no way to remove the effects of the load factor upon cost per plane mile. The use of cost per revenue ton-mile, on the other hand, permits the use of statistical techniques to remove all effects of the load factor.

Four independent variables. Figures on the length of the trip are given in Table II, Column (1). Figures on the speed of the plane are given in Table II, Column (2). Plane utilization was defined in Chapter I as the average number of hours per day that a plane is in flight. Data for this is given in Table II, Column (3). While load factor and speed of the plane are in a sense measures of plane utilization, the term "plane utilization" will be used throughout this thesis to refer to the average number of hours per day that a firm uses its planes. Metropolitan

population served is given in Table II, Column (4). Data on each of the variables discussed here were taken without changes from Roadcap's compilation.

Load factor. Ton-mile load factor was used in this study. Data are given in Table III, Column (2). It is customary to express load factors in percentage form rather than in decimal form. For instance, the load factor of American Airlines, as usually written, is 68.9. Column (2) was presented in decimal form so that it might be conveniently used to obtain a measure of available tons per aircraft mile. The previous discussion relative to the use of cost per revenue ton-mile instead of cost per passenger mile may be applied to ton-mile load factor. If cost per revenue ton-mile is to be used, it is logical to use the ton-mile load factor. If cost per passenger mile is to be used, the corresponding passenter-mile load factor should be used.

Capacity of the plane. A measure of available tons per aircraft mile is used in this study. Data are given in Table III, Column (3). No measure of this nature could be found in published form. For this reason, it was necessary to use available data to derive such a measure. It seems obvious that available tons per aircraft mile could be multiplied by the ton-mile load factor to obtain revenue tons per aircraft mile. It follows that revenue tons per aircraft mile may be divided by ton-mile load factor to obtain available tons per aircraft mile. Revenue tons per aircraft mile and ton-mile load factor are to be found among Roadcap's data.

population served is given in Table II, Column (8). Data on each of the variables discussed here were taken without changes from Roadcap's compilation.

Load factor. Ton-mile load factors used in this study. Data are given in Table III, Column (2). It is customary to express load factors in percentage form rather than in decimal form. For instance, the load factor of American Airlines, as usually written, is 60.9. Column (2) was presented in decimal form so that it might be conveniently used to obtain a measure of available tons per aircraft mile. The previous discussion relative to the use of cost per revenue ton-mile instead of cost per passenger mile may be applied to ton-mile load factor. If cost per revenue ton-mile is to be used, it is logical to use the ton-mile load factor. If cost per passenger mile is to be used, the corresponding passenger-mile load factor should be used.

Capacity of the plane. A measure of available tons per aircraft mile is used in this study. Data are given in Table III, Column (3). No measure of this nature could be found in published form. For this reason, it was necessary to use available data to derive such a measure. It seems obvious that available tons per aircraft mile could be multiplied by the ton-mile load factor to obtain revenue tons per aircraft mile. It follows that revenue tons per aircraft mile may be divided by ton-mile load factor to obtain available tons per aircraft mile. Revenue tons per aircraft mile and ton-mile load factor are to be found among Roadcap's data.

In Table III, Column (3) was obtained by dividing each respective element of Column (1) by the corresponding figure in Column (2).

Size of the firm. If the term "size of the firm" has any meaning whatsoever, it must be a measure of the investment of each firm, i.e., it must depend on the amount of flight equipment and other assets owned by a firm. This is the concept adopted by the Civil Aeronautics Board.[8] Adjustment must be made to total

---

8 Ibid., p. 75

---

assets to obtain a measurement suitable for the purposes here under consideration. The reason for this is that firms often set aside large sums of money for special funds, or hold large investments in other companies. Such sums are not properly part of the investment of a given company. That is, these amounts are not required for that company to engage in the transportation business. Except for these sums, the remaining assets are invested in equipment, contracts, tied up in current assets, etc. In Table IV, investments and special funds were deducted from total assets. The column "Adjusted Assets" therefore represents the assets that are used by each company for transportation purposes.

Closely tied to the data that has been presented here is the theory and techniques by which this data was manipulated to arrive at a least-squares equation. Accordingly, attention is now directed to the general techniques by which the data here presented was utilized.

III. STATISTICAL THEORY

Nature of prediction. The word "prediction" should be understood to refer to what is going to happen under given conditions, not to what is going to happen in the future. If one variable increases, what can be said about the effect of this increase upon another variable. If there is some relationship between changes in two variables such that a change in one will usually produce some effect upon the other, the necessary conditions for prediction are met. Suppose, for example, that the relationship between agricultural production and rainfall is under consideration. The fact that an increase in farm production is generally associated with an increase in rainfall will provide the necessary condition for statistical prediction. This type of prediction is fundamental in all science.

A priori nature of causality. The conditions necessary for statistical prediction do not include the conditions sufficient for causality. In order to make statements about causality it is necessary to introduce reasons for causality which are not contained in statistical theory. In the case of rainfall and farm production, it is necessary to introduce the hypothesis that rainfall is responsible for farm production. Usually such a theory of causality will originate from sources outside of the data under investigation. It was because of the need for theories of causality that the material in Chapter I was presented.

Statistical theory may test a priori theory. Two variables are independent of each other when changes in one variable are not associated with changes in the other variable. Similarly, two variables are said to be correlated when changes in one variable are associated with changes in the other variable. The degree of correlation is subject to statistical measurement. Dependence may be established only by a priori methods. Dependence can, however, be disproven by statistical methods. For example, suppose that a priori theory asserts that changes in rainfall are responsible for changes in farm production. Suppose it can be shown by statistical methods that changes in rainfall are independent of changes in farm production. The a priori theory and the statistical test cannot at the same time be true. The statistical test is more direct and is therefore subject to less error. Hence it will be accepted in preference to a priori theory. Dewey[9] has contributed to the philosophical

---

9 John Dewey, Theory of Valuation (Chicago: University of Chicago Press, November 4, 1947), pp. 39-40

---

significance of this choice by showing that a priori theory tends to be independent of the real world. Assumably, statistical measures represent estimates of the real world. Mathematically, the above choice is justified on the ground that two independent variables are always uncorrelated. Unfortunately it cannot always be said that two uncorrelated variables are necessarily

independent.[10] This means that it is usually not practical to

---

10 Harold Cramèr, Mathematical Methods of Statistics (Princeton: Princeton University Press, 1951), p. 278

---

disprove a hypothesis. Usually the statistician will be able to show only that a given situation could have occurred by chance. This is taken as sufficient evidence on which to refuse to accept a hypothesis. Under such circumstances, the statistician can show only that the a priori theory cannot be proven. On the basis of Dewey's empirical philosophy, it can be argued that a theory which cannot be proven empirically is not true in the real world. This means that in effect no a priori theory is accepted until it is proven empirically.

The fact that statistical theory verifies a priori theory proves only that the a priori could be true. As mentioned previously, the fact that two variables are highly correlated does not prove that one is responsible for the other. It is possible that a third variable may be responsible for all changes in the two variables under investigation. The statistician may adopt the philosophy that his correlation represents the real world. A priori theory is valuable primarily for the extent to which it explains statistical findings and paves the way for further investigation. To accept a priori theory per se means that an absolute truth is reached. Such an acceptance is contrary to the scientific method.

Effects of all variables except one removed. In this study the effect upon unit cost of each of seven variables is under consideration. It is possible to remove the effect of all variables except one. The question may then be asked, what is the correlation between unit cost and that variable after the effects of all other variables have been removed? This question is answered by the coefficient of partial correlation. A similar question is, how much better is the coefficient of multiple correlation where this variable is used than where it is not used in a least-squares regression equation. The mathematical appendix shows that both of these questions may be answered with Student's *t* test.

Why least-squares regression. The method of least-squares regression is subject to considerable criticism, but in general it is as effective as any other method available. In the first place, it is not necessary to make any assumptions about population distribution in order to use least-squares regression methods. According to the Markoff theorem, the method of least-squares gives the best unbiased estimate of the dependent variable.[11] The most

---

11 Gerhard Tintner, Econometrics (New York: John Wiley & Sons, 1952), p. 84

---

complete analysis of least-squares regression considered in connection with this study was the investigation by Wold. This analysis is perhaps as authoritative as any in the field. Wold

has come to the following conclusions:

> In the end there does not really remain much to be said against the method except that it does not always come up to the optimal efficiency. Much of the recent criticism of regression methods has no doubt been inspired by a desire to extend the triumphs of modern statistical methods to the field of economic and social statistics. We have seen, however, that such a programme meets great difficulties, for in applications to nonexperimental data we lack the experience and knowledge required for the proper specification of disturbances and residuals. This does not exclude the possibility that better methods may exist or could be devised. When it comes to practical application, however, their advantage will always have to be balanced against the substantial advantages of the least-squares method of being highly flexible as regards the underlying assumptions and very simple as regards the numerical computations. For the present the theoretical and practical evidence available in this direction is not strong enough to warrant the abandonment of the traditional methods.[12]

---

12 Wold, op. cit., p. 59

IV. REGRESSION TECHNIQUE

Steps in computation. Computations of the usual problems in regression analysis involve two major steps: (1) obtaining the raw data for a system of simultaneous equations; (2) solving these equations. Three types of data are necessary for such equations: (1) the mean of each variable involved; (2) the sum of the squares for each variable involved; (3) the summation of the cross products of each variable with every other variable. For the

accumulation of data of this nature a calculating machine is almost necessary.

Doolittle method. The Doolittle method of solving systems of simultaneous equations has the almost unanimous recommendation of textbooks on the subject. An important feature of this method is its utilization of the symmetrical aspects of least-squares determinants. The standard least-squares determinant has the same elements across the top as it has down the left side. The Doolittle method uses this feature as an aid in laying out the computation so that there is a great uniformity in operation. Once the computer has become familiar with this method, he can perform a rather difficult computation in a routine manner. In fact, significant operational mistakes are almost impossible.

The moment form of the Doolittle method. The moment form of the Doolittle method was the method of computation used in this study. Where Z_{ij} was a moment as outlined in the mathematical appendix, the form nZ_{ij} was used. As Klein[13] has indicated, this

---

13 Klein, op. cit., p. 145

---

method has the advantage of greater accuracy than the method in which the moment is not multiplied by the number of observations. The effect of this method upon all the simultaneous equations is that each of them is multiplied by the number of observations.

The difference between the moment method and the correlation

method is that with the correlation method all data must be divided by the standard deviation and the square root taken before computational work begins. This means that, with the correlation method, a single error in dividing or extracting the square root will cause the final least-squares equation to be in error. Under the moment method, the computation begins with the respective moments. The first step is the division of the first row by the first element of that row. Under the correlation method outlined by McNemar,[14] the principal difference is that the efficiency of

---

14 Quinn McNemar, Psychological Statistics (New York: John Wiley & Sons, 1950), p. 156-160

---

repeated operations is obtained. Moreover, the correlation method does not furnish a check where this operation is performed. Under the moment method, an error in any dividing operation will be detected.

The necessity of extracting the square root of a large number of variables is more time-consuming than it would appear. If an error is made in looking up a square root in the tables, no check is available. At the best there is a rounding error in extracting a square root. Klein[15] has pointed out that small

---

15 Klein, loc. cit., p. 145

---

errors in the original equations can result in substantial errors in the solution. The moment method avoids such errors. Not only

may the extracting of the square root be avoided before the solution of the simultaneous equations is made, but the F test may be used instead of the Student's t test, thus avoiding the extraction of a square root before testing the significance of each variable. The mathematical appendix outlines the theory for this.

Eliminating curvilinearity. Tintner[16] has remarked that the

---

16 Tintner, op. cit., p. 84

---

system of linear least-squares regression may be used in some cases where the relationship between the independent variable and the dependent variables is not linear. It is sufficient that the parameters to be estimated appear in linear form. Thus $\log X_j$, $\sin X_j$, etc., may be used if the substitution of these gives a linear function. A particular case of this occurs where both the logarithms of independent and dependent variables are linear. For example, the equation might appear as follows:

$$X_1 = (X_2)^{a_2}(X_3)^{a_3} \ldots (X_n)^{a_n} C$$

$$\log X_1 = a_2 \log X_2 + a_3 \log X_3 + \ldots + a_n \log X_n + C$$

Here X_j is replaced by $\log X_j$ and the computation proceeds in a normal manner. Before this form can be used, it is necessary to show that given data is linear in logarithms. This can be accomplished most easily by plotting on log-log paper. The

dependent variable, X_1, is first plotted on log-log paper with X_2. It is then plotted with X_3, X_4, etc., If the relationship between the dependent variable and the independent variable is linear, each of these plottings will approximate a straight line. If the data is curvilinear and logarithms are not used, it is necessary to use equations of degree greater than one. While the technique is primarily the same, it is easier to look up the logarithms of each individual variable than to solve an equation of higher degree.

Shifting the decimal point. If the square moment of one variable is many times as large as the moment of some other variable, this results in a rapid loss of accuracy in computation. For this reason, it is desirable to multiply or divide each variable by an appropriate power of 10 so that the resulting moments are all of approximately the same magnitude. Such operations have no effect upon the coefficient of multiple correlation, but they do effect the regression coefficients. If a variable is multiplied by 10, this has the effect of dividing the regression coefficient by ten. In other words, the regression coefficient is smaller, by a power of 10, than it would have been had the regression equation been solved without multiplying that variable by 10. Similarly, if a variable is divided by 10, this has the effect of multiplying the regression coefficient by 10.[17]

---

17 Klein, op. cit., pp. 145-146

Consider the situation in which it is desirable to shift the decimal point of one or more variables, solve the simultaneous equations of which these variables are a part, and obtain the same regression coefficients as though no shift in decimal point had been effected. If certain variables have been divided by 10, this will mean that the coefficients obtained after this division are 10 times as large as they would otherwise have been. To obtain the coefficients that would have been obtained had no shift in decimal been effected, it is merely necessary to divide the resulting coefficients by 10. Suppose the logarithmic equations of the type previously described, but with the base e, are to be fitted. As previously indicated, least-squares equations may be determined by computing the proper coefficient of each logarithm. Dividing one or more logarithms by 10 prior to solving the simultaneous equations causes the power of each independent variable to be multiplied by 10. Therefore, if the resulting exponent is divided by 10, the quotient will be the coefficient which would have resulted had the decimal point of the logarithm never been divided by 10.

This chapter has presented data that may be used in the solution of an economic problem. It has outlined statistical theory and technique which may be used on that problem. The next chapter will describe the computations connected with this problem.

CHAPTER III

COMPUTATIONS

The problem, data, and techniques for manipulating that data have been presented in previous chapters. This chapter will present the results of applying least-squares techniques to the data. Two major stages were involved in this application. In the first stage, a correlation was made between cost per revenue ton-mile and seven independent variables. Only three variables were found to be significant. In the second stage, three significant variables were fitted in a least-squares equation. This equation was used to predict the results of load factors at various levels.

I. SEVEN INDEPENDENT VARIABLES

Graphing. The first step performed was the drawing of seven scatter diagrams. Each diagram showed the relationship between cost per revenue ton-mile and one of the independent variables. When these diagrams were plotted on ordinary graph paper, the indicated curves were all curvilinear. Semi-logarithmic paper gave a straight line in only one case - the relationship between the logarithm of cost per revenue ton-mile and load factor. In plotting all other factors, a straight line was obtained only when log-log paper was used. This graphing indicated that a linear relationship existed between the logarithm of cost per revenue ton-mile and the logarithm of six other variables. In

the case of load factor, the use of raw data instead of logarithms was indicated. Following the suggestions for eliminating curvilinearity given in Chapter III, this data could be indicated by the following equation.

$$\log X_1 = b_2 \log X_2 + b_3 \log X_3 + b_4 X_4 + b_5 \log X_5 + \ldots + b_8 \log X_8$$

The only variable with which logarithms is not used is ton-mile load factor, X_4. The base of logarithms used throughout this study is e. The coefficients, b_j, $j = 2, 3, \ldots 8$, were found by the least-squares technique. The variables are as follows: X_1 is cost per revenue ton-mile (in cents); X_2 is the capacity of the plane (in tons); X_3 is the flight stage (length of non-stop flight in miles); X_4 is the ton-mile load factor (in percentage); X_5 is flight time (in average mimutes per plane per day); X_6 is the metropolitan population served (in hundreds of thousands); X_7 is average speed of each plane (in miles); and X_8 is the net assets of each firm (in hundreds of thousands of dollars). Each of these variables is expressed in deviations from its mean in accordance with the symbols set forth in the mathematical appendix.

Examples of the extent to which these scatter diagrams were linear are given in Diagrams, I, II, and III. The remaining diagrams were so similar to these that it did not seem worth while to reproduce them here. No measures were used to establish whether or not these relationships were more linear than curvilinear. Because of the small number of samples, it was assumed that suffi-

the case of load factor, the use of raw data instead of logarithms was indicated. Following the suggestions for eliminating curvilinearity given in Chapter III, this data could be indicated by the following equation.

$$\log X_1 = b_2 \log X_2 + b_3 \log X_3 + b_4 X_4 + b_5 \log X_5 + \ldots + b_8 \log X_8$$

The only variable with which logarithms is not used is ton-mile load factor, X_4. The base of logarithms used throughout this study is e. The coefficients, b_j, $j = 2, 3, \ldots 8$, were found by the least-squares technique. The variables are as follows:

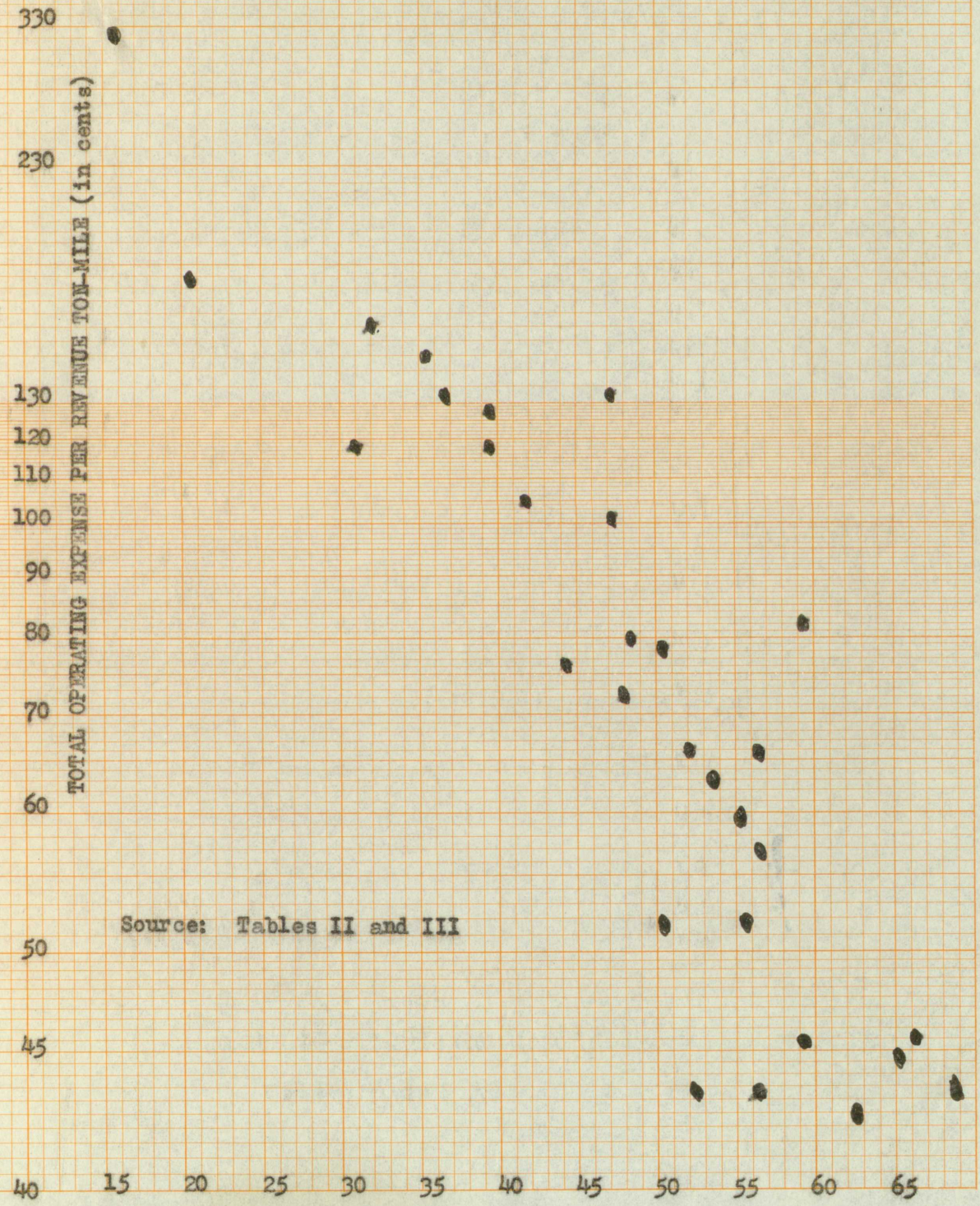
X_1 is cost per revenue ton-mile (in cents); X_2 is the capacity of the plane (in tons); X_3 is the flight stage (length of non-stop flight in miles); X_4 is the ton-mile load factor (in percentage); X_5 is flight time (in average minutes per plane per day); X_6 is the metropolitan population served (in hundreds of thousands); X_7 is average speed of each plane (in miles); and X_8 is the net assets of each firm (in hundreds of thousands of dollars). Each of these variables is expressed in deviations from its mean in accordance with the symbols set forth in the mathematical appendix.

Examples of the extent to which these scatter diagrams were linear are given in Diagrams I, II, and III. The remaining diagrams were so similar to these that it did not seem worth while to reproduce them here. No measures were used to establish whether or not these relationships were more linear than curvilinear. Because of the small number of samples, it was assumed that multi-

SCATTER DIAGRAM I

OPERATING EXPENSE AND LOAD FACTOR

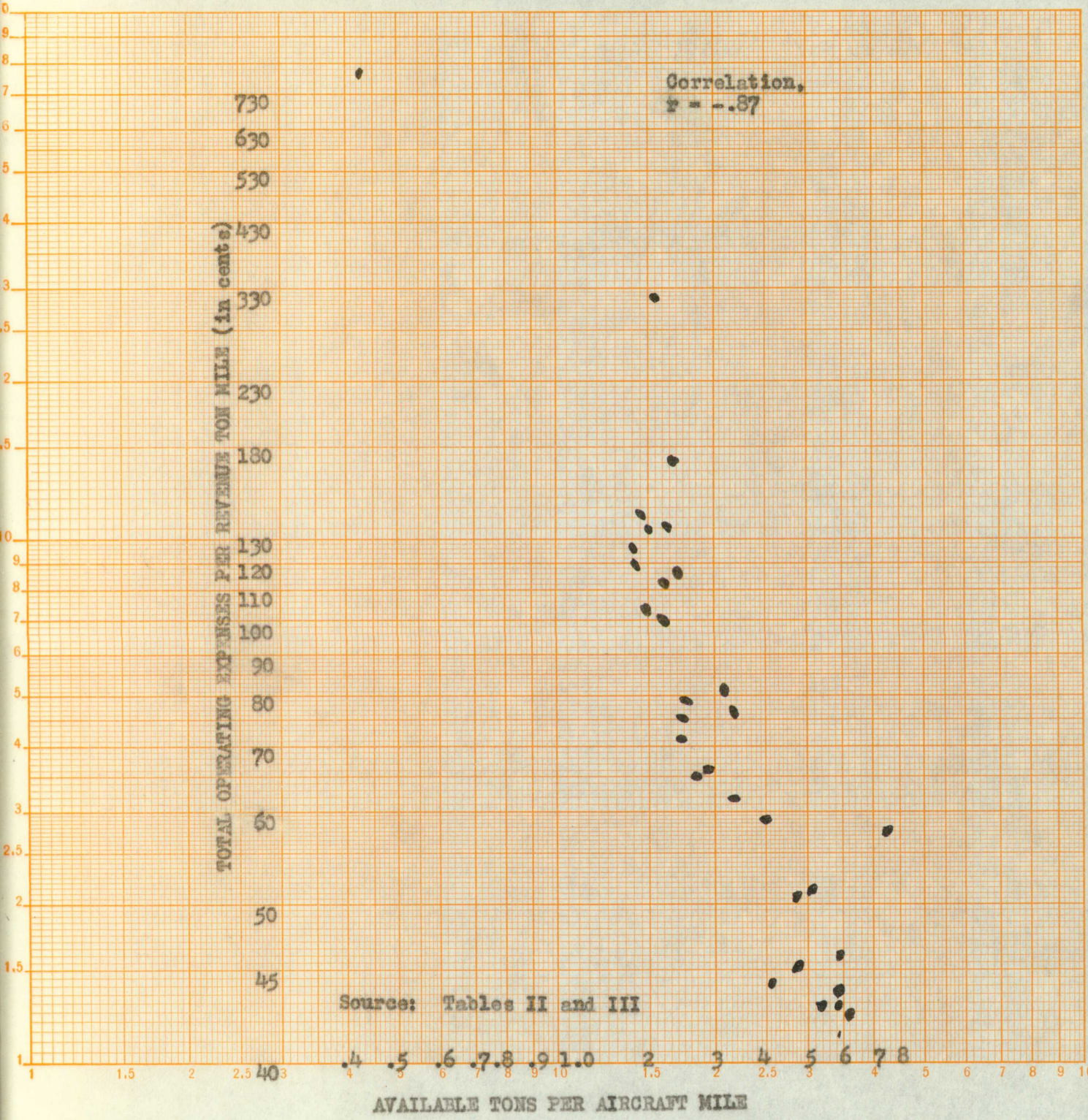
Correlation,
r = -.91.



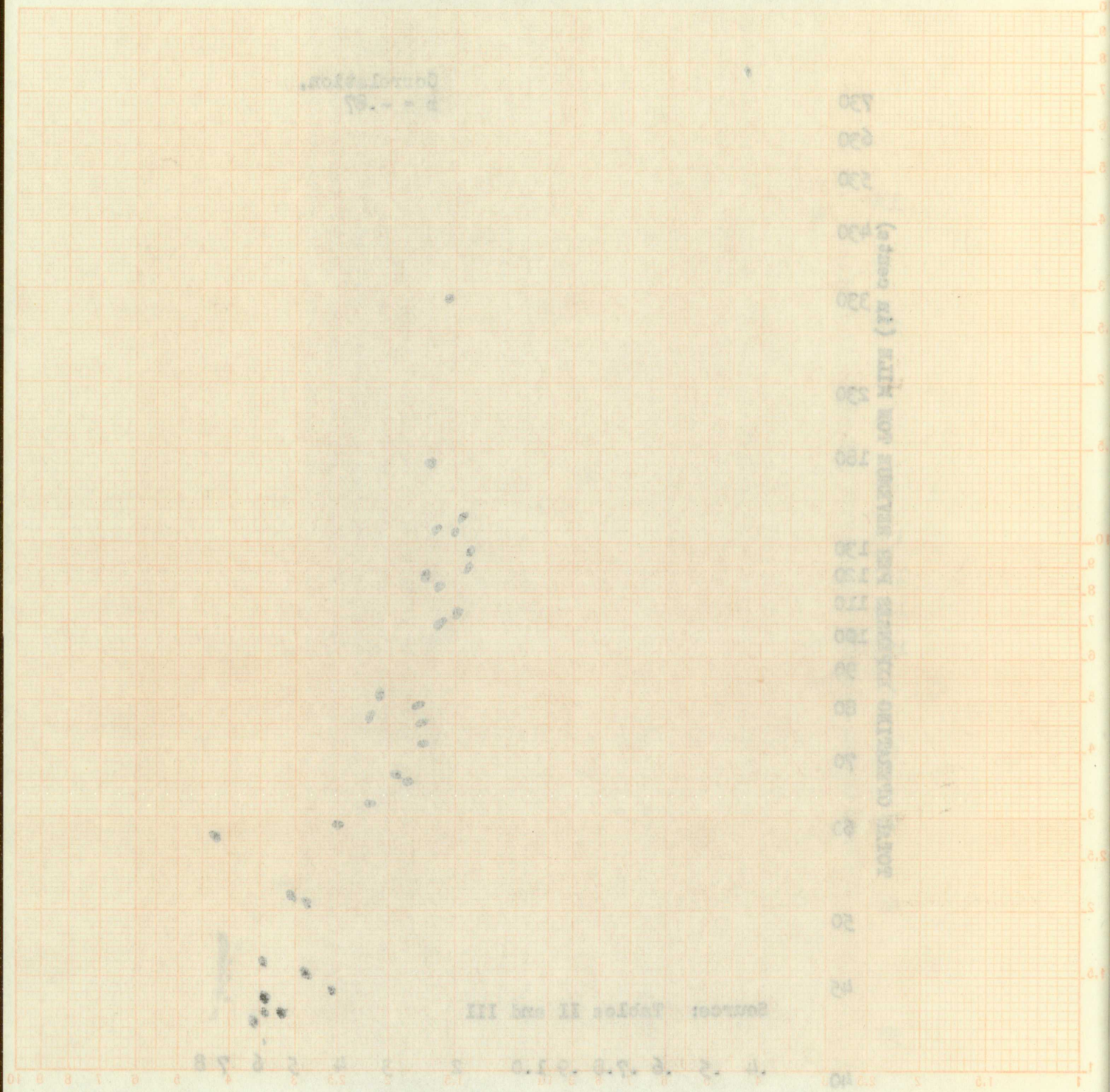
Source: Tables II and III

SCATTER DIAGRAM II

OPERATING EXPENSE AND CAPACITY OF PLANES



Source: Tables II and III



SCATTER DIAGRAM III

DAILY FLIGHT TIME PER AIRCRAFT

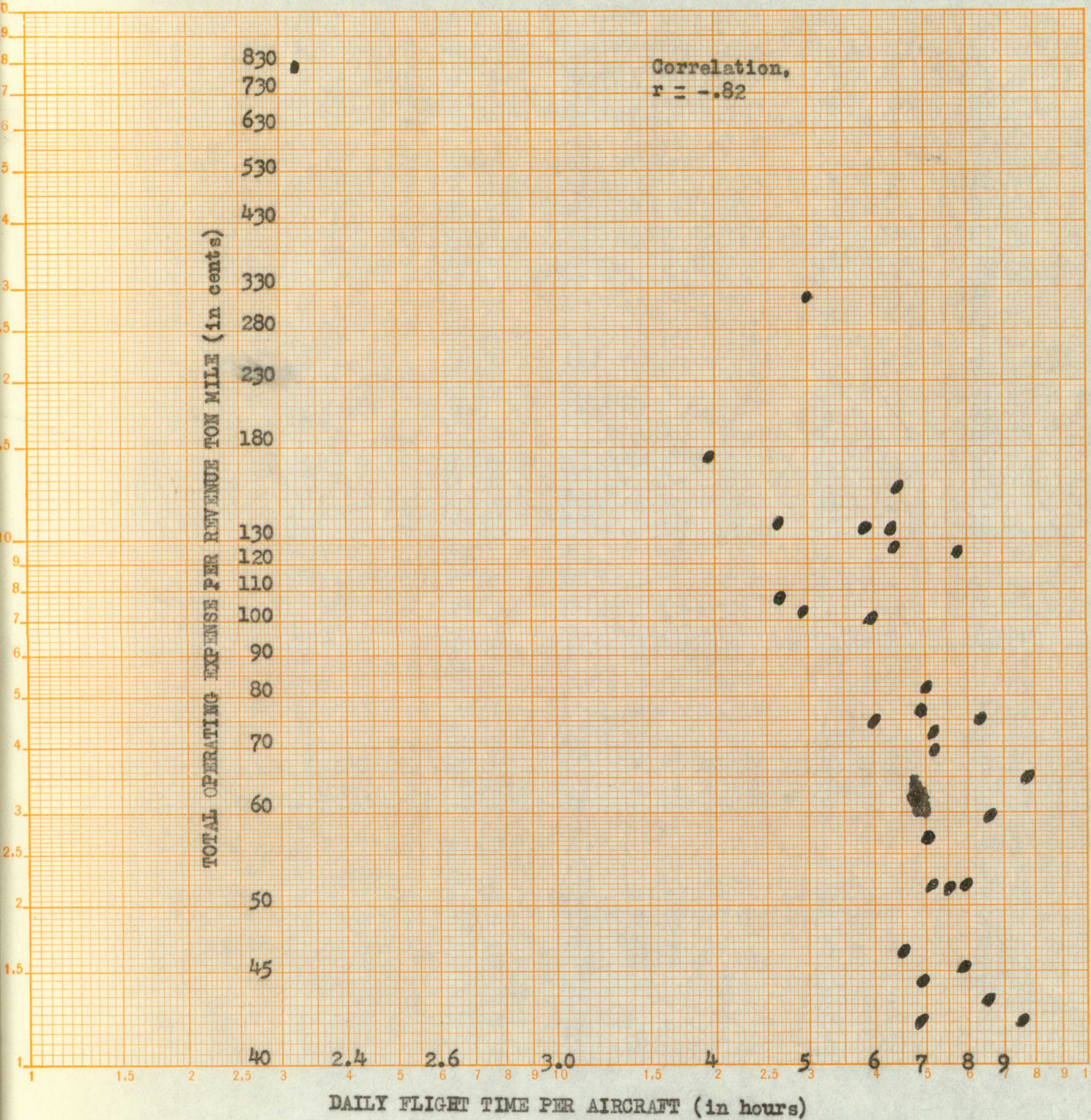


DIAGRAM NUMBER III

DAILY FLIGHT TIME PER AIRCRAFT

cient information could be gained from the diagrams. Correlation coefficients appeared to verify this.

Correlation coefficients. The logarithms of each observation for each variable were copied from the tables. The raw score for the ton-mile load factor was used instead of the logarithm of this variable. The logarithms of all other variables were used. The correlations between cost per revenue ton-mile and each of the indicated variables was found to be as follows: capacity of the plane (in tons), - .87; flight stage (in miles), - .81; ton-mile load factor (in percentage), - .91; plane utilization (in average minutes per plane per day), - .82; metropolitan population served, - .56; average speed of each plane, -.82; size of the firm (in assets), - .81.

It was observed that the correlation with ton-mile load factor is greater than the correlation with any other factor. Capacity of the plane gives the second highest correlation, while flight time gives the third highest.

It is difficult to extract any useful information from these correlations. All the correlations are definitely significant. It is impossible to say, from this evidence, that one factor is significant and another factor is not significant. It is impossible, for example, to say that load factor is more significant than capacity of the plane. It is impossible to establish statistically that capacity of the plane is more significant than plane utilization, or that plane utilization is more significant than speed

of the plane. It can be shown that six of the coefficients are more significant than metropolitan population served. It can be shown that load factor is on the border line of being more significant than length of flight and assets employed. In general, all of these variables appear to be significant.

It would be impossible to imply from these correlations that any one of them is responsible for variations in cost per revenue ton-mile. All of them are obviously highly correlated. For example, the correlation between length of the flight and capacity of the plane is .80. The correlation between length of the flight and load factor is .83. In general, the correlation between any independent variable and any other independent variable is somewhere in the 80's. It is difficult to say to what extent changes in one variable are responsible for changes in another variable. It is difficult to determine, for example, to what extent changes in load factor and length of the flight are responsible for changes in speed of the plane. Statistically, it is possible from these correlations that length of the flight is extremely significant. On the other hand, it is also possible that changes in some of the other variables are responsible for all changes in the length of the flight. For example, suppose that load factor causes changes in cost and at the same time causes changes in the length of the flight. It would then be true that cost and length of the flight are correlated, but that their correlation does not mean that changes in cost are occasioned by changes in the length of the flight.

For this reason, one of these correlations by itself is worthless as a means for determining cause. To determine cause, a different technique must be employed. The use of least-squares regression of partial correlation for determining cause was discussed in Chapter II. With this recognition of the need for least-squares regression, attention will be directed to the solution of a least-squares equation based on the data previously presented.

The moment matrix. Any element of the moment matrix may be designated nZ_{ij} where n is the number of observations and Z_{ij} if defined in the mathematical appendix. As indicated, the moment Z_{ij} is defined in terms of deviation from the mean. Except for the case of load factor, all means were the means of logarithms and the variables used in computation were the logarithms of the indicated variables. The raw observations were used in the case of the load factor. The reason for multiplying each element of the moment by n was to avoid rounding errors.[1] Elements of the

---

1 Lawrence R. Klein, Econometrics (Evanston, Illinois: Row, Peterson and Company, 1953), p. 145

---

moment matrix are contained in the first seven lines and first seven columns of Table V. In order to obtain this matrix, the decimal point of each moment was shifted according to the method outlined in the previous chapter. Load factor (X_4) was divided by 100 and assets (X_8) were divided by 10.

Computations. Table V gives the major part of the computations necessary to obtain the coefficients for the equation arising in this study. The initial set of simultaneous equations to be solved is given in lines (1) through (7). Line (8) is a reproduction of line (1). Line (9) is obtained by dividing line (8) by its first element. The method of making the computations shown in lines (8) through (29) is given by Klein.[2] Line (22) gives the

---

2 Klein, op. cit., pp. 152-157

---

regression coefficients which were obtained by a back solution. Elements of the inverse matrix contained in lines (23) through (29) were obtained in a similar manner by back solution. It should be noted that the columns of this computation which were used in computing the inverse matrix are not reproduced here. The two check columns are also omitted. Line (30) gives the regression coefficients computed from the inverse matrix (lines 23 through 29). Lines (22) and (30) are equal to within three decimal places. Line (31) contains the t ratio for determining the significance of the regression coefficients. This ratio is explained in the mathematical appendix. Line (32) shows the levels at which each of these regression coefficients is statistically significant.

Significance tests. Student's t distribution was used for determining the levels at which each of these coefficients is significant. This test tells whether or not each of the

TABLE V

CALCULATION OF ESTIMATED LINEAR EQUATION

	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_1
(1)	319.05	246.32	57.03	107.87	50.97	73.20	82.86	-310.33
(2)		284.64	59.40	87.90	46.62	76.78	83.36	-272.48
(3)			18.01	26.06	11.29	16.58	18.62	- 77.07
(4)				73.58	16.13	26.62	28.99	-141.12
(5)					18.92	18.89	19.07	- 49.19
(6)						24.87	24.73	- 81.75
(7)							29.40	- 88.12
(8)	319.05	246.32	57.03	107.87	50.97	73.20	82.86	-310.33
(9)	1.0	.772042	.178749	.338097	.159756	.229431	.259709	- .972669
(10)		94.470615	15.370444	4.619829	7.269019	20.266526	19.388600	- 32.892206
(11)		1.0	.162701	.048902	.076945	.214527	.205234	- .972669
(12)			5.315158	6.026696	.996487	.198189	.654313	- 16.247228
(13)			1.0	1.133869	.187480	.037288	.123103	- 3.056772
(14)				30.050074	-2.588159	.655506	- .714764	- 16.167635
(15)				1.0	- .086128	.021814	- .023786	- .538023
(16)					9.808188	.655754	4.156530	4.571515
(17)					1.0	.066858	.423782	.466092
(18)						3.662402	1.273265	- 2.841546
(19)						1.0	.347658	- .775870
(20)							1.599640	- 1.107872
(21)							1.0	- .067406

TABLE V

CATEGORIZATION OF RESPONDENTS BY EDUCATION

TABLE V (Continued)

CALCULATION OF ESTIMATED LINEAR EQUATION

	X_2	X_3	X_4	X_5	X_6	X_7	X_8
			(Regression Coefficients)				
(22)	-.423440	-.228620	-2.582553	-.476276	+.544964	-.752436	-.067406
			(Inverse Matrix)				
(23)	+.014632	-.002284	+ .002918	-.008524	-.001345	-.005542	-.022673
(24)	-.002285	+.033827	- .033226	+.005966	+.023949	-.036542	-.059110
(25)	+.002918	-.033226	+ .239846	-.040224	-.021302	+.008624	-.019692
(26)	-.008524	+.005966	- .040224	+.034248	+.010067	+.006848	-.001955
(27)	-.001345	+.023949	- .021302	+.010067	+.203468	+.068796	-.250393
(28)	-.005542	-.036542	+ .008624	-.006848	+.068796	+.348603	-.217335
(29)	-.022673	-.059110	- .019692	-.001955	-.250393	-.217335	+.625141
			(Regression coefficients by inverse matrix)				
(30)	-.423224	+.228759	-2.582550	-.476474	+.544828	-.752259	+.067500
			(Application of t test)				
(31) t	3.39	1.20	5.09	2.49	1.17	1.23	.082
			(Significance levels)				
(32)	.01	.2	.01	.02	.2	.2	.5 (approx)

TABLE V (Continued)

ORTHOGONALIZATION OF MINIMIZED TRIGONOMETRIC EQUATION

(Eliminated Coefficients)

	x_1^2	x_1^3	x_1^4	x_1^5	x_2^0	x_2^1	x_2^2	x_2^3	x_2^4	x_2^5	x_3^0	x_3^1	x_3^2	x_3^3	x_3^4	x_3^5
(1)	-.4.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500
(2)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(3)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(4)	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500
(5)	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500
(6)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(7)	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500
(8)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

(Eliminated Coefficients)

	x_1^2	x_1^3	x_1^4	x_1^5	x_2^0	x_2^1	x_2^2	x_2^3	x_2^4	x_2^5	x_3^0	x_3^1	x_3^2	x_3^3	x_3^4	x_3^5
(1)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(2)	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500
(3)	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500
(4)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(5)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(6)	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500
(7)	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500
(8)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

(Eliminated Coefficients)

	x_1^2	x_1^3	x_1^4	x_1^5	x_2^0	x_2^1	x_2^2	x_2^3	x_2^4	x_2^5	x_3^0	x_3^1	x_3^2	x_3^3	x_3^4	x_3^5
(1)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(2)	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500
(3)	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500
(4)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(5)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
(6)	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500
(7)	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500	-.2500
(8)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

coefficients is significantly different from zero. As indicated in the mathematical appendix, this test also shows what would be the effect upon the residual if each of these coefficients in turn were taken as zero. If the 5% level of significance is necessary before a regression coefficient is accepted as significant, there are only three variables with significant coefficients. These three variables are: capacity of the plane (X_2); ton-mile load factor (X_4); and plane utilization or flight time (X_5). As indicated by Table V, all other variables were found to be insignificant. Both plane capacity and load factor were found to be significant at the 1% level. Plane utilization was found to be significant at the 2% level. Coefficients of the magnitude of those obtained for flight stage, metropolitan population, and speed of the plane would have been obtained by chance one time out of five. It appears that the true regression coefficient of each of these variables is zero.

Signs of the coefficients. The algebraic signs of all the coefficients were negative except the signs of the coefficients for flight stage and metropolitan population served. The signs of both of these coefficients was positive. As indicated above, there is no good reason for believing that either of these coefficients is different from zero. In that case, it must be argued that an increase in the length of the flight will be responsible for an increase in cost per revenue ton-mile. The same statement applies to an increase in metropolitan population served. It is an accepted fact that the greater the population of a city,

the longer will be the distance people desire to travel.[3] This

---

3 Joseph L. Nicholson, Air Transportation Management (New York: John Wiley & Sons, 1951), p. 90.

---

suggests that as the population becomes greater and the length of the flight increases, the cost of serving a given number of people increases (after the effects of other factors are considered). This hypothesis could hardly be accepted without verification from other sources, however, because of the probability of one in five that this coefficient arose by chance.

Effects of other variables removed. The phrase "after the effect of other variables have been removed" will be used here in a way that requires explanation. This phrase is used where "other variables" appear in the regression equation. To remove the effects of "other variables" from a given regression coefficient means that the given regression coefficient appears in an equation in which "other variables" also appear. It is the presence of the other variables in the regression equation that remove their effect from the given regression coefficient. Conversely, if a given variable does not appear in a regression equation, the regression coefficients in this equation contain the effects of that variable. The theory developed in the mathematical appendix shows that the regression coefficient depends upon the relationship between a given independent variable and the dependent variable after the effects of all other independent variables have been

removed. To be more exact, this is the type of relationship measured by the Student's t test. The coefficient of partial correlation measures roughly the same thing. The algebraic sign of the coefficient of partial correlation and the regression coefficient are the same where the same variable is under consideration. The simple correlation coefficient, on the other hand, does not have the effects of the other variables removed.

The simple correlation coefficient between cost per revenue ton-mile and length of the flight was found to be -.81. But this coefficient does not tell how much of the relationship between cost and the length of the flight is due to other factors, such as load factor or capacity of the plane. When the effect of the other six factors was removed, the correlation between cost per revenue ton-mile and length of the flight changed to +.06. This implies that as the length of the flight increases, cost per revenue ton-mile also increases. As previously indicated, this conclusion is not too unreasonable when it is remembered that the effects of planes and load factors are removed. If an attempt is made to fly a small plane beyond a certain minimum distance, the cost per revenue ton-mile does increase. (See Chapter IV). Statistically, the partial correlation +.06 appears to be insignificant. Consequently it is unnecessary to explain why cost should increase as the length of the flight increases.

American Airlines has advanced the theory that the length of the flight is responsible for an increase in the speed of the

plane and that this is one of the reasons the length of the flight is important. An experiment was made to test this theory. The regression coefficient for speed of the plane, where the effect of five other variables was removed, amounted to -.32. The effects of the following five variables were removed: capacity of the plane; ton-mile load factor; flight time; metropolitan population served; size of the firm. With the effect of these variables removed and with a regression coefficient of -.32, the experiment attempted to show the change in regression coefficient that would result from removing the effects of the length of the flight. When the effect of the length of the flight was removed, the regression coefficient for the speed of the plane changed from -.32 to -.77. The amount of change was statistically insignificant. If this had been a meaningful correlation, it would have shown that the length of the flight caused less correlation between the speed of the plane and cost, not more correlation as American Airlines had hypothesized. Another experimental approach was made to the same problem. The effects of the five variables mentioned above were removed from length of flight. The regression coefficient for length of the flight amounted to .06. When the effect of the speed of the plane was removed, the regression coefficient changed to .23. This change was also statistically insignificant. These experiments seem to indicate that the speed of the plane does not have a significant effect upon cost, and that it does not significantly effect the correlation between

length of the flight and cost.

The data presented up to this point has indicated that there are only three significant variables. Although some of the other variables may be significant, they do not appear to be important in comparison with the variables singled out as most important. A closer examination of those three variables should give a more accurate picture of the industry.

II. THREE INDEPENDENT VARIABLES

Computation. If all variables except three are ignored, a new least-squares equation results. This equation will give regression coefficients which are slightly different from the regression coefficients for the same variables previously computed. Because these differences are small and statistically insignificant, details of computation will not be given here. Table VI shows the results of fitting a least-squares equation where only three variables are involved. The regression coefficients are not significantly different from those previously computed in Table V. The reasoning behind shifting the decimal point of regression coefficients was presented in Chapter V. The decimal point of the regression coefficient applicable to the load factor shown in Table V was not shifted to compensate for the fact that the logarithm of this data had been divided by 100. In Table VI, this coefficient was divided by 100.

Omission of four variables. In Table VI, the three variables

TABLE VI

STATISTICS ON THREE VARIABLES

	Ton-mile Load Factor	Capacity of the plane (in tons)	Plane Utilization (minutes per day)
Regression coefficient	-.02278398	-.376317	-.559280
Standard deviation of regression coefficient	.096	.004	.189
Confidence limits of regression coefficient	.014 to .030	.18 to .58	.2 to .9
Coefficient of partial correlation	.733	-.607	-.500
t ratio	5.54	3.92	3.96

Source: Computations

were selected from the seven variables given in Table V. Each of the four variables omitted was found to be insignificant. A question arises as to whether or not the simultaneous omission of four variables might not be significant. In other words, there is a possibility that the elimination of four independent variables might decrease the ability of an equation to make predictions and that this decrease might be more than could be accounted for by chance. A statistical "F" test for this possibility is given in equation (33) of the mathematical appendix. This test takes into consideration the possibility that three variables chosen as

TABLE VI

SUMMATION OF CHECK VARIABLES

	Ton-mile Load Factor	Capacity of the plane (in tons)	Plane Utilization (minutes per day)
Regression coefficient	-.0227838	-.376317	-.559280
Standard deviation of regression coefficient	.095	.004	.189
Confidence limits of regression coefficient	.014 to .030	.12 to .58	.2 to .9
Coefficient of partial correlation	.723	-.607	-.500
t ratio	5.58	3.02	3.56

Source: Computations

significant might have been chosen by chance. Out of seven variables, there is a fairly good chance that some might have been high because of the operation of chance. Applied to this data, the *F* ratio amounts to 1.43. To be significant at the 5% level, this ratio would have to amount to 2.80.[4] All this means

---

[4] Paul G. Hoel, Introduction to Mathematical Statistics (New York: John Wiley & Sons, 1947), p. 250

---

that the omission of four variables has not changed the residue (error of the equation of prediction) by more than could have occurred by chance. This is another way of saying that the omission of the four variables was unimportant. The equation in three variables remaining is as good as the original equation in seven variables. This concept simplifies both theory and prediction.

Comparative importance of variable. It might be supposed that the regression coefficients would give an indication of the relative importance of each variable. Unfortunately, this is not true. The magnitude of a regression coefficient depends upon the magnitude of the type of unit in which a variable is presented. For example, had the data for plane utilization been presented in hours instead of minutes, the regression coefficient would have been 60 times as large. Since there is no logical way in which miles can be compared with time, comparisons of regression coefficients are meaningless.

The coefficients of partial correlation, where the effects of two other variables are held constant, are given in Table VI. These coefficients indicate that capacity of the plane is more important than plane utilization and that load factor is more important than capacity of the plane. The coefficients of partial correlation do not give a comparative measurement of the importance of each variable. Because each respective t ratio has the same base, the t ratio gives a more comparable measurement of the importance of each variable. The meaning of this ratio is explained in the mathematical appendix. On the basis of the t ratio, capacity of the plane is 1.3 times as important as plane utilization.

Suppose the ton-mile load factor is omitted from the least-squares equation. The t test gives the level at which the residue would be effected by this omission. On the other hand, the effects of both capacity of the plane and plane utilization may be removed from the equation. In this case, prediction would depend upon ton-mile load factor only. The question may then be asked, how does the residue compare in these two cases. This experiment was made in terms of correlation. The correlation between cost per revenue ton-mile and load factor was previously found to be -.92. The multiple correlation coefficient between cost per revenue ton-mile and the two factors, capacity of the plane and plane utilization, was found to be -.92. In other words, the two factors when taken together were the equivalent of cost per revenue ton-mile

taken singularly.

Multiple correlation. The multiple correlation for this equation in three independent variables amounts to .96. By Snedecor's table[5] this equation would have reached the 1% level

---

5 George W. Snedecor, Statistical Methods, (Ames, Iowa: Iowa State College Press, 1940), p. 286

---

of significance with a correlation of .573. This indicates that the correlation here obtained is significant far beyond the 1% level.

A multiple correlation of .96 indicates that 92% of the variation in cost is due to the three factors here analysed. The remaining 8% is due to other factors. The other factors may consist of managerial skill, variations in accounting methods, variations due to regional differences in cost, etc. It does not appear from this figure that variations in accounting methods, the equivalent of errors in observations, are sufficient in magnitude to invalidate the use of least-squares techniques. For this reason, little consideration was given to the subject of errors in observations. On the basis of this correlation, a rough indication of the value of this equation for prediction purposes may be gained from the assertion that it will give a 73% improvement over chance. This statement is based on the assumption that a prediction arising from no improvement over chance would be the mean. If an individual wished to make predictions but had no

equation for this purpose, his best prediction would in every case be the mean. The technical effect of the multiple correlation coefficient appearing here is to reduce the error of estimate by 73%.[6]

---

6 Quinn McNemar, Psychological Statistics (New York: John Wiley & Sons, 1950), p. 112

The equation for prediction. When the three regression coefficients previously presented are placed in a least-squares equation, the following equation emerges.

$$Y = e^{9.25-.02t_R-.4_S-.6}$$

In this equation, Y is the cost per revenue ton-mile (in cents); e is the base of natural logarithms; R is the capacity of the plane (in tons); S is a measure of plane utilization (average flight time per plane per day); and t is the ton-mile load factor (in percentage). Where the universe is limited to the data in this study, the exponents are accurate to three decimal places. Table VII gives a comparison of the predicted costs and the observed costs for the year 1951. No comparisons were made based for years other than 1951. In consideration of the wide confidence limits that exponents of this equation evidence, it is to be expected that occasionally large errors in prediction will occur.

Validity of equation. This equation seems to be more valuable for theoretical purposes than it is for making predictions.

TABLE VII

SELECTED COMPARISONS OF ESTIMATED
AND
OBSERVED COSTS PER REVENUE TON-MILE

Company	Cost Estimated by Equation	Cost Reported by Company	Error of Estimation	Percentage of Error
All American	111.0	116.3	5.3	4.5
American	38.4	43.0	4.6	10.7
Bonanza	150.0	159.7	9.7	6.5
Continental	69.8	73.6	3.8	5.1
Eastern	46.4	46.7	.3	.1
National	46.3	48.0	1.7	3.5
Trans-World	43.1	48.7	5.6	11.5
United	45.4	43.7	1.7	3.9
Wiggins	655.0	938.2	283.2	30.2

TABLE VII

SELECTED COMPARISONS OF ESTIMATED
AND
OBSERVED COSTS PER REVENUE TON-MILE

Company	Cost Estimated by Equation	Cost Reported by Company	Error of Estimation	Percentage of Error
All American	111.0	116.3	5.3	4.5
American	98.4	93.0	4.6	10.7
Bonanza	150.0	159.7	9.7	6.5
Continental	69.8	73.6	3.8	5.1
Eastern	46.4	46.7	.3	.1
National	46.3	48.0	1.7	3.5
Trans-World	43.1	48.7	5.6	11.5
United	45.4	43.7	1.7	3.9
Virginia	653.0	998.2	289.2	30.2

The computations leading to this equation suggest that four factors do not play an important part in airline economics. These four factors are: flight stage, size of the firm, speed of the plane, and magnitude of metropolitan population served. This equation shows that the most important factors are ton-mile load factor, capacity of the plane, and the frequency with which each plane is used. Although the confidence limits for the exponents could have been improved by taking larger samples, the overall effectiveness of the equation is high. Only 8% of the variation in cost is unaccounted for, and the equation represents a 73% improvement over chance. This is sufficient reason for examining further the implications of the equation.

Further predictions. It is realized that there is no mathematical justification for using a least-squares equation outside of the range of observations from which the equation was constructed. However, in this case there seems to be a fairly good reason for assuming that the equation will continue to hold true outside this range. This equation is based on load factors from within the range of approximately 20 to 70. Considering that the equation holds true over this range, it would not seem too great a departure from the probable to ask what would happen should the load factor be extended to 90. This would suggest that equipment larger than the average used by most companies could be used. It should generally mean a greater number of average hours of plane utilization. To make the results of the

higher load factors reasonable, the highest plane capacities and number of minutes of plane utilization were used. The capacity of equipment used was 7.39 tons, which was less than that reported for 1951 by Northwest Airlines. The estimate was based on the number of hours of plane utilization experienced by Inland for the year 1951, which was nine hours and 23 minutes. This was less than the average number of hours Eastern Airlines utilized its aircraft. Based on these figures and the load factors indicated, the equation gives the costs shown in Table VIII.

TABLE VIII

ESTIMATED COSTS AT HIGH LOAD FACTORS

Load Factor	Cost per R.T.M.
70	28.7
75	25.4
80	22.6
85	20.0
90	17.9

It will be noted that the cost in this estimate for a load factor of 70 is considerably below that experienced by American Airlines, which operated during 1951 with an average load factor of almost 70. The reason that American Airlines had such high costs was that its equipment was of low capacity, in contrast with the equipment assumed in this estimate. Also, American Airlines used its equipment an average of only 6 hours per day, while the estimate was based on a use of equipment for 9 hours per day. This equation indicated that American would have had considerably

lower cost had it used larger equipment and used its equipment more effectively. This study has not, of course, taken cognizance of the fact that certain characteristics of American's route structure may have made the use of larger equipment impractical.

The principal way in which the findings of this study differ from the findings of American Airlines, as shown in Table I, is in the emphasis on the significance of the length of the flight. This difference will receive further attention in the next chapter.

CHAPTER IV

LENGTH OF THE FLIGHT

The statistical work previously described seems to indicate that the length of the flight is insignificant. Before this conclusion is accepted, it is necessary to inquire whether or not it is reasonable. It is necessary to explain why this study disagrees with the findings of American Airlines. A criticism of the application of accounting methods to the solution of economic problems will first be attempted. Certain pertinent conditions of the airline industry will be described. Last, a direct criticism of the findings of American Airlines will be made.

It will be indicated that American Airlines failed to compare the relative importance of reduction in cost due to increase in the size of equipment with the relative importance of reduction in cost due to an increase in the length of the flight. American Airlines failed to consider that long cargo hauls may often force a company to operate 24 hours per day, while a short-haul company may be able to operate during the daylight hours only. American Airlines failed to consider that the cost of operating each station decreases as the length of the flight decreases.

I. ACCOUNTING

The logic of accounting. The logic in accounting is a

reflection of custom. As Dewey has indicated, all logic tends to be a reflection of custom. Primitive man used the logic of his day to show that each tree has a soul. The Greek philosopher used the logic of his day to show that each type of living being is fixed and that mutation is impossible.[1] Both of these ideas

---

1 Eugene G. Bewkes and others, A Survey in Philosophy and Religion (New York: Harpers and Brothers, 1940), p. 258

---

are erroneous in the light of present-day knowledge. Had the logic of primitives never been subject to question, we would still be concerned about the soul of every tree. Just as primitive logic had its limitations, so does accounting logic have its limitations. Just as it was difficult for a primitive to find the error in his logic, so is it difficult for an accountant to find the error in his logic. Just as the logic of the primitive was useful in that it helped to hold the tribe together and give them common understanding, so is the logic of the accountant useful in that helps business men to understand each other.

The accountant is highly trained in the logic of accounting. As long as the accountant thinks within the boundaries of every-day business situations, his logic is useful. Similarly, the logic of the primitive is useful as long as he lives under primitive conditions. When he attempts to explain an airplane in terms of primitive logic, his logic breaks down. When an accountant departs from the situations for which his thought patterns were

designed, his tools for thinking are no longer reliable.

Accounting is inadequate to identify cause. Accounting as a tool of logic was not designed to identify cause. When accounting is used for this purpose, it is no longer serving its traditional function and for this reason the logic of accounting techniques is faulty.

In the first place, the accountant assumes that the cause of every expenditure is known at the time it is made. For example, he might suppose that an increase in the cost of a pilot is due to the fact that the pilot flew short flights and spent more time on the ground. In making this decision he is using his accumulated experience, or logic. It seems axiomatic that if accountants, or management, knew the answer to this question when the study began, there would be no reason for making the study. All distribution of such cost is based on the assumption that the accountant can look at each individual item of cost and know the reason for that item. In the case of the increase in pay of a pilot, this ignores the fact that in the long run the pilot might prefer to spend more time on the ground and that there might be a tendency for pilots to work for less where the flight is shorter. It is highly unlikely that the accountant will investigate this possibility. In short, the accountant is ignoring many circumstances which attend the short flight. The reduction in salary which the pilot is willing to undergo was not cited as a realistic explanation. It was cited only to show

the arbitrary manner in which the accountant makes decisions about causation. There are numerous ramifications to changes in the length of the flight. Management is aware of some of these, and it seems likely that there are relationships of which management is not aware. For this reason, the best measure of what happens in a business is to be found by observing what happens under a given circumstance rather than by manipulating figures so that they will conform to a preconception.

Another example of circumstances under which accounting is inadequate to determine cause is to be found in the way in which overhead expenses are prorated. Suppose the accountant must decide the basis on which expenses will be prorated to Departments A and B. Departments A and B accrue the same number of man hours during the month, but the wages per hour in Department A are higher than they are in Department B. If the firm decides to prorate overhead on the basis of departmental wages, this will mean that Department A will be charged more than Department B. If overhead is prorated on the basis of man hours, both departments will receive an equal amount of overhead. The conclusion that costs are greater in Department A than in Department B is no more valid than the assumption on which overhead is prorated. There is no specific accounting technique to decide such a question. The method of correlation analysis is, of course, available. If it should be determined that the correlation between wages and overhead is greater than the correlation between man hours and overhead, the accountant might have sufficient reason for using

wages rates. Suppose, however, that the accountant should consider ten factors, instead of two. Out of these ten factors, the accountant might choose man hours because it appeared to correlate well with cost. As shown in earlier sections, it might well be that some other factor is responsible for changes in man hours, and that the correlation between overhead and man hours would actually be negative after the effects of other factors were removed. In many cases, the accountant cannot afford to use multiple correlation or he does not realize the need for it. If he has learned that either man hours or the length of the flight is important, he investigates this variable by simple correlation, finds an apparently high correlation, and apportions his costs accordingly. His apportionment of costs can never be better than the investigation which he used as a basis for determining the manner in which apportionment should be made. For this reason, if costs are apportioned by accounting methods, there is no particular reason to assume that such apportionment has any resemblance to cause.

This argument has indicated that there are two possible ways of determining cause by accounting methods — by allocating each individual expense among certain classifications or by apportioning a group of expenses among these classifications. If individual allocations are used, these represent merely the opinion of the accountant. Such allocations are subject to all the cultural bias that the accountant may have accumulated in his association

with other accountants or with business men. If it is assumed that the accountant knows the cause of all variations of expense in advance, there is no reason for the accounting study. At the best, the study shows only the carefully weighed decision of the accountant. Considering that the accountant is a specialist in custom, it is highly likely that such a study is nothing more than reflection of custom. On the other hand, if the study by American Airlines represents an apportionment of certain groups of expenses, the basis of the apportionment is no more valid than the study which justified this apportionment. Since American Airlines gave no evidence of a study on which to base apportionment, there is no reason why apportionment should be accepted. Since there is no reason to believe that the distribution of individual expenses represents more than custom and since there is no reason to believe that the apportionment of expenses in groups has empirical validity, there is no reason why the study of American Airlines deserves serious consideration.

Statistics versus accounting. Accounting and statistical methods have one feature in common. Under both methods, assumptions must be made about causality when the study begins. Under both accounting and statistical methods, the investigator must choose the variables, or expense classifications that he wishes to consider. Both methods of determining causation assume that the investigator, or accountant, has knowledge about the subject adequate to choose the correct classifications. Note the tremen-

dous difference in method that exists after this one choice has been made. In regression analysis, the question of normality of distribution and level of significance give the investigator a small margin of choice. For the most part, whether or not the outcome of a statistical experiment will show a given variable as significant depends upon the data available. In accounting, the amount of human choice has only begun when the classifications have been chosen.

In accounting, there are frequently three or more different methods by which the same object may be accomplished. For example, inventories may be evaluated on the basis of first-in-first-out, last-in-first-out, or average cost. If drapery is a part of a building, it may be capitalized; while if it is not attached to a building, it is an expense. The accountant must decide on whether or not it is a part of the building. Depreciation may legitimately be made on a number of different bases and considerable variation in the useful life of the property is possible. Variations in the depreciation of aircraft were so great that the Civil Aeronautics Board made its analysis of cost standards without reference to depreciation.[2] This great variation in

---

2 Civil Aeronautics Board, Cost Standards - Domestic Scheduled Air Carriers (Washington, D. C.: CAB, July 1950), p. 40

---

methods means that the accountant has ample opportunity to unconsciously adopt procedures that will verify his preconceptions.

---

2 Civil Aeronautics Board, Coast Standards - Domestic Scheduled Air Carriers (Washington, D. C.: CAB, July 1959), p. 40

It should be pointed out, however, that under ordinary circumstances the accountant's adherence to custom is an advantage. Under ordinary circumstances, the business man is not utilizing expense distribution to determine the unknown -- causation.

In the light of the foregoing discussion, it appears that accounting studies for the purpose of determining causation are primarily reflections of group attitudes. The science of multiple correlation is subject to cultural bias only in the selection of variables. In the light of this consideration, statistical correlation as a method of investigation seems to have every advantage over accounting methods. A statement of the Civil Aeronautics Board is pertinent to this question.

> There would seem to be no reasonable choice between standards based, as here, upon statistical correlation and standards developed by "some other method." The common experience of the carriers affords the only reliable evidence of validity.[3]

---

3 Ibid., p. 13

It seems likely that "some other method" applies to accounting methods. Much of the outcome of the issue here under debate depends upon whether the investigator prefers to accept findings made by accounting methods or by statistical methods. The issue is whether or not the investigator prefers the empirical, scientific methods of statistics or *a priori* opinions which reflect group attitudes.

It should be pointed out, however, that under ordinary circumstances the accountant's adherence to custom is an advantage. Under ordinary circumstances, the business man is not utilizing expense distribution to determine the unknown -- causation.

In the light of the foregoing discussion, it appears that accounting studies for the purpose of determining causation are primarily reflections of group attitudes. The science of multiple correlation is subject to cultural bias only in the selection of variables. In the light of this consideration, statistical correlation as a method of investigation seems to have every advantage over accounting methods. A statement of the Civil Aeronautics Board is pertinent to this question.

> There would seem to be no reasonable choice between standards based, as here, upon statistical correlation and standards developed by "some other method." The common experience of the carriers affords the only reliable evidence of validity.[3]

---

3 Ibid., p. 19

It seems likely that "some other method" applies to accounting methods. Much of the outcome of the issue here under debate depends upon whether the investigator prefers to accept findings made by accounting methods or by statistical methods. The issue is whether or not the investigator prefers the empirical, scientific methods of statistics or <u>a priori</u> opinions which reflect group attitudes.

Due to the fact that conclusions made by accounting methods depend upon the attitudes and previous experience of the accountant, it is doubtful if accounting often leads the investigator to change his opinion. Correlation analysis may force the investigator to change his opinion. This thesis was commenced with the purpose of proving that the length of the flight is extremely important in airline economics. When the statistical computations indicated otherwise, a strong urge to destroy the computations and start anew was felt. Only after considerable thought and analysis was the validity of the computations accepted.

This section has sought to show that accounting, as a method for determining causation, is inadequate. Some specific conditions of the airline industry that are not immediately obvious from accounting studies will next be presented.

II. CONDITIONS OF THE INDUSTRY

Scope of this study. The data used in this thesis is based on both trunk lines and feeder lines. The reason for basing the study on both types of firms was to investigate the effect of increasing the length of the flight from an extremely short average flight to a comparatively long average flight. Had the study been confined to one type of company, the variation in the average flight stage would not have been sufficient to give a meaningful picture. Such a study faces a practical disadvantage. Conditions in the feeder lines are different from conditions in

the trunk lines. Instead of measuring the effects of an increase in the length of the flight, a statistical study may measure the difference in conditions prevailing in the two industries. This has the practical advantage of providing the background for a theory that applies to the entire industry. It may be reasoned that conditions of the carriers gradually change as the size of the carrier increases, load factor increases, etc., so that the difference between trunk and feeder lines is only a special case of the graduated differences existing within each division.

Distance of Transportation. The trunk lines are engaged in the transportation of cargo for long distances. Much of the cargo of the trunk lines must move from one coast to the other. This makes it both necessary and desirable that a trunk line should have comparatively few stops, that cargo should move as directly as possible. The feeder lines, on the other hand, are characterized by comparatively short interstation distances. During 1949 the average length of interstation flight distance for 10 feeders was 74 miles. During this period the average length of interstation flight distance for 15 trunks was 175 miles.[4] Since the average length of the flight of the trunk

---

4 Computed from data given by the Civil Aeronautics Board, Ibid., p. 81

---

lines is almost 2½ times that of the feeders, a comparison of the operating characteristics of the two divisions for purposes

of studying the length of the flight should be valid.

Effect of through cargo upon the intermediate station. Passenger cargo constitutes a major part of the load carried by the trunk lines and is easily the determining factor. The time of departure and arrival must be adjusted to the passenger's convenience. It is convenient and desirable for passengers to leave Chicago in the evening and arrive on the West Coast the next morning. Unfortunately, not many passengers desire to begin a trip late in the evening, and there is a smaller number who are willing to catch a plane after midnight. However, the through plane must stop at intermediate points to discharge passengers, refuel, and pick up a few passengers. This necessitates the maintenance of an intermediate station which is equipped to handle the maximum number of passengers that may appear at any time. It means that a radio station must be kept in operation throughout the night to maintain communication with craft in flight. This has the effect of requiring the operation of an intermediate station 24 hours per day. Of particular importance in the utilization of crews is the fact that two or more planes may arrive at the same time. This is partially responsible for the large number of personnel that must be available. Although it is true that the airports are usually operated by municipalities, each airline operates its own trucks, ramps, ticket counters, etc. The Air Transport Association has found that each plane is usually

met by five to eight of the airline terminal personnel. The average length of time between the arrival and departure of planes amounts to 40 minutes. The ground crew is busy only one-third of the time.[5] The fact that a crew is required to operate a

---

5 Joseph L. Nicholson, Air Transportation Management (New York: John Wiley & Sons, 1951), pp. 117-118

---

station for a small number of flights means that there is no great saving in cost to be achieved from failing to stop at that station on certain flights.

The feeder-line schedule. The feeder lines operate under conditions quite different from the trunk lines. The theory of the feeder lines has been set forth by the Civil Aeronautics Board. One purpose of these lines is to permit the population of small towns to travel to larger towns in the morning and return in the afternoon.[6] For this reason, most feeder lines operate

---

6 Ibid., p. 88

---

during the daytime when transportation from these smaller towns is desirable. The fact that it is practical to operate feeder lines primarily during the daylight hours has important implications so far as cost is concerned. It means that the feeder lines can operate an intermediate station for comparably less than the trunk line can operate such a station. The advantage from a cost stand-

point of limiting operations to certain hours of the day has been recognized by trunk lines. It has been suggested that American Airlines reached a low-cost position during the years 1949 and 1950 because flight equipment was withdrawn during the early morning hours.[7] As Josh Lee of the Civil Aeronautics Board has

---

7 Ibid., p. 237

---

suggested, a town which would be a loss to a trunk line whose scheduled planes came through at 2 A.M. might be a profit to a feeder line.[8] It seems that the advantage the feeder line has in

---

8 Ibid., p. 88

---

lower station cost is offset by the disadvantage of lower load factors, smaller capacity of equipment, and a smaller number of hours per day during which equipment may be utilized.

Difficulty of proving effect of long flight. It is difficult to demonstrate the effect of the length of the flight upon cost without using multiple correlation or regression methods. If the investigator is not prepared to accept the outcome of a statistical experiment such as that described in this monograph, he may be inclined to accept some of the group thinking about the length of the flight that was previously described. The difficulty with using ordinary methods of investigation is that there is no way to show how much of the variation in expense is due to load

point of limiting operations to certain hours of the day has been recognized by trunk lines. It has been suggested that American Airlines reached a low-cost position during the years 1949 and 1950 because flight equipment was withdrawn during the early morning hours.[7] As head of the Civil Aeronautics Board has suggested, a town which would be a loss to a trunk line where scheduled planes came through at 2 A.M. might be a profit to a feeder line.[8] It seems that the advantage the feeder line has in

---

7 Ibid., p. 297

---

lower station cost is offset by the disadvantage of lower load factors, smaller capacity of equipment, and a smaller number of hours per day during which equipment may be utilized.

Difficulty of proving effect of long flight. It is difficult to demonstrate the effect of the length of the flight upon cost without using multiple correlation or regression methods. If the investigator is not prepared to accept the outcome of a statistical experiment such as that described in this monograph, he may be inclined to accept some of the crude thinking about the length of the flight that was previously described. The difficulty with using ordinary methods of investigation is that there is no way to show how much of the variation in expense is due to load

factor, and other causes, and how much is due to length of the flight. As previously mentioned, the effects of load factor may be removed by studying cost per available ton-mile instead of cost per revenue ton-mile. This study is not satisfactory, however, because the firms with large load factors tend to be the firms which have long flights. These large load factors are associated with large equipment. There is no simple way of divorcing the cost reduction that accrues from the use of large equipment from the cost reduction that results from the longer flight. Cost per plane mile is no answer to this question because of the fact that it costs so much more to operate a large plane than it does to operate a small plane. If cost per plane mile is considered, firms which operate the smaller planes will have a tremendous advantage. Because of the difficulty of extracting information about the length of the flight from measurements such as those here described, this study directed attention to the cost of operating each station.

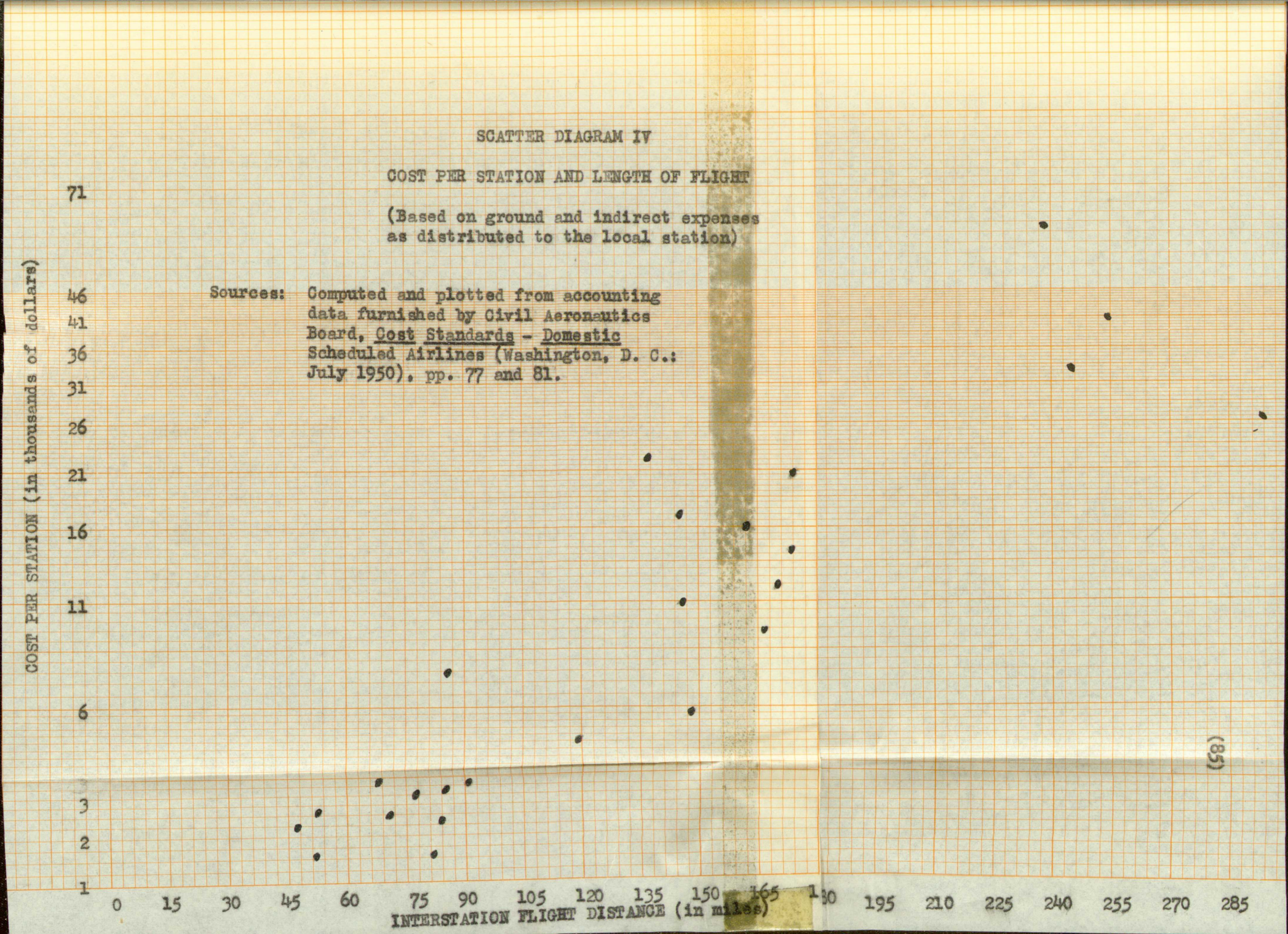
Cost of station operation. The cost of station operation is not an ideal measurement for investigating the length of the flight, but it appears to be as efficient as any other simple measurement. The question is, what effect does a reduction in interstation flight distance have upon the cost of operating each station? Scatter Diagram IV provides a simple and direct approach to this question. This diagram shows that as the length of the interstation flight distance increases, the cost per station

SCATTER DIAGRAM IV

COST PER STATION AND LENGTH OF FLIGHT

(Based on ground and indirect expenses as distributed to the local station)

Sources: Computed and plotted from accounting data furnished by Civil Aeronautics Board, Cost Standards - Domestic Scheduled Airlines (Washington, D. C.: July 1950), pp. 77 and 81.



increases. The logarithm of the increase in cost per station is proportional to the increase in flight distance. This means that when comparatively long flights are involved, an increase in flight distance will be associated with a proportionately greater increase in cost per station. For example, if the flight distance is 200 miles, an increase in flight distance of 50 miles (25%) will be associated with an increase in cost from $21,000 to $41,000 (almost 100%).

The expenses in Diagram IV represent approximately 25% of all operating expenses. This data was obtained by distributing general and overhead expenses between two classifications -- local stations or overhead expenses. General overhead expenses represent slightly more than 50% of all operating expenses. Local station expenses represent approximately 50% of ground and indirect expenses. In general, local station expenses represent a higher percentage of expense for the trunk lines than they represent for the feeders.

It is estimated from Diagram IV that for a firm with flight distances comparable with American, a reduction of 50% in interstation flight distance will be associated with a reduction of 85% in cost per station. With this estimate, consider the effect of adding an intermediate station which would reduce interstation flight distance by 50%. It would mean that the cost of operating the two terminal stations would amount to only 15% of their previous cost. The intermediate station would cost 15% of the cost of one terminal station. An overall cost reduction of at

least 75% would apparently result. The reason for this cost reduction is that companies with longer flight distances tend to have the larger load factors. A reduction in flight distance is generally associated with a comparable reduction in the number of passengers carried and a resulting reduction in operating expense. It seems likely that when the variations in load are accounted for, the cost of operating each station will vary in proportion to interstation flight distance. This is the implication of Diagram IV.

What is implied by the observation that cost per station varies in proportion to interstation flight distance? It implies that a certain unit cost of operating each station is associated with the distance between stations. As the distance between stations decreases, the operating cost per station decreases proportionately. Another way of saying this is that certain costs are apportioned on the basis of distance. If the distances between stations is great, a large burden of costs falls upon each station. If the distance between stations is small, a comparatively small burden of costs falls upon each station. This is definitely the implication of Diagram IV. It follows from this that a reduction in the distance between stations will be associated with a comparable reduction in station cost. The elimination of a station will not have the effect of reducing over-all expense. Such an elimination will cause the cost of operating each remaining station to be increased proportionately. So far as the cost of operating stations is concerned, the firm with the shorter flight has a distinct advantage. The reason for this advantage will next

be shown.

The effect of the length of the flight upon cost can be understood further by analysing station cost in terms of tons lifted. Table IX is presented for this purpose. Table IX shows that the cost per ton lifted is greater for the trunk lines, with the longer flights, than it is for the feeder lines. This data supports the argument previously made to the effect that station cost is greater for the trunk lines than it is for the feeder lines.

TABLE IX

COST OF TONS LIFTED FOR FEEDERS AND TRUNKS

	Feeders	Trunks
Station Expense, Ground and indirect	$2,597,926	$102,793,924
Station Expense, Ground Operations	1,533,102	47,591,071
Total Local Station Expense	$4,131,028	$150,384,995
Tons lifted	173,914	3,829,712
Cost per ton lifted	$23.75	$39.29

Ratio of cost of trunks to cost of feeders — 1.66

Note: This data represents the average of 10 trunk lines and 15 feeder lines, for year ending September 30, 1949.

Source: Civil Aeronautics Board, Cost Standards - Domestic Scheduled Air Carriers (Washington, D. C., July 1950), pp. 77 and 81

Although Table IX does show that short-flight firms enjoy an advantage so far as station costs are concerned, this advantage is underestimated. Suppose the feeder lines should operate their stations and equipment as near to capacity as the trunk lines operate. This would involve comparatively little increase in cost. It would, however, mean a tremendous increase in tonnage. If present costs were spread over this tonnage, the feeder lines would have an even greater advantage in comparison with the trunk lines. This increase in tonnage might cause the costs, per ton lifted, of the feeder lines to shrink to one-third of the cost experienced by the trunk lines.

III. THE COST STUDY BY AMERICAN AIRLINES

Techniques of investigation inadequate. The study made by American Airlines indicates that as the length of the flight increases, cost of operations are substantially reduced. This study adheres to principles that seem quite logical in the light of cost accounting techniques. It was pointed out that there is no empirical reason for accepting the cost allocations made by American Airlines. There is no indication that American Airlines made a statistical investigation that would justify the apportionment of expenses in groups. Apparently most expenses were individually allocated by accountants. This was equivalent to assuming that the accountant knew all the complicated interrelationships of expenses before the study was commenced. In the light of these considerations, there is little, if any

reason for accepting the study by American Airlines.

Conditions of the industry not considered. In addition to the use of questionable methods, the study by American Airlines failed to consider certain conditions of the airline industry. This study failed to consider that a firm which operates under conditions of long haul must operate on a 24-hour basis. This creates a disadvantage from a cost standpoint. The study by American Airlines failed to consider that a trip which is profitable for the feeder lines may be unprofitable for trunk lines. It was previously shown that the cost of station operation tends to decrease as the length of the flight decreases. The failure to consider these conditions of the industry invalidate the study by American as far as its application to the entire industry is concerned. These conditions throw considerable doubt on the validity of the study so far as it is applicable to American Airlines. If other firms tend to have a decrease in station expenses as the length of their flight decreases, it is difficult to see why this condition would not hold true with American Airlines. To show that this principle does not hold true with American Airlines, the company would need to prove that conditions with American are essentially different from conditions among other airlines.

American Airlines versus the Civil Aeronautics Board. The study on cost standards made by the Civil Aeronautics Board[9]

---

9 Civil Aeronautics Board, op. cit.

---

used statistical methods instead of accounting methods. This study did not consider the theoretical significance of the capacity of equipment. In the absence of this consideration, it seems natural that the length of the haul and length of the flight would both be considered important. Also, the purpose of this study was to provide an equation to predict the results of a merger. The tools for this purpose would not necessarily be concerned with causation. In spite of the fact that the length of the flight was accepted as a factor in making predictions by the Civil Aeronautics Board, the study by this board seemed to show that the length of the flight was of remarkably small importance.

The study by the Civil Aeronautics Board showed the results of eliminating five stations. Two companies were given as examples. Stations having the smallest amount of traffic were selected. The elimination of the five smallest stations of Western Inland reduced operating expenses per traffic ton-mile from 58.60¢ to 57.89¢ per traffic ton-mile. The elimination of the five smallest stations of Pioneer Airlines increased operating expenses from 95.84¢ to 97.77¢ per traffic ton-mile. All of the stations eliminated were intermediate stations. They also represented minimum traffic stations.[10] The study by the board contains the following comment:

---

10 Ibid., pp. 35-37

> Tests of the probable effect of remedial action which might be taken respecting

> notably uneconomic route systems seems to indicate total discontinuance or merger as the courses offering the possibility for the most substantial saving. Discontinuance of service to certain stations of very light traffic also offers opportunity for a degree of saving, but less, perhaps, than might be expected.[11]

---

11 Ibid., p. 31

Under the circumstances described here, the length of the flight seems to have little significance. As previously mentioned, The Civil Aeronautics Board did not find it expedient to consider the importance of capacity of equipment. Had they done so, they would have found the length of the flight of even less importance. As it stands, the study made by the Civil Aeronautics Board seems to indicate that the significance of the length of the flight has been exaggerated by American Airlines.

Type of equipment. The study by American Airlines was based on two planes, the Convair 240 and the DC-6B. Both of these are large planes and are designed for long trips. It is only natural that they should show the most efficient operating costs as the trip becomes longer. Another plane would have given entirely different results. Consider for example the plane which gave rise to the data given in Table X. For this plane, these figures are just as authoritative as those given by American Airlines. The reason that this study is exactly opposite to the findings of American's study is that the capacity of this plane is small so

that the increase in the length of the flight necessitates a proportionately larger amount of gasoline. This reduces the space available for passengers or other cargo. This is responsible for a decrease in load and in turn an increase in operating costs per ton-mile.

TABLE X

RELATIONSHIP OF OPERATING COST PER TON-MILE TO LENGTH OF FLIGHT

Length of Flight (miles)	Estimated Total Operating Cost per Ton-Mile (cents)
600	22
1,000	24
1,500	26
2,000	30
2,500	36
3,000	47
3,500	75

Source: Joseph L. Nicholson, Air Transportation Management (New York: John Wiley & Sons, 1951), p. 196

The foregoing table should demonstrate that costs do not necessarily decrease as the length of the flight increases. This does occur with certain planes within certain ranges. The extent to which it is general in the industry can be determined only by examining the cost figures of the industry. The question cannot be determined by examining two specific planes and making conclusions on this basis.

that the increase in the length of the flight necessitates a proportionately larger amount of gasoline. This reduces the space available for passengers or other cargo. This is responsible for a decrease in load and in turn an increase in operating costs per ton-mile.

TABLE X

RELATIONSHIP OF OPERATING COST PER TON-MILE TO LENGTH OF FLIGHT

Length of Flight (miles)	Estimated Total Operating Cost per Ton-Mile (cents)
500	22
1,000	24
1,500	26
2,000	30
2,500	36
3,000	47
3,500	75

Source: Joseph D. Nicholson, Air Transportation Management (New York: John Wiley & Sons, 1951), p. 196

The foregoing table should demonstrate that costs do not necessarily decrease as the length of the flight increases. This does occur with certain planes within certain ranges. The extent to which it is general in the industry can be determined only by examining the cost figures of the industry. The question cannot be determined by examining two specific planes and making conclusions on this basis.

Capacity of equipment not considered. Even if all the criticisms previously made of the study by American Airlines should prove erroneous, another serious limitation to American's study remains. American has failed to compare the significance of savings from the use of large planes with the savings which will accrue from operating long flights. Because of this failure American's study gives a biased picture of the significance of the length of the flight. American Airlines is somewhat in the same position as a business executive who tries to compare the cost policy of two business firms, basing his observations solely on the wages paid to janitors. Just as the wages of janitors must be considered in relation to their significance on the profit and loss statement, so must the length of the flight be considered in relation to other factors which influence cost. It is necessary to consider types of planes, load factors, number of hours per day planes are used, and any other factor which there is reason to believe important. In addition, it is necessary to consider the effect of any one of these factors upon the other factors. The study by American Airlines did not give sufficient weight to the necessity for large load factors as a prerequisite to the use of the DC-6B. This interrelationship of load factor and equipment appears to be extremely important. It is the removal of the effect of load factor from the length of the flight that causes the length of the flight to become unimportant. There is apparently no way to remove the effect of load factor from the length of the flight by the use of accounting

Capacity of equipment not considered. Even if all the criticisms previously made of the study by American Airlines should prove erroneous, another serious limitation to American's study remains. American has failed to compare the significance of savings from the use of large planes with the savings which will accrue from operating long flights. Because of this failure American's study gives a biased picture of the significance of the length of the flight. American Airlines is somewhat in the same position as a business executive who tries to compare the cost policy of two business firms, basing his observations solely on the wages paid to janitors. Just as the wages of janitors must be considered in relation to their significance on the profit and loss statement, so must the length of the flight be considered in relation to other factors which influence cost. It is necessary to consider types of planes, load factors, number of hours per day planes are used, and any other factor which there is reason to believe important. In addition, it is necessary to consider the effect of any one of these factors upon the other factors. The study by American Airlines did not give sufficient weight to the necessity for large load factors as a prerequisite to the use of the DC-6B. This interrelationship of load factor and equipment appears to be extremely important. It is the removal of the effect of load factor from the length of the flight that causes the length of the flight to become unimportant. There is apparently no way to remove the effect of load factor from the length of the flight by the use of accounting

techniques.

This discussion has shown that American Airlines made its study on the length of the flight by the use of faulty techniques. It failed to consider economic conditions of the industry. The study was made with only two planes and the effect of changes in capacity of equipment was not considered. Any one of these criticisms appears sufficient to invalidate the study made by American. In light of these considerations, it seems safe to accept the conclusion that variations in the length of the flight are not responsible for significant variations in cost per revenue ton-mile. The acceptance of this conclusion suggests that the meaning of it should be further explored. Chapter V will further explore the meaning of this conclusion. In addition, it will attempt to formulate an over-all theory of airline cost based on the findings of this study.

techniques.

This discussion has shown that American Airlines made its study on the length of the flight by the use of fuel by techniques. It failed to consider economic conditions of the industry. The study was made with only two planes and the effect of changes in capacity of equipment was not considered. Any one of these criticisms appears sufficient to invalidate the study made by American. In light of these considerations, it seems safe to accept the conclusion that variations in the length of the flight are not responsible for significant variations in cost per revenue ton-mile. The acceptance of this conclusion suggests that the meaning of it should be further explored. Chapter V will further explore the meaning of this conclusion. In addition, it will attempt to formulate an over-all theory of airline cost based on the findings of this study.

CHAPTER V

CONCLUSIONS

Previous chapters have presented a research study in the factors which determine airline cost. The purpose of this chapter is to interpret that study in the light of transportation problems. Before such interpretation is made, it seems fitting that the limitations of this study should be examined. The need for further research develops from such weaknesses. For this reason, the first section of this chapter will be devoted to the possibility of further research. With the weaknesses of the study in mind, the theoretical implications will next be presented. Finally, a theory of the way in which this study affects broad transportation policy will be formulated.

I. FURTHER RESEARCH

Repetition of this study. In spite of the fact that the results of this study show considerable agreement with other studies of the subject, there is a possibility that it contains serious error. This possibility should be investigated by a regression study similar to the one here made. Such a study should contain at least 50 observations. It would be desirable to omit certain companies from this study. For example, Wiggins Airlines operates a helicopter service. Comparison of the costs of Wiggins with those of other companies operating under normal conditions is unjustified. When the data for this thesis were

CHAPTER V

CONCLUSIONS

Previous chapters have presented a research study in the factors which determine airline cost. The purpose of this chapter is to interpret that study in the light of transportation problems. Before such interpretation is made, it seems fitting that the limitations of this study should be examined. The need for further research develops from such weaknesses. For this reason, the first section of this chapter will be devoted to the possibility of further research. With the weaknesses of the study in mind, the theoretical implications will next be presented. Finally, a theory of the way in which this study affects broad transportation policy will be formulated.

I. FURTHER RESEARCH

Repetition of this study. In spite of the fact that the results of this study show considerable agreement with other studies of the subject, there is a possibility that it contains serious error. This possibility should be investigated by a regression study similar to the one here made. Such a study should contain at least 50 observations. It would be desirable to omit certain companies from this study. For example, Wiggins Airlines operates a helicopter service. Comparison of the costs of Wiggins with those of other companies operating under normal conditions is unjustified. When the data for this thesis were

selected, considerable care was given to include all data available. In the case of Wiggins, this seems to have been a mistake. A closer observation of the operating conditions of small companies would probably reveal valid reasons why others should be omitted. In order to obtain an adequate population sample, observations should be taken from more than one year. For example, the years 1951 and 1952 might conveniently be used. This would mean that American Airlines would be represented twice, once for the year 1951 and once for the year 1952. In a repetition of this investigation, considerable time could be saved by omitting variables. In the light of the comments of Koontz, and in light of the findings of this study, there seems to be no valid reason for considering the size of the firm. In this study, the size of the firm appeared to have practically no significant correlation with cost. Furthermore, there was no logical reason for including metropolitan population as one of the variables. If metropolitan population were an extremely important factor, most of its effects would be accounted for by load factor. With the omission of the size of the firm and metropolitan population, the following variables remain for investigation: load factor; capacity of the plane; plane utilization; length of the flight; and speed of the plane. Instead of the length of the flight, the variable length of the haul might be used. Although the correlation between length of the haul and length of the flight is great, it is possible that length of the haul may correlate much better with cost per revenue ton-mile than does length of the flight. In this respect, the study here described

has a definite limitation. It applies only to the length of the flight. If the assumption is made that the length of the haul correlates more perfectly with cost than does the length of the flight, the limitations of this study become important. It can be logically assumed that the length of the haul would be affected by terminal cost while the length of the flight would not be affected by terminal cost. That is, length of the flight would not be affected by the cost of loading and unloading cargo, while length of the haul would be affected by changes in this cost.

Use of cost per available ton-mile. In the study in connection with this thesis, an experiment was made to investigate the influence of ton-mile load factor on the length of the flight. When the effects of five variables were removed (through use of these variables in the regression equation), the regression coefficient between cost per revenue ton mile and length of the flight amounted to -.15. When the effect of the sixth variable - the ton mile load factor - was removed, the regression coefficient for the length of the flight changed to +.23. The question posed is whether or not the change in this coefficient from -.15 to +.23 was due to chance. By use of the t test and the hypothesis that the true value of the regression coefficient for the length of the flight is zero in both of the above cases, the statistical conclusion may be made that both of these regression coefficients are due to chance. A method of checking this conclusion is needed. Such a method could be based on the relationship between cost per

available ton-mile and four of the factors previously mentioned. The ton-mile load factor should not appear in the investigation here recommended. More specifically, the effects of capacity of the plane, plane utilization, and speed of the plane should be removed from the regression coefficient between cost per available ton-mile and the length of the flight. As previously explained, this means that a least-squares equation should be computed in which length of the flight and the three variables mentioned above appear.

Load factor as the dependent variable. An attempt should be made to explain load factor by use of other independent variables. Although details for this experiment are not clear from the material presented in this thesis, it might be supposed that metropolitan population served is one of the factors responsible for load factor. A more important consideration appears to be the number of flights a company makes between certain cities. Koontz[1] comments on this.

---

1 Harold D. Koontz, "Domestic Air Line Self-Sufficiency", The American Economic Review, Volume XLII, No. 1., March, 1952, pp. 118-121

A measurement of the number of flights a company makes between cities with a high community interest should be effective in explaining the ton-mile load factor. Advertising expenditures might also show a correlation with load factor. In the study previously cited, the Civil Aeronautics Board showed a high correlation between cost and number of tons originated. While the number of tons originated closely resembles the load factor,

there is a possibility that it also explains the load factor. In other words, it is possible that correlation between the load factor and the number of tons originated is so high that the number of tons originated rather than the load factor should be the dependent variable.

Capacity of equipment as the dependent variable. This study does not show the exact way in which capacity of the equipment is dependent upon the length of the flight and upon load factor. Such data is desirable in order that the significance of the length of the flight may be more adequately explained. It is known that equipment of certain capacities can be used most effectively only within certain ranges. This suggests that the range within which equipment should be used might be obtained from the specifications of the manufacturer. This has a major difficulty. The manufacturer is unable to specify the importance and effect of other operating conditions upon the use of certain equipment. For this reason, it might be more desirable to make correlations between capacity of the equipment and other variables instead of accepting the specifications of the manufacturer. It is possible that the manufacturer would give the minimum range within which a given plane should be used as 300 miles. Conditions of the industry might be such that carriers would find this plane more efficient within a range of 200 miles. In the absence of data on the subject, it seems best to assert only that the capacity of the equipment is an important cost-determining factor. It must

be recognized that capacity of equipment depends somewhat upon the length of the flight, but the exact relationship of this dependence is beyond the scope of this thesis. It does not seem likely, however, that variations in the length of the flight have extremely important implications so far as the use of equipment is concerned.

Examination of the study by American Airlines. The study by American Airlines needs to be examined in detail before the type of criticism made here can be fully accepted. A close examination is likely to show that some of the arguments used here are not particularly applicable to American's study. The acceptance of American's arguments on *a priori* grounds is not, however, to be justified. Due to cultural bias, it is extremely difficult for any research worker to find the logical error involved in an accounting study of this nature. That which appears to be extremely logical may not be in accordance with fact. A thorough examination of such an accounting study means that the correlation must be made between cost and each element of expense over a large number of years. This will show whether or not the particular elements of expense vary with cost as the accounting theory of American Airlines infers. There are two ways in which this type of correlation should be run. In the first place, the study should be confined to data from American Airlines. This would mean that cost per revenue ton mile would be correlated with the various elements of expense, all the data being drawn from figures of American Airlines. Next, the same type of correlation should be made for averages in the entire

industry. A comparison could then be made between the distribution of American Airlines and the distribution of the entire industry. Such a study would provide a reliable basis for making cost accounting distributions. Once such a study was made, however, there would be no point in attempting to determine the outcome of an issue by accounting methods. If the accounting methods were at maximum efficiency, the best they could do would be to approximate the results obtained by statistical methods.

The foregoing suggestions on research that is needed to verify this study should indicate the limitations of the study. It must be admitted that the outcome of such a research program might reverse all the findings explained in this thesis. On the other hand, there is good reason to believe that the findings of this thesis would be supported by further research. The question of whether or not further research is justified depends upon the importance of the question under consideration. This question can best be determined by examining further the issues of airline economics. Justification of further research depends somewhat upon the implications that arise as a result of the work described in this study. With this objective, the theoretical results of the study will be further explored.

II. THEORY

Size of the firm. Koontz[2] has an excellent discussion on

---

2 op. cit., p. 114

II. INCOME

Size of the firm. Koontz[2] has an excellent discussion on

---

2 *op. cit.*, p. 118

the size of the firm. He has indicated that four of the smallest firms seem to have a handicap due to size. With this exception, he finds that smaller firms seem to be comparatively as efficient as larger firms. He had apparently used exclusively an analysis of cost per available ton mile to arrive at his conclusions. There seems to be considerable possibility that he had made an error by using this measure. American Airlines has argued that the cost of selling tickets and performing many of the other passenger services is important in determining costs connected with the length of the haul. If these costs are important in connection with the length of the haul, they should also be important when a large differential in load factor is involved. If the costs of these services is important, then it is to be expected that they will cause the large firms, with high load factors, to have comparatively higher costs per available ton-mile than the smaller firms. It is to be noted that cost per available ton-mile would contain the larger costs associated with the larger load factors, but no allowance would be taken for the extent that the larger firms are loaded almost to capacity. Correlation, where cost per revenue ton mile is used, does not have this disadvantage. The effects of load factor may be subtracted, so that only the effects of correlation between cost per revenue ton mile and the size of the firm remain.

Computations based on cost per revenue ton mile showed that the size of the firm is insignificant. It may reasonably be concluded from this study that there is no reason for further

the size of the firm. He has indicated that four of the smallest firms seem to have a handicap due to size. With this exception, the three that smaller firms seem to be comparatively as efficient as larger firms. He had apparently used exclusively an analysis of cost per available ton mile to arrive at his conclusions. There seems to be considerable possibility that he had made an error by using this measure. American Airlines has argued that the cost of selling tickets and performing many of the other passenger services is important in determining costs connected with the length of haul. If these costs are important in connection with the length of the haul, they should also be important when a large differential in load factor is involved. If the costs of these services is important, then it is to be expected that they will cause the large firms, with high load factors, to have comparatively higher costs per available ton-mile than the smaller firms. It is to be noted that cost per available ton-mile would consider the larger costs associated with the larger load factors, but no allowance would be taken for the extent that the larger firms are loaded almost to capacity. Correlation, where cost per revenue ton-mile is used, does not have this disadvantage. The effects of load factor may be subtracted, so that only the effects of correlation between cost per revenue ton mile and the size of the firm remain.

Computations based on cost per revenue ton mile showed that the size of the firm is insignificant. It may reasonably be concluded from this study that there is no reason for further

investigations relating to the size of the firm. The statements of Koontz on the size of the firm were verified. It is believed that this finding alone justifies the study and thesis.

It follows from this finding that no particular advantage is to be gained by consolidating the larger firms. The number and size of carriers is a question about which Congress and the Civil Aeronautics Board must make decisions. This study should throw some light on the question. Because of factors other than the size of the firm, the larger carriers have important economic advantage. This study shows without question, however, that this advantage is not due to size.

Other unimportant factors. In addition to size of the firm, this study showed that metropolitan population served, speed of the plane, and length of the flight were unsignificantly correlated with cost. Koontz has indicated that load factor depends upon the community interest between two cities. This means that load factor is not in proportion to population. To at least a limited extent, that is the finding of this study. It might be supposed that it was the removal of the effects of load factor from metropolitan population that caused metropolitan population served to become insignificant. An experiment showed that this was not the case. A regression equation containing the variables capacity of the plane, plane utilization, speed of the plane, metropolitan population, and length of the flight was computed. Notice that ton-mile load factor did not appear in this equation.

Nevertheless, the regression coefficient for metropolitan population served did not appear to be significant.

As might be expected, little is to be gained by increasing the speed of the plane. This increase apparently increases costs in proportion. The theoretical possibilities of the speed of the plane exist primarily in relation to the length of the flight. Both of these variables were discussed in the previous chapter.

Load factor. Load factor is by far the most important element in airline economics. If the small firms were able to operate with load factors of 70% or 80%, they could easily overcome the disadvantages they have when compared with the large firms. Load factor is known to vary with the size of the city. In general, a large city is a prerequisite to a proportionately larger number of people who desire to travel by air. The reasons for this have not been adequately explained. It is known that a large amount of air travel is done for purposes of business. Such business tends to concentrate in the large metropolitan cities. For this reason, business travel tends to occur chiefly between large metropolitan cities. Vacation travel accounts for another important segment of air travel. The per capita income of large cities is known to be greater than the per capita income of smaller cities. For this reason, there is a greater tendency for metropolitan people to take their vacations by air. Finally, the large city affords the advantage of a smaller deviation from day to day in the number of people who desire

to travel. It is easier to predict what a large number of people are going to do from day to day than it is to predict what a smaller number of people are going to do.

The extreme importance of load factor suggests the advantages of reduced rates, family fares, etc., as an element in cost reduction. Assuming that a function for the elasticity of demand is known, the equation developed here could be used to compare the reduction in cost due to increased load factor with the increase, or decrease, in revenue due to a reduction in ticket rate.

Capacity of the plane. Capacity of the plane depends upon the number of passengers available. Management will not find it expedient to use a large plane unless there are a large number of passengers available to fill that plane. For this reason, it is not surprising that the capacity of the plane is strongly associated with load factor. On the other hand, a large plane must in general be used over a longer route than a small plane. This study shows that the capacity of the plane is significant in airline economics, that substantial savings are made as the capacity of the plane is increased. It is known that the larger the plane the longer must be the flight in order to permit efficient operations. This does not answer directly the question of how long the flight has to be, however. It may be that in general a DC-6B is as effective at 300 miles as it is at 1,000 miles. A few of the operating conditions which are associated with an increase in the length of the flight have been presented in the previous chapter. There are probably

to travel. It is easier to predict what a large number of people are going to do from day to day than it is to predict what a smaller number of people are going to do.

The extreme importance of load factor suggests the advantages of reduced rates, family fares, etc., as an element in cost reduction. Assuming that a function for the elasticity of demand is known, the equation developed here could be used to compare the reduction in cost due to increased load factor with the increase, or decrease, in revenue due to a reduction in ticket rate.

Capacity of the plane. Capacity of the plane depends upon the number of passengers available. Management will not find it expedient to use a large plane unless there are a large number of passengers available to fill that plane. For this reason, it is not surprising that the capacity of the plane is strongly associated with load factor. On the other hand, a large plane must in general be used over a longer route than a small plane. This study shows that the capacity of the plane is significant in airline economics, that substantial savings are made as the capacity of the plane is increased. It is known that the larger the plane the longer must be the flight in order to permit efficient operations. This does not answer directly the question of how long the flight has to be, however. It may be that in general a DC-6B is as effective at 300 miles as it is at 1,000 miles. A few of the operating conditions which are associated with an increase in the length of the flight have been presented in the previous chapter. There are probably

others that have not been mentioned. Available studies do not furnish a clear picture of what happens as the length of the flight increases. On an industry-wide basis, the following facts may be surmised from the statistical calculations previously presented: (1) There is a greater correlation between unit cost and capacity of the equipment than there is between unit cost and length of the flight; (2) Capacity of the equipment is strongly correlated with load factor and from an a priori approach depends primarily upon load factor. If it is admitted that capacity of the plane is important, it is impossible to escape the admission that the length of the flight must be of some importance. The degree of association between capacity of the plane and length of the flight becomes the issue under debate.

Why should it be assumed that it is the change in capacity of the plane rather than the change in the length of the flight that is responsible for changes in unit cost? In the first place, this study shows that unit costs follow the capacity of the plane much more closely than they follow the length of the flight. When the effect of changes in the length of the flight is removed from changes in the capacity of equipment, changes in the capacity of equipment continue to be correlated with unit cost. On the other hand, when the effect of changes in the capacity of equipment is removed from changes in the length of the flight, the length of the flight is no longer correlated with unit cost. It is just as logical to say that the length of the flight depends upon the

capacity of the plane as it is to say that the capacity of the plane depends upon the length of the flight. In consideration of the increase in cost that results from operating a plane beyond its capacity, it is perhaps more logical to say that the length of the flight depends upon the capacity of the plane. If this surmise is accepted, the logic of removing statistically the effects of changes in the capacity of the plane from changes in the length of the flight becomes apparent. It seems natural enough that when the effects of capacity of equipment are removed, length of the flight will show no significance.

The foregoing considerations suggest that the capacity of the plane, not the length of the flight, is primarily responsible for changes in cost. This means that once an adequately large load factor is available to permit the use of large equipment, most of the reduction in cost that is possible has already been achieved. Management should be aware and the public should be aware that the primary savings arise from the fact that a large number of people is available to fill to near capacity a large plane.

The theory of plane capacity. A simplified analysis of the manner in which a change in capacity of equipment will effect unit cost should help to show why plane capacity is so important. Suppose Company A has two small planes which it is using to transfer passengers and cargo. A large number of passengers habitually travel over the route of Company A, so that the load

factors are usually high. Company A decides to exchange its two small planes for a large plane. Although the large plane will hold as much in passengers and cargo as the small planes, the company will make a profit on the exchange. In other words, the large plane will carry as much as the two small planes, but its depreciation will be less. Acquisition of the large plane has the immediate effect of reducing by almost 50% the crew requirements. It also reduces the amount of maintenance, since less maintenance is required for one large plane than was previously required for two small planes. Certain other expenses will be slightly reduced. The very day on which Company A begins to operate this larger plane, its unit costs will substantially decrease. Note that no assumption was made relative to a change in load factor or equipment utilization. With what will this decrease in cost be correlated? It will obviously be correlated with a change in capacity of equipment and nothing else.

Assume that Company A operates between stations 1, 2, and 3. Most of the passengers travel from Station 1 to Station 3. These passengers are dissatisfied because the stop at Station 2 requires 45 minutes of their time. Company A believes that it could achieve higher load factors by omitting Station 2 than it could by stopping there. When Company A was using the smaller planes, it was necessary to stop at Station 2 to refuel. Now that a larger plane is in use, there is no advantage to be obtained from this stop. Company A finds that the costs of operating Station

2 is approximately equal to the revenue derived therefrom. This station is therefore eliminated. Although there is a slight increase in load factor, there is no significant change in cost per revenue ton mile due to the elimination of that stop. Company A has now increased the length of its flight, but it has not accomplished a significant change in unit cost. The elimination of the intermediate stop does not immediately effect load factor. For this reason, there is no correlation between the increase in the length of the flight and changes in cost. In a year or more after the intermediate stop is eliminated, the public becomes aware of the increased speed with which they can travel between Station 1 and Station 3. The load factor slowly shows the slight increase that the company anticipated. Due to the lag between the time Station 2 was eliminated and the time the increase in load factor occurred, no reduction in unit cost is associated with the elimination of Station 2. Station 2 was eliminated primarily for the convenience of the through public. It was eliminated because the use of the larger plane made a longer flight possible, not because a substantial reduction in cost would result from the longer flight. The large load factor permitted the use of larger equipment. The use of larger equipment made longer flights convenient. It should be expected from this that a large load factor would be associated with the use of larger equipment and that the use of larger equipment would be associated with longer flights. The larger load factor was itself responsible

for reductions in cost. The use of larger equipment was also responsible for reductions in cost. Increase of the two variables - load factor and capacity of the plane - together made an increase in the length of the flight practicable.

Utilization of planes. Depreciation of flight equipment amounts to approximately 10% of total costs of operation. In light of the relative significance of depreciation, variations in the use of equipment would not appear to be of great importance. The relative importance of equipment utilization may be accounted for in two ways. First, most other factors considered in this study have shown comparatively little tendency to vary with cost. Although the cost of equipment is not particularly significant, its consistent variation with unit cost has correlated sufficiently so as to make it appear more important than it actually is. Second, there is a possibility that equipment utilization is only one indication of overall high operating efficiency. Companies which utilize their planes to a maximum capacity also utilize their ground equipment and personnel to a maximum capacity. Under this theory, the number of hours per day that a plane is used will be considered an indication of managerial efficiency. It seems likely that route characteristics also play an important part in determining whether or not it is possible for management to obtain the maximum utilization of equipment and personnel. Management must weigh the relative effects of securing a high equipment utilization against the effects of operating during hours when a

heavy load factor is difficult to obtain.

The suggestion that peaks in traffic loads make the maximum utilization of equipment significant is highly probable in the light of this study.

Route structure. This study suggests that the two most important factors in the determination of airline self-sufficiency are load factor and capacity of the equipment. This study seems to verify the work of Koontz as previously cited. It indicates that airlines are in a position to become self-sufficient if they can obtain high load factors and use large equipment. It indicates that important economies will probably not be obtained by omitting intermediate stations. Mergers of small companies will effect an element of cost reduction. The reason that some companies are prosperous while others are highly subsidized is simple enough. The high-cost carriers are those which operate over routes that few people care to travel.

Tapering fare. In the light of this study, the tapering fare principle seems subject to considerable question. Distance of travel is presently one of the factors used in determining the price per mile at which passenger tickets will be sold. This has been justified on the basis of economies resulting from the long flight. From the standpoint of this study, the attitude of the Civil Aeronautics Board appears to have been both reasonable and

justifiable. The Civil Aeronautics Board has been criticized for its failure to base freight rates more directly on the distance of the haul.

> The Board's attitude in allowing uneconomic rates to persist was difficult to understand. For example, why should rates have shown little relationship to distance with the result that air freight for short hauls, in most instances, was less expensive than rail express. . . .[3]

---

3 Joseph L. Nicholson, Air Transportation Management (New York: John Wiley & Sons, 1951), p. 190

The above-quoted author represents the viewpoint of the accountant. The Civil Aeronautics Board, on the other hand, uses statistical techniques to determine the significance of the length of the haul. It is realized that the Civil Aeronautics Board may have many reasons for giving small consideration to the length of the haul, the results of statistical investigation being only one of them. It appears from the material presented here that techniques of evaluation are one of the reasons for the differences in viewpoint that exist between certain managerial groups and the Civil Aeronautics Board. But the conflict between the Civil Aeronautics Board and certain groups of airline management probably is much more extensive than the conflict between accounting and statistics.

III. PUBLIC POLICY

Income distribution. The tapering fare represents a

charge that decreases as the distance traveled increases. In a sense, it represents a shifting of the burden of transportation from the group who travel a long distance to a group who travel a short distance. The question may be asked, is this not the same group? Unfortunately, no data seems to be available on this subject. It is known that generally the higher the income the more the individual spends for luxury purposes. As an example, the higher the income the more the individual spends for vacations. The long distances that American Airlines discusses in its accounting study are far beyond the economic possibilities for the average man in the low income brackets. The reduction in fares due to an increase in the distance traveled is apparently a shifting of the burden of transportation from the lower income group to the higher income group. It represents a movement to encourage those in the higher income brackets to travel more. To the extent that the decrease in ticket revenue associated with the longer flight must be compensated for by an increase in revenue associated with the shorter flight, the tapering fare principle represents a movement to make the low-income group travel less by air. From the standpoint of income groups, the opposite end is accomplished by air-coach service. This service represents an attempt to persuade the low income group to travel more by air. The principle of the tapering fare and the principle of air-coach service are thus seen to be antagonistic. American Airlines advocates the tapering fare principle primarily on the ground that it reflects unit cost.

In the light of the study here described, this argument is suspect. It should be noted that in the past the airlines have appealed largely to the higher income group. The operation of the luxury or "de luxe" service has been based primarily on this type of appeal. Once the idea of appealing to this income group is accepted, the idea of making differentiations is a logical extension. The airline business has been very selective from an income standpoint. The principle of tapering fare is a technique for achieving greater selectivity.

The question might well be asked, does a reduction in fare for people who are traveling long distances have a significant effect upon the number who travel? Would not larger increases in load factors be obtained by giving comparable decreases for short distances. Assuming that the higher income groups are the ones who travel long distances, the effect of the reduced cost in inducing more travel might be questioned. For the larger corporations which are in the higher income brackets, it seems rather doubtful that this decrease in cost will produce an increase in the number of business men who travel. Travel expense is subject to deduction for income-tax purposes, and for this reason a small reduction in expense due to the length of the flight may not be viewed with as much enthusiasm as might be expected.

A policy for future development. The finding that load factor and capacity of equipment are the two variables primarily

responsible for unit cost in the airline industry has significant implications so far as a policy of airline development is concerned. In the first place, it shows that costs can be expected to decline as the number of passengers or the amount of freight increases. This suggests that the primary policy of the Civil Aeronautics Board should be to induce more people to travel by air. How can this be accomplished?

Continuing the present policy of government subsidy is probably one of the ways in which the government can do most to promote travel by air. The charge has been made that this is subsidizing one means of travel at the cost of another. Although there is truth to this argument, the fact of the matter is that all means of travel are government subsidized. The growth of the railroads was aided in every way possible by the government. If airlines are to be placed on a competitive status comparable with other modes of transportation, they should first receive a comparable amount of aid. The trucking industry has the highways provided by government. Canals have traditionally been built by government. The fact of the matter is that the government has developed nearly every other form of transportation to the point where it could become relatively self-sufficient. This can happen throughout most of the airline industry, and it will doubtless happen eventually if present policies are continued. The fact that a few of the larger companies have become relatively self-sufficient is ample evidence that more carriers will do so in the future.

The assumption is usually made that most people travel by train or by bus primarily because of the economic advantages. While there is undoubtedly some truth in this statement, the foundamental problem is obscured by it. Sociologists have continually pointed out that much of what people do is done because of custom and habit. This no doubt holds true in the airline industry. It would seem that the problem of the airlines is to make air travel as acceptable among the masses as is travel by bus. This is one reason why subsidization is necessary. It is necessary to bring the cost of travel down to a point where a larger number of people will be induced to travel by air. Once the advantages of air travel are fully accepted, the economies of mass transportation will be obtained. This thesis has shown that such economies are practical and that they depend primarily upon large load factors. These load factors cannot be obtained in a day, but there is every reason to believe that over a long period of time they will be obtained. The argument to reduce present subsidy is primarily an argument that mass transportation by air is not worth the price.

Assuming that large load factors and mass transportation by air is the objective, how can this best be obtained. It would seem that every effort should be made to break down the resistance many people have to air travel. This can best be accomplished by getting them to travel by air a few times. This is where the advantages of reduced fares for the short haul are most important. Generally the short-haul passengers will be those in the lower

The assumption is usually made that most people travel by train or by bus primarily because of the economic advantages. While there is undoubtedly some truth in this statement, the fundamental problem is obscured by it. Sociologists have continually pointed out that much of what people do is done because of custom and habit. This no doubt holds true in the airline industry. It would seem that the problem of the airlines is to make air travel as acceptable among the masses as is travel by bus. This is one reason why subsidization is necessary. It is necessary to bring the cost of travel down to a point where a larger number of people will be induced to travel by air. Once the advantages of air travel are fully accepted, the economies of mass transportation will be obtained. This thesis has shown that such economies are practical and that they depend primarily upon large load factors. These load factors cannot be obtained in a day, but there is every reason to believe that over a long period of time they will be obtained. The argument to reduce present subsidy is primarily an argument that mass transportation by air is not worth the price.

Assuming that large load factors and mass transportation by air is the objective, how can this best be obtained. It would seem that every effort should be made to break down the resistance many people have to air travel. This can best be accomplished by getting them to travel by air a few times. This is where the advantages of reduced fares for the short haul are most important. Generally the short-haul passengers will be those in the lower

income brackets. There are larger numbers of them than there are people in the higher income brackets. Although most people who travel by air presently make more than $5,000 per year, it is nevertheless true that the further the income exceeds $5,000 the smaller will be the number of people in a given income bracket. An appeal to the masses essentially means an appeal to those in the lower income brackets. This generally means an appeal to people who expect to travel shorter distances. The alert reader will at once want to know why it is that the long hauls are presently the low-cost hauls. In the light of this argument, the answer is that the airlines have in the past appealed to high income groups. High income groups reside in metropolitan areas and travel longer distances. By pursuing their present policies, the airlines have tended to exclude the mass market which would ordinarily travel comparatively shorter distances. It is not denied that in light of costs experienced in the past, this has seemed to be the desirable course. It may be argued that the policy is not the best policy for the future, however. The future promises a higher purchasing power for the masses. A survey of the past development will show that this purchasing power has increased fairly uniformly over the years. As the purchasing power of the average man increases, the extent to which he is motivated primarily by economic considerations decreases. This means that the influence of custom and habit on spending will become stronger than it has been in the past. The airlines should be able to capture

larger percentages of the transportation market. Much of this market is short-haul traffic which can become profitable once a sufficiently large number of lower-income people accept air transportation. This is a long-run policy. It is not necessary to become over-optimistic about the future in order to justify this policy. Present economic trends justify it. The growth of the airline industry has been remarkably rapid, and the growth of the low-income purchasing power has been dependable. The function of long-run airline policy is to fuse these trends into a prosperous future.

BIBLIOGRAPHY

I. ECONOMICS AND ADMINISTRATION

Due, John F., Intermediate Economic Analysis. Chicago: Richard D. Irwin, 1950. 566 pp.

Duncan, Julian S., Transport Economics. Albuquerque: University of New Mexico, Department of Economics. (Mimeographed)

Keynes, John Maynard, The General Theory of Employment Interest and Money. New York: Harcourt, Bruce, and Company.

Koontz, Harold D., "Domestic Air Line Self-Sufficiency", The American Economic Review, Volume XLII, Number 1, March, 1952

_____ "Air Line Self-Sufficiency: Rejoinder," The American Economic Review, Volume XLIII, Number 9, June, 1953

Nicholson, Joseph L., Air Transportation Management. New York: John Wiley and Sons, 1951. 446 pp.

II. GOVERNMENTAL PUBLICATIONS AND OTHER SOURCES OF DATA

Air Transport Association of America, data on domestic airlines. Approximately 20 pages of this data are on file in the University of New Mexico, Department of Economics. This data includes both operating and cost data for each airline and for each type of aircraft. The association charges for this data.

Cherington, Charles R., "The Essential Role of Large Irregular Airlines in United States Air Transportation Industry", Hearings Before A Subcommittee of the Select Committee on Small Business, United States Senate, March 31, May 1, 4, 5, 6, 7, and 8, 1953.

Civil Aeronautics Board, Transcontinental Coach Type Service Case, Brief of American Airlines, Inc., Docket Nos. 3397 et al., Washington, D. C., Civil Aeronautics Board, April 20, 1950.

Civil Aeronautics Board, Transcontinental Coach Type Service Case, Report of William J. Madden, Examiner, Docket Nos. 3397 et al., Washington, D. C., Civil Aeronautics Board, Served November 22, 1950.

Civil Aeronautics Board, General Passenger Fare Investigation, Exhibits of American Airlines, Inc., Docket No. 5509, Washington, D. C., Civil Aeronautics Board, April 1, 1953.

Civil Aeronautics Board, Uniform System of Accounts for Air Carriers, Form 41 Manual, Washington, D. C., Civil Aeronautics Board, 1951, effective January 1, 1947.

Civil Aeronautics Board, Recurrent Report of Financial Data, Washington. At the time of writing a free copy of this report may be obtained by writing to the Civil Aeronautics Board. This report constitutes a basic source of data for the study of financial data.

Civil Aeronautics Board, Cost Standards - Domestic Scheduled Air Carriers, Washington, D. C., Civil Aeronautics Board, July, 1950. 83 pp.

Roadcap, Roy R., and associates, World Airline Record. Chicago: published by Roy R. Roadcap, compiler, 1952. 396 pp.

III. MATHEMATICS AND STATISTICS

Arkin, Herbert, and Raymond R. Colton, An Outline of Statistical Methods. New York: Barnes and Noble, 1950. 224 pp.

Cramer, Harold, Mathematical Methods of Statistics. Princeton: Princeton University Press, 1951. 575 pp.

Ezekiel, Mordecai, Methods of Correlation Analysis. New York: John Wiley & Sons, 1930. 427 pp.

Feller, William, Probability Theory and its Applications, Volume One. New York: John Wiley & Sons, 1950. 419 pp.

Ferrar, W. L., Algebra. Oxford: Clarendon Press, 1949. 202 pp.

Frazer, R. A., Elementary Matrices. Cambridge University Press, 1938. 524 pp.

Hoel, Paul G., Introduction to Mathematical Statistics. New York: John Wiley & Sons, 1947. 258 pp.

Klein, Lawrence R., Econometrics. Evanston, Illinois: Row, Peterson and Company, 1953. 355 pp.

Kenney, E. F. and E. S. Keeping, Mathematics of Statistics. New York: D. Van Nostrand Company, 1951. 429 pp.

McNemar, Quinn, Psychological Statistics. John Wiley & Sons, 1950. 388 pp.

Snedecor, George W., Statistical Methods. Ames, Iowa: Iowa State College Press, 1940. 422 pp.

Tintner, Gerhard, Econometrics. New York: John Wiley & Sons, 1952. 370 pp.

Wold, Herman, Demand Analysis. New York: John Wiley & Sons, 1953. 358 pp.

IV. PHILOSOPHY

Bewkes, Eugene G., and others, A Survey in Philosophy and Religion. New York: Harper and Brothers, 1940. 649 pp.

Dewey, John, Reconstruction in Philosophy. New York: The New American Library, 1950. 168 pp.

_____ Theory of Valuation. Chicago: The University of Chicago Press (International Encyclopedia of Unified Science, Volume II, Number 4), 1947. 63 pp.

MATHEMATICAL

APPENDIX

MATHEMATICAL APPENDIX

I. VOCABULARY

Covariance: The covariance of x_1 and x_2 is defined by:

$$\text{cov}(x_1,x_2) = SX_1X_2/n = S(x_1-M_1)(x_2-M_2)/n - M_1M_2$$

The above definition is meaningful whenever x_1 and x_2 have finite variances.[1] Note that the difference between

---

1 William Feller, Probability Theory and its Applications (New York: John Wiley & Sons, 1950), p. 180.

$\text{cov}(x_1,x_2)$ and Z_{12} is that $\text{cov}(x_1,x_2) = Z_{12}/n$.

Deviance means the sums of the squares.[2] The term

---

2 F. F. Kenney and E. S. Keeping, Mathematics of Statistics (New York: D. Van Nostrand Company, 1951), p. 239.

"residual deviance" will be used to denote the sums of the squares of the deviances of the observed x's from the x's predicted by means of a least-squares equation. Note that the difference between deviance and variance is that variance is deviance divided by n.

As an example of the use of deviance, consider the

following case. The deviance of x_1 will be written

$$D = SX_1^2$$

The variance of x_1 will be written

$$q^2 = SX_1^2/n$$

or

$$nq^2 = D.$$

A Normalized variable. If X_1 has the mean M_1 and the variance q, then x_1-M_1 has a zero mean and hence the variable

$$X_1 = (x_1-M_1)/q$$

has zero mean and variance 1. It is called the normalized variable corresponding to X_1. The physicist would interpret this as the introduction of dimensionless quantities.[3]

---

3 Feller, op. cit., p. 179

In normalized covariance, the deviations of both variables from their mean is divided by their respective standard deviations.

following case. The deviance of x^1 will be written

$$D = \Sigma x^1{}_2$$

The variance of x^1 will be written

$$\sigma_2 = \Sigma x^1{}_2/n$$

or

$$n\sigma_2 = D.$$

A Normalized variable. If X^1 has the mean M^1 and the variance σ, then x^1-M^1 has a zero mean and hence the variable

$$X^1 = (x^1-M^1)/\sigma$$

has zero mean and variance 1. It is called the normalized variable corresponding to X^1. The physicist would interpret this as the introduction of dimensionless quantities.[3]

---

3 Feller, op. cit., p. 179

In normalized covariance, the deviations of both variables from their mean is divided by their respective standard deviations.

II. SYMBOLS

Symbol	Meaning
A	A determinant used for theoretical purposes in connection with least-squares equations. Where i and j denote the row and column, each element of this determinant may be designated Z_{ij}, i, j = 1,2, . . . n. See the definition of Z. The determinant, A, is defined in equation 3.
A_{ij}	The cofactor of the element in the i'th row and j'th column of A.
A*	A determinant consisting of the cofactors of each respective element of A.
B	The vector of the coefficients of the least-squares equation where b_j is the coefficient of the j'th column, j = 2, 3 . . . n. See equation 1.
C	$\begin{bmatrix} Z_{22} & Z_{23} & Z_{24} & \cdots & Z_{2m} \\ Z_{32} & Z_{33} & Z_{34} & \cdots & Z_{3m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ Z_{m2} & Z_{m3} & Z_{m4} & \cdots & Z_{mm} \end{bmatrix}$

Note that the determinant of this matrix may be expressed in terms of A. This determinant will

II. SYMBOLS

Symbol	Meaning
A	A determinant used for theoretical purposes in connection with least-squares equations. Where i and j denote the row and column, each element of this determinant may be designated Z_{ij}, i, j = 1,2, . . . n. See the definition of Z. The determinant, A, is defined in equation 3.
A_{ij}	The cofactor of the element in the i'th row and j'th column of A.
A*	A determinant consisting of the cofactors of each respective element of A.
B	The vector of the coefficients of the least-squares equation where b_j is the coefficient of the j'th column, j = 2, 3 . . . n. See equation 1.
C	$\begin{bmatrix} Z_{22} & Z_{23} & Z_{24} & \cdots & Z_{2n} \\ Z_{32} & Z_{33} & Z_{34} & \cdots & Z_{3n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ Z_{n2} & Z_{n3} & Z_{n4} & \cdots & Z_{nn} \end{bmatrix}$

Note that the determinant of this matrix may be expressed in terms of A. This determinant will

Symbol	Meaning

be given subsequently in expression (4).

C^{-1} a matrix inverse to C.

c^{ij} An element of the inverse matrix C^{-1}, where i, j = 1, 2, 3, . . . m. In the particular case where i = j, elements of the diagonal are given by c^{jj}.

D SX_1^2. The sum of the squares of the deviations of the dependent variable, x_1, from its mean. See also Z_{ij}/n.

F The ratio used in Snedecor's F test.[4] As defined

---

4 Kenney and Keeping, op. cit., p. 182

---

by McNemar:[5]

---

5 Quinn McNemar, Psychological Statistics (New York: John Wiley & Sons, 1949), p. 266

$$F = \frac{(R_1-R_2)/(m_1-m_2)}{(1-R_1)/(n-m_1)}$$

where R_1 is defined as in R in these definitions. R_1 is based on m_1 variables and R_2 is based on m_2 variables selected from among the m_1 variables. The number of degrees of freedom is $N_1 = m_1-m_2$ and $N_2 = n-m_1$.

Symbol	Meaning
	be given subsequently in expression (b).
C^{-1}	a matrix inverse to C.
c_{ij}	An element of the inverse matrix C^{-1}, where i, j = 1, 2, 3, . . . m. In the particular case where i = j, elements of the diagonal are given by c_{ii}.
D	SX_j^2. The sum of the squares of the deviations of the dependent variable, x_j, from its mean. See also $\Sigma x_j^2/n$.
F	The ratio used in Snedecor's F test.[b] As defined

---

[b] Kenney and Keeping, op. cit., p. 252

---

by McNemar:[c]

---

[c] Quinn McNemar, Psychological Statistics (New York: John Wiley & Sons, 1949), p. 328

$$F = \frac{(R_j^2 - R_s^2)/(m_j - m_s^2)}{(1-R_j^2)/(n-m_j)}$$

where R_j^2 is defined as in R in these definitions. R_j^2 is based on m_j variables and R_s^2 is based on m_s^2 variables selected from among the m_j variables. The number of degrees of freedom is $N_j^2 = m_j - m_s^2$ and $N_s^2 = n - m_j$.

Symbol	Meaning
G	A column matrix consisting of the elements Z_{12}, Z_{13}, . . . Z_{1m}. Note that when the vector, G is multiplied by the vector B, the result is a scaler indicated by definition H. Where B is known, this procedure is used to calculate H.
H	$b_2Z_{12} + b_3Z_{13} + \ldots + b_mZ_{1m}$. $(H = SX_1W)$ $H = n\ \mathrm{cov}(x_1,w)$, where w is the predicted x_1.
h_j	$b_2Z_{12} + b_3Z_{13} + \ldots + b_{j-1}Z_{1,j-1} + b_{j+1}Z_{1,j+1} + \ldots + b_mZ_{1m}$ The symbol h_j is similar to H except that in h_j the term b_jZ_{1j} has been omitted. This amounts to omitting the effects of one variable from the determinant A.
i	The row to which each element of the determinant A belongs.
j	The column to which each element of the determinant A belongs. This is also the column to which each coefficient of the vector B belongs.
K	$1 - (1-R)(n-1)/(n-m)$, R adjusted.
k	$1 - (1-r)(n-1)/(n+1-m)$, r adjusted. (The square of the multiple correlation coefficients must be adjusted for shrinkage[6]. The above definitions

Symbol	Meaning
G	A column matrix consisting of the elements $r_{12}, r_{13}, \ldots r_{1m}$. Note that when the vector, G is multiplied by the vector B, the result is a scalar indicated by definition H. Where B is known, this procedure is used to calculate R.
H	$b_2 r_{12} + b_3 r_{13} + \ldots + b_m r_{1m}$. $(H = BX_1R)$ $H = n\ \mathrm{cov}(x_1, w)$, where w is the predicted x_1.
H_j	$b_2 r_{12} + b_3 r_{13} + \ldots + b_{j-1} r_{1,j-1} + b_{j+1} r_{1,j+1} + \ldots + b_m r_{1m}$ The symbol b_j is similar to H except that in H_j the term $b_j r_{1j}$ has been omitted. This amounts to omitting the effects of one variable from the determinant A.
i	The row to which each element of the determinant A belongs.
j	The column to which each element of the determinant A belongs. This is also the column to which each coefficient of the vector B belongs.
K	$1 - (1-R)(n-1)/(n-m)$, R adjusted.
k	$1 - (1-r)(n-1)/(n-1-m)$, r adjusted. (The square of the multiple correlation coefficients must be adjusted for shrinkage[6]. The above definitions

Symbol | Meaning

6 Mordecai Ezekiel, Methods of Correlation Analysis (New York: John Wiley & Sons, 1930), p. 177

make the required adjustment.[7]

7 See also McNemar, op. cit., p. 161, for a discussion of the reason why this adjustment must be made.

m — The number of variables, both dependent and independent, in a least-squares equation. There are m-1 independent variables. In a matrix where one variable has been deleted there are m-2 independent variables.

M_j — The mean of the observed variable, x_j. $M_j = Sx_j/n$.

n — The number of population or samples taken of each variable.

N — The number of degrees of freedom.

p — A number of variables selected from among m variables.

P_j — The square of the coefficient of partial correlation between x_1 and x_j where the remaining variables are held constant.

q — The standard deviation of the dependent variable x_1.

Symbol	Meaning

Symbol	Meaning
	See also the term "deviance" in the section on vocabulary.
R	The square of the coefficient of multiple correlations between x_1 and $x_2, x_3, \ldots x_m$. The coefficient of multiple correlation in an unsquared form will seldom be used.
r_j	The square of the coefficient of multiple correlation between x_1 and $x_2, x_3, \ldots x_{j-1}, x_{j+1}, \ldots x_m$. Note that r_j is used where one variable has been omitted from A. If no variable has been omitted, R is used. By definition, $r_j = h_j/D$.
S	The sum over n observations.
s	The sum over m-1 or m-2 independent variables, as the case may be.
t_j^2	The ratio $(n-m)b_j^2U_j/V_j$. The statistic, t_j, has a Student's t distribution with n-m degrees of freedom.[8]

8 Harald Cramér, Mathematical Methods of Statistics (Princeton: Princeton University Press, 1951), p. 410

The following symbols will be used to indicate measures of residual deviance under the circumstances indicated.

The following symbols will be used to indicate measures of residual deviance under the circumstances indicated.

Symbol	Meaning

U_j The residual deviance of x_j with respect to x_2, x_3 . . . x_{j-1}, x_{j+1} . . . x_m . This means that x_1 is omitted, and that the least-squares equation between x_j and the remaining variables is taken. The sum of the squares of the deviations of the observed x_j's from the estimated x_j's is denoted by U_j . The omission of x_1 will result in the deletion of the first row and column from A. This situation may be represented by the cofactor A_{11}.[9] Cramer[10] shows

---

9 Ibid., p. 109
10 Ibid., p. 305

---

that the residual deviance of x_1 with the remaining variables is A/A_{11}. This may be generalized to A/A_{ii}. Application of this rule to the case in question, where the determinant is A_{11} will give: $U_j = A_{11}/A_{11 \cdot jj}$. Here $A_{11 \cdot ij}$ is a cofactor of the $i = j$'th element selected from the cofactor A_{11} .

V The residual deviance of x_1 with respect to the independent variables x_2, x_3 . . . x_m . This may be represented: $V = (x_1 - w)^2 = A/A_{11}$. The difference between U and V is that in U, x_1 is omitted while it is included in V.

Symbol	Meaning
U_j	The residual deviance of x_j with respect to $x_2, x_3 \ldots x_{j-1}, x_{j+1} \ldots x_m$. This means that x_j is omitted, and that the least-squares equation between x_j and the remaining variables is taken. The sum of the squares of the deviations of the observed x_j's from the estimated x_j's is denoted by U_j. The omission of x_j will result in the deletion of the first row and column from A. This situation may be represented by the cofactor A_{jj}.[9] Cramer[10] shows
V	The residual deviance of x_j with respect to the independent variables $x_2, x_3 \ldots x_m$. This may be represented: $V = (x_j - u)^2 = A/A_{jj}$. The difference between U and V is that in U, x_j is omitted while it is included in V.

---

9 Ibid., p. 109
10 Ibid., p. 305

---

that the residual deviance of x_j with the remaining variables is A/A_{jj}. This may be generalized to A/A_{ii}. Application of this rule to the case in question, where the determinant is A_{jj} will give $U_j = A_{jj}/A_{jj \cdot ij}$. Here $A_{jj \cdot ij}$ is a cofactor of the $i = ij$th element selected from the cofactor A_{jj}.

Symbol	Meaning
v_j	The residual deviance of x_1 with respect to the independent variables $x_2, x_3, \ldots x_{j-1}, x_{j+1}, \ldots x_m$. The difference between V and v_j is that in v_j, the j'th variable has been omitted from the system of equations. The omission of this variable results in a determinant which is equal to the cofactor A_{jj}, i.e., a determinant in which the j'th row and column have been omitted. The residual deviance of this determinant is,[11] $v_j = A_{jj}/A_{jj\cdot 11}$.
w	The value of the dependent variable, x_1, estimated, by the least-squares equation, from $x_2, x_3 \ldots x_m$.
W	The deviation of w from the mean of the observed x_1's.
x_j	The observed value of the j'th variable
X_j	The deviation of the observed x_j from its mean. $X_j = x_j - M_j$
X_j^2	The square of the deviation of x_j from its mean.
Z_{ij}	The sum over n samples of the deviation of one variable from its mean multiplied by the deviation of another variable from its mean. $Z_{ij} = S(X_1 - M_{Xi})(X_j - M_{Xj})$.

---

11 Ibid., p. 306

Symbol	Meaning
v_j	The residual deviance of x_j with respect to the independent variables $x_2, x_3, \ldots x_{j-1}, x_{j+1}, \ldots x_m$. The difference between V and v_j is that in v_j the j'th variable has been omitted from the system of equations. The omission of this variable results in a determinant which is equal to the cofactor A_{jj}, i.e., a determinant in which the j'th row and column have been omitted. The residual deviance of this determinant is.[11] $v_j = A_{jj}/A_{jj \cdot 12}$
w	The value of the dependent variable, x_1, estimated, by the least-squares equation, from $x_2, x_3 \ldots x_m$.
W	The deviation of w from the mean of the observed x_1's.
x_j	The observed value of the j'th variable
X_j	The deviation of the observed x_j from the mean. $X_j = x_j - \bar{X}_j$
X_j^2	The square of the deviation of x_j from the mean.
Σ_{ij}	The sum over n samples of the deviation of one variable from the mean multiplied by the deviation of another variable from the mean. $\Sigma_{ij} = \Sigma(X_j - \bar{X}_j)(X_i - \bar{X}_i)$.

---

11 Ibid., p. 306

Symbol	Meaning
Z_{ij}/n	$= \text{cov}(x_i x_j)$. When $i = j$, $Z_{jj} = SX_jX_j = SX_j^2$. In the particular case where $i = j = 1$, $Z_{11} = SX_1X_1 = SX_1^2 = D$. The definition of Z used here is similar to the definition of S used by Tintner.[12]

12 Gerhard Tintner, Econometrics (New York; John Wiley & Sons, 1952), p. 84

$\underline{/}$ The inequality sign used to show that the first quantity is less than the second.

III. FUNDAMENTAL CONCEPTS

The least squares technique is based on a choice of constants b_j so that

$$(X_1 - sb_jx_j)^2 = \text{minimum.}$$

Differentiating this function with respect to $b_2, b_3, \ldots b_m$ gives rise to the following system of equations.

$$(1) \qquad \begin{aligned} b_2Z_{22} + b_3Z_{23} + \ldots + b_mZ_{2m} &= Z_{12} \\ b_2Z_{32} + b_3Z_{33} + \ldots + b_mZ_{mm} &= Z_{13} \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots & \\ b_2Z_{m2} + b_3Z_{m3} + \ldots + b_mZ_{mm} &= Z_{1m} \end{aligned}$$

With the definitions previously set forth, (1) may be

Symbol	Meaning
Σ_{ij}/n	$= \mathrm{cov}(x^i x^j)$. When $i = j$, $\Sigma_{jj} = S x^j x^j = S x^j{}_2$. In the particular case where $i = j = 1$, $\Sigma_{11} = S x^1 x^1 = S x^1{}_2 = 0$. The definition of Σ used here is similar to the definition of S used by Tintner.[12]

---

[12] Gerhard Tintner, Econometrics (New York; John Wiley & Sons, 1952), p. 81.

$\leq$ The inequality sign used to show that the first quantity is less than the second.

III. FUNDAMENTAL CONCEPTS

The least squares technique is based on a choice of constants b^j so that

$$(X^1 - b^j x^j)_2 = \text{minimum}.$$

Differentiating this function with respect to $b^2, b^3, \ldots b^m$ gives rise to the following system of equations.

$$\begin{aligned} b^2\Sigma_{22} + b^3\Sigma_{23} + \ldots + b^m\Sigma_{2m} &= \Sigma_{12} \\ b^2\Sigma_{32} + b^3\Sigma_{33} + \ldots + b^m\Sigma_{mm} &= \Sigma_{13} \\ \ldots\ldots\ldots\ldots\ldots\ldots & \\ b^2\Sigma_{m2} + b^3\Sigma_{m3} + \ldots + b^m\Sigma_{mm} &= \Sigma_{1m} \end{aligned} \tag{1}$$

With the definitions previously set forth, (1) may be

represented in matrix form as follows:[13]

---

13 Ibid., p. 335

$$(2) \qquad CB = G.$$

Note that the system of equations in (1) and (2) is based on moments, not simple correlation coefficients. It can be shown that a system of equations based on correlation coefficients can be made to yield the same coefficients as the b's in (1).

The Doolittle method of computation is based on (1) and (2). For computational purposes, this is the form that will be used. For theoretical purposes the following determinant will be used:

$$(3) \quad A = \begin{vmatrix} Z_{11} & Z_{12} & Z_{13} & \cdots & Z_{1m} \\ Z_{21} & Z_{22} & Z_{23} & \cdots & Z_{2m} \\ Z_{31} & Z_{32} & Z_{33} & \cdots & Z_{3m} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ Z_{m1} & Z_{m2} & Z_{m3} & \cdots & Z_{mm} \end{vmatrix}$$

The cofactor of Z_{11} is A_{11}. The cofactor A_{11} is the determinant of the matrix C.

$$(4) \quad A_{11} = \begin{vmatrix} Z_{22} & Z_{23} & \cdots & Z_{2m} \\ Z_{32} & Z_{33} & \cdots & Z_{3m} \\ \cdots & \cdots & \cdots & \cdots \\ Z_{m2} & Z_{m3} & \cdots & Z_{mm} \end{vmatrix}$$

---

13 Ibid., p. 335

The least squares coefficient, b_j, where x_1 is the dependent variable and x_2, x_3, . . . x_m are independent variables may be represented as follows:[14]

$$b_j = - A_{1j}/A_{11} \tag{5}$$

---

14 Cramér, op. cit., p. 303

In a similar manner, the square of the coefficient of partial correlation between x_1 and x_j with the remaining variables held constant may be represented:[15]

$$P_j = (A_{1j})^2 / A_{11}A_{jj} \tag{6}$$

---

15 Ibid., p. 306

The residual deviance of x_1 with respect to x_2, x_3, . . . x_m is written:[16]

$$V = A/A_{11} \tag{7}$$

---

16 Ibid., p. 305

The square of the coefficient of multiple correlation, R,

is written:[17]

(8) $$R = 1-V = 1-A/DA_{11}$$

---

17 Ibid., p. 308

IV. MULTIPLE CORRELATION

It will now be shown that the square of the coefficient of multiple regression may be developed with the concept of covariance.

The covariance between x_1 and the estimated x_1 is given by

(9) $$cov(x_1,w) = SX_1W/n = H/n.$$

When the fit of x_1 on x_2, x_3 . . . x_m is perfect, each W is exactly equal to the corresponding X_1. Consequently SW^2 approaches D, as the fit becomes better and, the two are exactly equal under conditions of perfect fit. It follows from this that, as the least-squares equation approaches a perfect fit, the standard deviation of w approaches the standard deviation of x_1. Under conditions of perfect fit, the standard deviation of <u>both</u> w and x_1 may be designated by q.

In the case of perfect fit, the normalized covariance between x_1 and w is:

$$q^{-2}\text{cov}(x_1,w) = H/nq^2$$

$$= q^{-2}\text{cov}(x_1,x_1) = D/nq^2 = nq^2/nq^2 = 1$$

This shows that in a special case where a perfect fit is obtained, the normalized covariance between x_1 and w is 1.

Consider next the case where the fit of x_1 on x_2, x_3 . . . x_m is absolutely imperfect. This occurs where x_1 and w are completely independent. In this case $\text{cov}(x_1,w)$ is zero.[18]

---

18 Feller, op. cit., p. 180

It is apparent from the above that the maximum covariance between w and x_1 occurs where

(10) $$\text{cov}(w,x_1) = \text{cov}(x_1,x_1) = D/n.$$

The reason for this is that the normalized covariance approaches 1 when the fit is perfect and approaches zero when the fit is imperfect. All other cases must lie somewhere between a perfect and an absolutely imperfect fit so that the normalized covariance lies between zero and 1. This shows that the maximum $\text{cov}(x_1,w)$ occurs when $x_1 = w$.

It will be observed that the above discussion avoided the question of what happens to the standard deviation of w when the covariance between w and x_1 is zero or near zero. There is, as a matter of fact, no reason for believing that this

statistic will be small, or that it will be significantly different from the standard deviation of w where w is identical with x. The factor SX_1W determines the magnitude of covariance. The magnitude of this factor depends upon whether or not X_1 and W are large or small at the same time, not upon the magnitude of spread possessed by either statistic.

A most meaningful ratio is the ratio of $cov(x_1, w)$ to the maximum covariance of X_1. A most meaningful ratio is therefore the ratio of H/n to D/n or

$$R = SX_1W/D = H/D \tag{11}$$

which shows that R is equal to the ratio of $cov(x_1, w)$ to the maximum covariance.

Since the maximum covariance is D/n, it is obvious that

$$H/n \leq D/n \tag{12}$$

or

$$H \leq D \tag{13}$$

and that $R \leq 1$.

Another way of interpreting R is to multiply R by 100. If H = D, R = 1 and 100R = 100. It can be said that 100% of the variance of x_1 is explained by covariance between x_1 and w.

elasticity will be small, or that it will be significantly different from the standard deviation of w where w is identical with x. The factor SX_jW determines the magnitude of covariance. The magnitude of this factor depends upon whether or not X_j and W are large or small at the same time, not upon the magnitude of spread possessed by either elasticity.

A most meaningful ratio is the ratio of $cov(x_j, w)$ to the maximum covariance of X_j. A most meaningful ratio is therefore the ratio of H/n to D/n or

$$R = SX_jW/D = H/D \tag{11}$$

which shows that R is equal to the ratio of $cov(x_j, w)$ to the maximum covariance.

Since the maximum covariance is D/n, it is obvious that

$$H/n \leq D/n \tag{12}$$

or

$$H \leq D \tag{13}$$

and that $R \leq 1$.

Another way of interpreting R is to multiply R by 100. If $R = D$, $R = 1$ and $100R = 100$. It can be said that 100R of the variance of x_j is explained by covariance between x_j and w.

If 100R = 80, this means that the covariance between x_1 and w is 80% the variance of x_1.

The above percentages may be interpreted as the amount of the variance of x_1 that is explained by the covariance between x_1 and w. If x_1 and w are independent in a statistical sense, their covariance will be zero. If w can be used to predict or explain x_2, they must have significant covariances. As the extent to which w can explain x_1 increases, the covariance of the two will increase. The ratio, R, has as its denominator the maximum covariance between x_1 and w. The ratio R thus shows the ratio of a given covariance to the maximum covariance. But covariance is itself a measure of the extent to which one variable explains another. Hence R must show the percentage of the total variance of x_1 that is explained by its covariance with w.

The ratio, R, is the square of the coefficient of multiple correlation. If the square root of R is taken, however, R no longer may be used to show the percentage of explained variance. For that reason, R will be used in this discussion instead of its square root.

V. RESIDUAL DEVIANCE

It was shown that R is the proportion of variance explained by w. The remaining proportion, 1-R, must be the part of variance that is unexplained. This may be written

If 100R = 80, this means that the covariance between x_j and w is 80% the variance of x_j.

The above percentages may be interpreted as the amount of the variance of x_j that is explained by the covariance between x_j and w. If x_j and w are independent in a statistical sense, their covariance will be zero. If w can be used to predict or explain x_j, they must have significant covariances. As the extent to which w can explain x_j increases, the covariance of the two will increase. The ratio, R, has as the denominator the maximum covariance between x_j and w. The ratio R thus shows the ratio of a given covariance to the maximum covariance. But covariance is itself a measure of the extent to which one variable explains another. Hence R must show the percentage of the total variance of x_j that is explained by the covariance with w.

The ratio, R, is the square of the coefficient of multiple correlation. If the square root of R is taken, however, R no longer may be used to show the percentage of explained variance. For that reason, R will be used in this discussion instead of the square root.

V. RESIDUAL DEVIANCE

It was shown that R is the proportion of variance explained by w. The remaining proportion, 1-R, must be the part of variance that is unexplained. This may be written

(14) $$1-R = 1-H/D = (D-H)/D = V/D$$

The numerator, V, will be called the "residual deviance". It tells the amount of deviance which is unexplained by covariance.

The unexplained part of X is the difference between X_1 and w. The residual deviance is the sum of the squares of this difference. In symbols

(15) $$V = S(X_1-W)^2$$

It can be shown[19] that (15) reduces to the form

(16) $$V = SX_1^2 - b_2X_1X_2 - b_3X_1X_3 - \ldots - b_mX_1X_m$$
$$= D-H.$$

---

19 Cramer, op. cit., p. 305

The above equation is consistent with (14). This verifies the previous statement that V is the deviance unexplained by w. It is possible to reason backward from this and show that R is the proportion of deviance explained by w. Thus (13) gives assurance that the interpretation previously made of (11) is correct.

The residual deviance V can also be expressed in terms of determinants.[20]

(17) $$V = A/A_{11}$$

20 Loc. cit.

The above identity of course applies to the case where x_1 is taken as the dependent variable and V is the sum of the squares of the deviations of x_1 from an x_1 estimated from x_2, x_3, . . . x_m. If any other variable had been chosen as the independent variable instead of x_1, the residual deviance would have been determined in a similar manner. In that case, V_j could have been defined:

(18) $$V_j = A/A_{jj}$$

VI. OMISSION OF A VARIABLE

Suppose that any independent variable, x_j, were omitted from the system of equations. An important question is, what effect would the omission have on R and the residual. If the solution to a least-squares equation is correct, such an omission cannot reduce the residual deviance, V. In most cases, omission of a variable will increase V. The question may also be asked in another way. How much did the inclusion of x_j affect the residual deviance? Two measures of the effect of omitting a variable will be considered here -- the coefficient of partial correlation and the F ratio.

Suppose that x_2 is omitted from the least-squares equation and x_1 is fitted with the remaining variables. Let

20 Loc. cit.

w* be the estimated x_1 based on x_3, x_4, . . . x_m. The change in x_1 that is not due to w* will be (x_1-w^*).

In a similar manner, let x_2 be the dependent variable with x_3, x_4, . . . x_m the independent variable. The change in x_2 that is not due to x_3, x_4, x_m will be (x_2-w^{**}).

Now that the effects of the other variables has been removed, x_1-w^* will be correlated with x_2-w^{**}. The correlation between x_1-w^* and x_2-w^{**} is called the coefficient of partial correlation.[21]

---

21 Herbert Arkin and Raymond R. Colton, An Outline of Statistical Methods (New York: Barnes and Noble, 1950) p. 96.

Like the coefficient of multiple correlation, the square of the coefficient of partial correlation is the statistic easiest to interpret. For this reason, a symbol for the square of the coefficient of partial correlation will be used throughout this discussion.

VII. PARTIAL CORRELATION

When the number of variables equals the number of observations, perfect correlation will result. That is to say, when $m = n$, the coefficient of multiple correlation will always be 1. When the number of variables approaches the number of observations, the coefficient of multiple correlation will approach 1. For this reason, when the number of variables approaches the number of observations, a positive bias results.

The coefficient of multiple correlation must be adjusted to eliminate the effects of this bias.[22]

---

22 McNemar, op. cit., p. 161

The symbol R was previously defined as the square of the multiple correlation coefficient of x_1 with respect to x_2, x_3, . . . x_m. Let r_j be the same coefficient where the effects of x_j have been eliminated from the least-squares equation. In other words, x_j has been omitted from equation (1). Ezekiel[23]

---

23 Ezekiel, op. cit., pp. 177-180

---

gives the following formula for adjustment to eliminate bias where K is R after adjustment and k_j is r_j after adjustment. The appropriate definitions of R follow.

(19) $$R = H/D$$

(20) $$r_j = h_j/D$$

(21) $$D - H = V$$

(22) $$D - h_j = v_j$$

(23) $$K = 1 - (1-R)(m-1)/(n-m)$$

(24) $$k_j = 1 - (1-r_j)(n-1)/(n-1-m)$$

(25) $$1 - K = (1-R)(n-1)/(n-m)$$

The coefficient of multiple correlation must be adjusted to eliminate the effects of this bias.[22]

---

22 McNemar, op. cit., p. 169

The symbol R was previously defined as the square of the multiple correlation coefficient of x_1 with respect to x_2, x_3, . . . x_m. Let r_1^2 be the same coefficient where the effects of x_j have been eliminated from the least-square equation. In other words, x_j has been omitted from equation (1). Wallis[23]

---

23 Wallis, op. cit., pp. 177-180

---

gives the following formula for adjustment to eliminate bias where K is R after adjustment and k_j is r_j after adjustment. The appropriate definitions of R follow.

(19) $R = 1 - \lambda$

(20) $r_j^2 = 1 - \lambda_j$

(21) $D - R = V$

(22) $D - h_j = v_j$

(23) $K = 1 - (1-R)(n-1)/(n-m)$

(24) $k_j = 1 - (1-r_j^2)(n-1)/(n-1-m)$

(25) $1 - K = (1-R)(n-1)/(n-m)$

(26) $$1 - k_j = (1-r_j)(n-1)/(n\not{-}1-m)$$

With the foregoing definitions, Ezekiel's formula for the square of the partial correlation coefficient may be represented as follows.

(27) $$P_j = 1 - \frac{(1-K)}{(1-k_j)} = 1 - \frac{(1-R)(n-1)/(n-m)}{(1-r)(n-1)/(n\not{-}1-m)}$$

Substituting (17) and (18):

(28) $$P_j = 1 - \frac{(1 - H/D)(n\not{-}1-m)}{(1 - h_j/D)(n-m)} = 1 - \frac{(D-H)(n\not{-}1-m)}{(D-h_j)(n-m)}$$

$$= 1 - \frac{V(n\not{-}1-m)}{v_j(n-m)}$$

(29) $$P_j = \frac{v_j(n-m) - V(n\not{-}1-m)}{v_j(n-m)}$$

Where n is fairly large, and m is small, the difference between n-m and n$\not{-}$1-m is insignificant and (29) becomes:

(30) $$P_j = \frac{v_j - V}{v_j} = 1 - V/v_j$$

Substituting (21) and (22) in the numerator of the second expression:

(31) $$P_j = (H-h_j)/v_j$$

The numerator of P_j is the increase in covariance due to inclusion of x_j in a least-squares equation. The square

of the coefficient of partial correlation is the ratio of the increase in covariance to the residual variance before introduction of the variable x_j.

VIII. THE Q RATIO

It will often be desirable to compare the effect of omitting one variable with the effect of omitting another variable. The coefficient of partial correlation is a rough measure of this effect. The weakness of the coefficient of partial correlation is that the denominator changes from variable to variable. The numerator of P_j gives an accurate measure of the increase in covariance that results from the use of the variable x_j. The denominator v_j is the residual deviance where x_j is not included in a least-squares equation. For this reason, v_j does not depend upon x_j in any manner. The magnitude of v_j depends on the variables other than x_j in the least-squares equation.

The above criticism suggests the advantage of dividing $H-h_j$ by V rather than v_j. In an equation containing a large number of variables, v_j and V will be approximately the same in most cases. The advantage of using V will be that the changes in residual due to the use of each respective variable will be strictly comparable. Let such a measure be defined:

$$Q_j = (v_j - V)/V = v_j/V - 1 = (H-h_j)/V \tag{32}$$

With Q_j, the interpreter can make accurate statements about the effect upon covariance of omitting each respective variable. For example, he might be able to say that the omission of one variable effects a reduction in covariance twice as much as does another variable. A statement of this nature cannot be made using the coefficient of partial correlation.

The Q coefficient indicates the proportion which V will increase if any given variable, x_j, is omitted. The square of the coefficient of partial correlation indicates the proportion which v_j will decrease if a given variable is added. Both of these are valid measures. It seems, however, that the Q coefficient has the advantage where a comparison of the significance of each variable is the objective.

Just as the square root of P_j was taken to give the coefficient of partial correlation, so should the square root of Q be taken to give a comparable measure of the significance of each variable. It must be remembered that both P_j and Q_j are based on the squares of the deviations of the predicted values from the observed values. Suppose that one deviation is numerically equal to 2 and the other is numerically equal to 4. One deviation will be 50% of the other. When deviations are squared, one will be only 25% of the other. To compensate for the tendency of the sum of the squares to exaggerate differences in deviations, the square root of measurements based on the sum of squares should be taken.

IX. THE F RATIO

Another measure of the contribution of each individual variable to the least-squares equation is the F ratio. Let the definition of $r = h/D$ be extended so that r is a value based on p variables selected from among the m variable. Then McNemar[24] gives an F test which may be expressed in the

---

24 McNemar, op. cit., p. 266

---

following manner:

(33) $$F = \frac{(R-r)/(m-p)}{(1-R)/(n-m)}$$

Snedecor's F table may be used with this ratio. The numerator has m-p degrees of freedom and the denominator has n-m degrees of freedom. Throughout this appendix F will be restricted to the case where the effects of one variable are omitted from the system of least-square equations. With this understanding, r_j will be used to show that the variable x_j has been omitted. Similarly F_j will denote (33) where only one variable, x_j, has been omitted. Where only one variable has been omitted,

(34) $$(m-p) = m - (m-1) = 1.$$

When (34) is used in (33), the F ratio becomes:

(35) $$F_j = (R-r_j)(n-m)/(1-R)$$

IX. THE F RATIO

Another measure of the contribution of each individual variable to the least-squares equation is the F ratio. Let the definition of $r = h/D$ be extended so that r is a value based on p variables selected from among the n variables. Then McNemar[24] gives an F test which may be expressed in the following manner:

---

24. McNemar, op. cit., p. 246

$$F = \frac{(R-r)/(n-p)}{(1-R)/(n-n)} \tag{33}$$

Snedecor's F table may be used with this ratio. The numerator has n-p degrees of freedom and the denominator has n-n degrees of freedom. Throughout this appendix F will be restricted to the case where the effects of one variable are omitted from the system of least-squares equations. With this understanding, r_j will be used to show that the variable x_j has been omitted. Similarly F_j will denote (33) where only one variable, x_j, has been omitted. Where only one variable has been omitted,

$$(n-p) = n - (n-1) = 1. \tag{34}$$

When (34) is used in (33), the F ratio becomes:

$$F_j = (R-r_j)(n-n)/(1-R) \tag{35}$$

Substituting (19) and (20) for R and r_j:

(36) $$F_j = (H-h_j)(n-m)/(D-H)$$

From (21) and (22),

(37) $$H - h_j = v_j - V$$

With this substitution in the numerator and with the substitution D-H = V in the denominator, (37) becomes:

(38) $$F_j = \frac{(v_j - V)(n-m)}{V}$$

When the F ratio given by McNemar is presented in form (38), it has a numerator that differs from the square of the coefficient of partial correlation (30) by a constant. The F ratio is even closer to the Q ratio. The entire ratio differs from the Q ratio only by the factor (n-m).

It must be emphasized that the discussion here pertains to the least-squares equation where first one variable then another is omitted. It does not change from one variable to another. Thus it may be looked upon as a constant. It follows from this, and a comparison of (30) and (36), that the Q ratio differs from the F ratio only by a constant. From this it follows that the F test is a measurement of the significance of the Q ratio. If F is significant, then the Q ratio, which differs only by a constant, is significant. The Q ratio is

the part of the F test that is used to compare one variable with another. The F test indicates whether or not a given change in residual was due to chance. This properly suggests the question, did the difference between two linear regression coefficients occur by chance. To answer this, a measure for the confidence limits of regression coefficients must be used.

X. A CONFIDENCE TEST FOR REGRESSION COEFFICIENTS.

With the variables given in equation (1), construct a least-squares equation where x_2 is the dependent variable and $x_3, x_4, \ldots x_m$ are independent variables. Just as the relationship between x_1 and $x_2, x_3, \ldots x_m$ was for convenience expressed by means of the determinant A, so can the relationship between x_2 and $x_3, x_4, \ldots x_m$ be expressed by means of the cofactor A_{11}. Just as the residual variance of x_1 was given by A/A_{11}, so will the residual variance of x_2 be given by $A_{11}/A_{11\cdot22}$. The cofactor $A_{11\cdot22}$ means that the cofactor of the first element will be taken, then from this cofactor, the cofactor of the first element (x_2) will be taken. In the determinant A_{11} it is easy to let x_3 replace x_2. To do this the second row of (4) must be moved to the position of the first row. If an odd number of rows intervenes, this will change the sign of the determinant to minus. Then the second column will be moved to the position of the first column. This will change the sign of the determinant back to plus. The residual deviance x_3 with $x_2, x_4, x_5, \ldots x_m$ will then become $A_{11}/A_{11\cdot33}$. The residual deviance of the other

variables may be determined in the same way. Let U_j be the residual deviance of the variable x_j, $j > 1$, where x_j is the independent variable and all the remaining variables in A are dependent. In general, the residual deviance of x_j is indicated by:[25]

(39) $$U_j = A_{11}/A_{11 \cdot jj}$$

---

25 Cramér, op. cit., p. 306

An inverse matrix may be represented as follows:[26]

(40) $$C^{-1} = (A_{ki}/A)$$

---

26 Ibid., p. 307

A_{ki} is the transpose of the cofactor A_{ik}. Since least-squares equations of the type discussed here are symmetrical, $A_{ki} = A_{ik}$, and (40) may be written:

(41) $$C^{-1} = (A_{ik}/A)$$

Assume now that the inverse matrix is taken of a matrix whose determinant is A_{11}. (Such a matrix was defined in Section II of the appendix.) By (41), the elements of the inverse

matrix will be:

(42) $$C^{-1} = (A_{11 \cdot ik}/A_{11})$$

In a diagonal, $j = k$.

In general, the elements of a diagonal of (40) will be:

(43) $$c^{jj} = A_{11 \cdot jj}/A_{11}$$

Note that the equation of least squares is expressed by (2). The solution to (2) is given in matrix form by:[27]

(44) $$B = C^{-1}G$$

---

27 Tintner, op. cit., p. 338

From equations (39) and (43):

(45) $$U_j = 1/c^{jj}$$

From (42) and (44) it will be noted that c^{jj} is an element from the inverse matrix used in the computation of the b coefficients. The residual deviance, U_j, may be obtained from the inverse matrix without computing a separate least-squares equation.

The square root of the following ratio is given by

Cramer:[28]

(46) $$t^2 = (n-m)(b_j^2 U_j)/V$$

28 Cramér, op. cit., p. 410

where b_j is given in (1) and is defined in vector notation in (44); where U_j is defined by (39) and V is defined in (15). The square root of (46) t, has a "Student's" t distribution with n-m degrees of freedom. If (45) is substituted for U_j and both the numerator and denominator of (46) are divided by (n-m), (46) becomes:

(47) $$t^2 = \frac{b_j^2}{Vc^{jj}/(n-m)}$$

The square root of (47) is given by Ezekiel[29] and Kenney and

29 Ezekiel, op. cit., p. 258

Keeping[30] as the test for the significance of b_j. The test may

30 Kenney and Keeping, op. cit., p. 314

be used to ascertain whether or not b_j differs significantly from some constant, k. In particular, if k = 0, the test is used to determine whether or not b_j differs significantly from zero. This is equivalent to answering whether or not a least-

squares equation would be just as effective if the variable x_j and its coefficient b_j were omitted.

XI. SIGNIFICANCE OF THE t TEST

The question that next presents itself is whether or not the t test is significantly different from the F test. What does the F test show that is not shown by the t test. It is proposed to answer this question by showing that the square of the t ratio (46) is identically equal to F_j (38). This identity might be expected from the nature of the F and t tests. McNemar has commented:

> "When we take $n_1 = 1$ and $F = t^2$, the equation of F can be transformed to that of t with df or $n = n_2$." 31

---

31 McNemar, op. cit., p. 343

To establish the proposed identity, a special case of Jacobi's theorem on complementary minors will be employed.[32]

---

32 For a treatment of Jacobi's theorem on complementary minors, see W. L. Ferrar, Algebra (Oxford: Clarendon Press), pp. 56-58.

Cramér[33] gives this case as follows:

squares equation would be just as effective if the variable x_j and its coefficient b_j were omitted.

XI. SIGNIFICANCE OF THE t TEST

The question that next presents itself is whether or not the t test is significantly different from the F test. What does the F test show that is not shown by the t test. It is proposed to answer this question by showing that the square of the t ratio (46) is identically equal to F_j (38). This identity might be expected from the nature of the F and t tests. McNemar has commented:

"When we take $n_1 = 1$ and $F = t^2$, the equation of F can be transformed to that of t with df or $n = n_2$."[31]

---

31 McNemar, op. cit., p. 319

To establish the proposed identity, a special case of Jacobi's theorem on complementary minors will be employed.[32]

---

32 For a treatment of Jacobi's theorem on complementary minors, see W. L. Ferrar, Algebra (Oxford Clarendon Press), pp. 56-58.

Cramer[33] gives this case as follows:

(48) $$\begin{vmatrix} A_{11} & A_{i1} \\ A_{1j} & A_{ij} \end{vmatrix} = A_{11}A_{ij} - A_{i1}A_{1j} = AA_{11\cdot ij}$$

In the pafticular case where i = j,

(49) $$A_{11}A_{jj} - A_{1j}^2 = AA_{11\cdot jj}$$

and

(50) $$A_j^2 = A_{11}A_{jj} - AA_{11\cdot jj}$$

---

33 Cramér, op. cit., pp. 110 - 111.

The residual deviance, v_j, may be expressed:

(51) $$v_j = A_{jj}/A_{jj\cdot 11} = A_{jj}/A_{jj\cdot 11}$$

The symbol v_j was defined in equation (22) in terms of r_j. The square of the coefficient of correlation, r_j, was used where one variable, x_j, was omitted from the least-squares equation. This means that in (3), the row and column arising from x_j will be deleted. This deletion may be accomplished by taking the cofactor, A_{jj}, of Z_{jj}. The cofactor will be equivalent to (3) except that the effects of x_j have been removed. Just as the residual deviance of x_{11} was taken from A to give $V = A/A_{11}$, so may the residual deviance be taken from A_{jj} to give $v_j = A_{jj}/A_{jj\cdot 11}$.

Substituting (5) and (39), (46) may be expressed as follows:

$$(52) \qquad t^2 = \frac{(n-m)(A_{1j})^2 A_{11}}{V(A_{11})^2 A_{11\cdot jj}} = \frac{(n-m)(A_{1j})^2}{V A_{11} A_{11\cdot jj}}$$

Substituting (50) for the numerator of the last expression,

$$(53) \qquad t^2 = \frac{(n-m)(A_{11}A_{jj} - AA_{11\cdot jj})}{V A_{11} A_{11\cdot jj}}$$

$$= \frac{(n-m)}{V} (A_{jj}/A_{11\cdot jj} - A/A_{11})$$

Substituting (51) and (7)

$$(54) \qquad t^2 = \frac{(v_j - V)(n - m)}{V}$$

Expression (54) is identical with (38). Therefore:

$$(55) \qquad F_j = t^2$$

Substituting (38) and (47) in (55),

$$(56) \qquad \frac{(v_j - V)(n - m)}{V} = \frac{(b_j)^2 (n-m)}{V c^{jj}}$$

and dividing both sides by $(n-m)/V$,

$$(57) \qquad v_j - V = (b_j)^2 / c^{jj}.$$

Solving for v_j,

$$v_j = V - (b_j)^2/c^{jj}. \tag{58}$$

If both sides of (57) are divided by v_j, the left side of (57) is identical with P_j as defined in (30). It follows that P_j may be written:

$$P_j = (b_j)^2/v_j c^{jj}. \tag{59}$$

Values for V can be computed by means of expression (16). This permits computation of the coefficient of partial correlation through the use of expressions (58) and (59).

XII. INTERPRETATION

The foregoing analysis has shown that the coefficient of partial correlation gives a measure of the effect of eliminating any given coefficient, x_j, from the least-squares equation. The magnitude of the coefficient of partial correlation depends primarily upon the change in residual that results from adding the variable x_j to the least-squares equation.

The Q coefficient was found to have an advantage over the coefficient of partial correlation in that it gave a more stable means of comparing the importance of variables which belong to a given equation. It was shown that the square root of both P_j and Q_j must be taken before a valid comparison of the importance of variables can be made. It was shown that

the Q ratio and F ratio differ by only a constant. It follows from this that the F ratio gives just as efficient a means of comparing variables as is given by Q_j or P_j. It was also shown that $F_j = t^2$. It follows from this that the t ratio gives just as good a means of comparing variables as does the Q ratio. In fact, the t ratio differs from the Q ratio only by a constant. For this reason, the t ratio has all the attributes necessary for comparing the importance of variables. If one variable is twice as important as another as indicated by the magnitude of the Q ratio, the fact that both variables are multiplied by the same constant will not destroy their comparative magnitudes.

Identity (56) shows that the t ratio contains within it the information essential for the coefficient of partial correlation. The decrease in residual due to the elimination (or addition) of a variable is contained in the term $v_j - V$. The residual deviance, V, is in a sense a measurement of multiple correlation. As V decreases, the coefficient of multiple correlation increases (assuming that D remains constant).

It follows from the foregoing argument that the t ratio has most of the qualities of the coefficient of partial correlation. The t ratio is preferable to the coefficient of partial correlation for the same reason that the Q ratio is preferable. The t ratio, like the F ratio, is a measure of the effect upon the residual of eliminating of a given variable. The t ratio

shows more than the importance of a given variable within an equation. It shows, in addition, the extent to which the the ability of the entire equation to make predictions would be decreased if a given variable were not used.

XIII. CALCULATION

Although the t ratio contains most of the information contained in the coefficient of partial correlation, all readers of research investigations may not be sufficiently aware of this fact. For this and other reasons, it may be desirable to have the coefficient of partial correlation. A method of computing this coefficient simultaneously with the computation of the t ratio is, therefore, needed.

The right side of (56) provides the basic formula for t^2. First, each regression coefficient is squared. The square of each regression coefficient is then divided by c^{jj}. Next, n-m is divided by V. The quotient (n-m)/V is then multiplied by each respective quotient $(b_j)^2/c^{jj}$. This gives t^2 for each x_j. At this point, the significance of each coefficient may be tested. Since $F_j = t^2$, the table for F may be used instead of the table for t. One degree of freedom and n-m degrees of freedom should be used. The square root of each ratio will then be taken to obtain the t ratio. The t ratios can be used to compare the relative importance of each variable.

The coefficient of partial correlation may be obtained

with a minimum of effort where the above order of computation is followed. To each quotient, $(b_j)^2/c^{jj}$ previously computed, V will be added. This gives v_j as shown by (58). Next, each of the quotients $(b_j)^2/c^{jj}$ will be divided by v_j. This gives equation (59). The square root of equation (59) will now be taken. The algebraic sign of each square root will be taken to correspond to the sign of the respective regression coefficient. This choice of signs is conventional.[34]

---

34 Paul G. Hoel, Introduction to Mathematical Statistics (John Wiley & Sons, 1947), p. 119

IMPORTANT!

Special care should be taken to prevent loss or damage of this volume. If lost or damaged, it must be paid for at the current rate of typing.

Date Due

JAN 5	1958		
FEB	5 RECD		
	DEC 13 1963		
NOV 18 RETD			

DIETER
BINDS
BETTER
DENVER COLO