FLIGHT TRANSPORTATION LABORATORY REPORT R85-6

TWA RESERVATIONS ANALYSIS: PROJECT UPDATE -- DEMAND DISTRIBUTION PATTERNS

Peter P. Belobaba

April 1985

FLIGHT TRANSPORTATION LABORATORY

R85-6

TWA RESERVATIONS ANALYSIS:
Project Update - Demand Distribution Patterns

Peter P. Belobaba

Flight Transportation Laboratory
Massachusetts Institute of Technology

April 1985

TWA RESERVATIONS ANALYSIS

Project Update - Demand Distribution Patterns

I. BACKGROUND

Research into airline capacity management and yield optimization at the Flight Transportation Laboratory is being performed with the support and funding of the Cooperative Research Program. The focus of this research over the past six months has been on patterns of demand for the different fare products available in airline markets. The objective of this work is to gain insight into the most complex component of the capacity management problem faced by airlines--variations in demand.

Trans World Airlines, a participant in the Cooperative Research Program, has given us access to reservations data and booking histories from its domestic operations over the past several years. A preliminary analysis of final reservations totals by day and by fare class for a sample of two transcontinental flights was undertaken to explore the database and identify its potential uses. The results of this preliminary analysis are discussed in an FTL report completed in April 1984.

The preliminary analysis examined trends in daily booking levels and their variability over the sample period. The reservations totals exhibited traditional seasonal and daily variations, except when disrupted by changes in product pricing and/or marketing. Of potential importance to the capacity management problem was the finding that the frequency distributions of demand over the sample period (and portions thereof) did not appear to be Normal (Gaussian) in shape, but rather were positively skewed. An intuitive explanation of such a distribution shape was suggested in the paper, namely that some base level of demand can generally be expected and that extreme values are more likely to be high relative to the mean number of reservations.

This issue of demand distribution patterns was pursued with further, more detailed, analysis of reservation data from a larger sample of TWA flights. This paper outlines the analysis that was undertaken and discusses the analysis results in the context of airline capacity management.

II. DATA SAMPLE

The original sample of two transcontinental flights was expanded to include additional flights in markets similar to those already analyzed. The data requested from TWA, as outlined in Table 1, included all TWA non-stop transcontinental flight segments that did not involve an origin or termination at JFK airport in New York. New York flights were excluded from the outset because the high proportion of international connecting traffic making use of TWA flights to/from JFK would have made the analysis of reservations data extremely complicated.

TWA provided reservations data and pre-departure booking histories for the ten flights shown in Table 1, for the period January 1 to December 31, 1983. The flights in the sample are comparable in terms of distance flown, frequency of service, nature of market served, and even departure times.

The analysis of this expanded data sample once again focussed on day-of-departure reservations totals by fare class for each flight operated during the sample period. Reservations totals could thus be calculated for each departing flight in each of the following categories: First Class, Ambassador (business) Class, and total coach compartment bookings, which could be further categorized as discount fare or full coach fare reservations.

TABLE 1: Data Sample Requested--TWA 1983 Reservations Data

1983 Flight Numbers	Origin-Destination	Distance (miles)	Depart Time	Aircraft Type
811/847	Boston-Los Angeles	2611	5:00 pm	L10
810/846	Los Angeles-Boston		9:00 am	L10
845/061	Boston-San Francisco	2704	6:30 pm	L10
754	San Francisco-Boston		9:00 am	L10
037	Philadelphia-Los Angeles	2401	6:00 pm	767
038	Los Angeles-Philadelphia		9:00 am	767
891	Washington-Los Angeles	2288	5:30 pm	L10/767
890	Los Angeles-Washington		9:00 am	L10/767
063	Washington-San Francisco	2419	5:00 pm	L10/767
064	San Francisco-Washington		8:30 am	L10/767

Source: Official Airline Guide, North American Edition, 1983

We should reiterate that these data consist of flight reservations totals, not actual traffic carried. Because the no-show/overbooking phenomenon has not been taken into account, these data represent a somewhat inflated estimate of demand, particularly for the fare classes with few booking and travel restrictions. Still, we continue to regard reservations totals as an indicator of passenger requests for flights that have been satisfied by the carrier and, therefore, as an acceptable proxy for demand in the capacity management context.

As is the case with most data collection efforts, not all the data provided by TWA proved to be useable for analysis. Data for two of the non-stop flights requested simply was not compiled in a useable format in the first place. One pair of flights was not operated from January through April of 1984. The resulting useable data sample is summarized in Table 2.

III. DATA ANALYSIS

The objective of this analysis was to produce and examine distribution plots of day-of-departure reservations by fare class for different markets and flight segments. Distribution shapes and parameters were to be compared among fare classes and markets to determine whether our preliminary findings of skewed demand distributions could be generalized.

Substantial editing of the data sample was required to create manageable subsets of the sample that could be used to produce demand distributions for flights with similar characteristics. Because reservation patterns differ from market to market, by direction of travel, by season and by day of week, the random variation in demand could only be identified with these systematic elements of variation eliminated. Analysis of demand distributions was thus undertaken as a nested process in which larger portions of the sample were examined first, so that smaller subsets with common characteristics could be identified for more detailed analysis.

Table 2 - DATA SAMPLE NOT USEABLE FROM OUTSET

FLIGHT	MARKET	J	F	M	A	M	J	J	A	S	O	N	D
037	PHL/LAX	X	X	X	X								
038	LAX/PHL	X	X	X	X								
063	IAD/SFO												
064	SFO/IAD												
890	LAX/IAD												
891	IAD/LAX												
810/846	LAX/BOS	X	X	X	X	X	X	X	X	X	X	X	X
811/847	BOS/LAX												
845/061	BOS/SFO												
754	SFO/BOS	X	X	X	X	X	X	X	X	X	X	X	X

X - Data not useable.

The entire 1983 sample period for 8 useable flights reflects variations over the course of the year which include seasonal peaks and troughs in demand, as well as exogenous (market pricing and other competitive) effects. Aggregation of the data from the entire sample would inevitably obscure the demand patterns of relevance to capacity management. Reservations data from any subset of the sample will necessarily include variations by day of week and direction of service, even if a single market (city-pair) is isolated. All these dimensions of variation can only be taken into account when individual flights or sets of flights from a relatively homogeneous period of the year are selected for analysis.

The analysis results will be discussed in two distinct sections below. In the first section, the process of identifying subsets of the sample that seem most suitable for more detailed examination is reviewed. As part of this discussion, several examples of the results generated in this exploratory phase are outlined. Comparisons are made among markets and among fare categories with respect to the systematic components of variation in reservations totals.

The subsequent section is devoted to the major questions addressed in this analysis: What can be conclude about the shapes of the distributions of reservation requests fulfilled for a particular sample of flight operations, and (how) do these distributions differ among fare classes? Discussion of these questions is based on the analysis of relatively homogeneous subsets of the entire sample, as identified in the exploratory data analysis phase.

IV. EXPLORATORY ANALYSIS

The initial data analysis phase involved examining the database and portions thereof in an attempt to identify the flights, periods of the year, and even days of the week that exhibited similar reservations patterns and characteristics.

Descriptive statistics of central tendency and spread were generated for sample subsets defined along a variety of combinations of time periods and markets in this search for homogeneity. The results of much of this exploratory analysis highlighted differences among flights and revealed patterns in reservations means and variation that should be accounted for in any analysis of demand distributions.

From the outset, no attempt was made to derive descriptive statistics for the <u>entire</u> sample of all 8 markets and the entire 1983 period, simply because these aggregate statistics would have revealed very little about similarities or differences in reservations means and variances. The exploratory analysis was therefore limited first to specific periods of the year and, ultimately, to specific flights operated during these periods. The periods considered were based on approximations of the traditional seasons of air travel demand, modified when necessary to account for major changes in the format of the data or significant changes in fare product pricing and/or restrictions. The flights included in each subset of the sample were based on similarities in the city-pairs served and/or direction of service.

The discussion in this paper of the exploratory phase of analysis includes the results obtained for three subsets of the data defined along the following dimensions:

(1) all flights for a homogeneous period of the year;

(2) flights operated in similar O-D markets for a homogeneous period; and

(3) flights operated in opposite directions in the same market, again for a homogeneous period of the year.

The analysis results for each type of data subset will be illustrated by a detailed discussion of one or more specific examples, in separate subsections below.

(a) All Flights, Homegeneous Sample Period

The most aggregate samples examined in the exploratory analysis of the database included reservations data from all of the flights, but only for selected periods of the year. The example to be discussed here is that of the August-December period of 1983. The last five months of the year were characterized in domestic transcontinental airline markets by a greater stability in fares and service levels, relative to the fare war period earlier in 1983. In fact, because of this stability, this period ultimately proved to be the most suitable for more detailed analysis of demand distributions, as will be discussed in Section V.

The patterns in reservations data aggregated over the eight available flights for the August-December period are worth outlining at this point as a basis for comparison with the smaller sub-samples to be examined. Daily means and K-factors were calculated for this sample period by day-of-week and class of service (Table 3), and summarized graphically in Figure 1. Each day-of-week value is based on a sample of approximately 155 to 170 data-points. Although the reservations data were in this case aggregated to the point that much of the variation in daily demand by flight may be obscured, several patterns that could well have been anticipated seem to emerge, and are described briefly below.

First, total reservations for all the fare classes peak significantly on Fridays and Sundays, and bottom out on Saturdays. This pattern is repeated in the three full-fare categories (First, Ambassador, and Coach classes). The

Table 3 - DAY OF WEEK VARIATION: 8 Transcontinental Flights August - December 1983

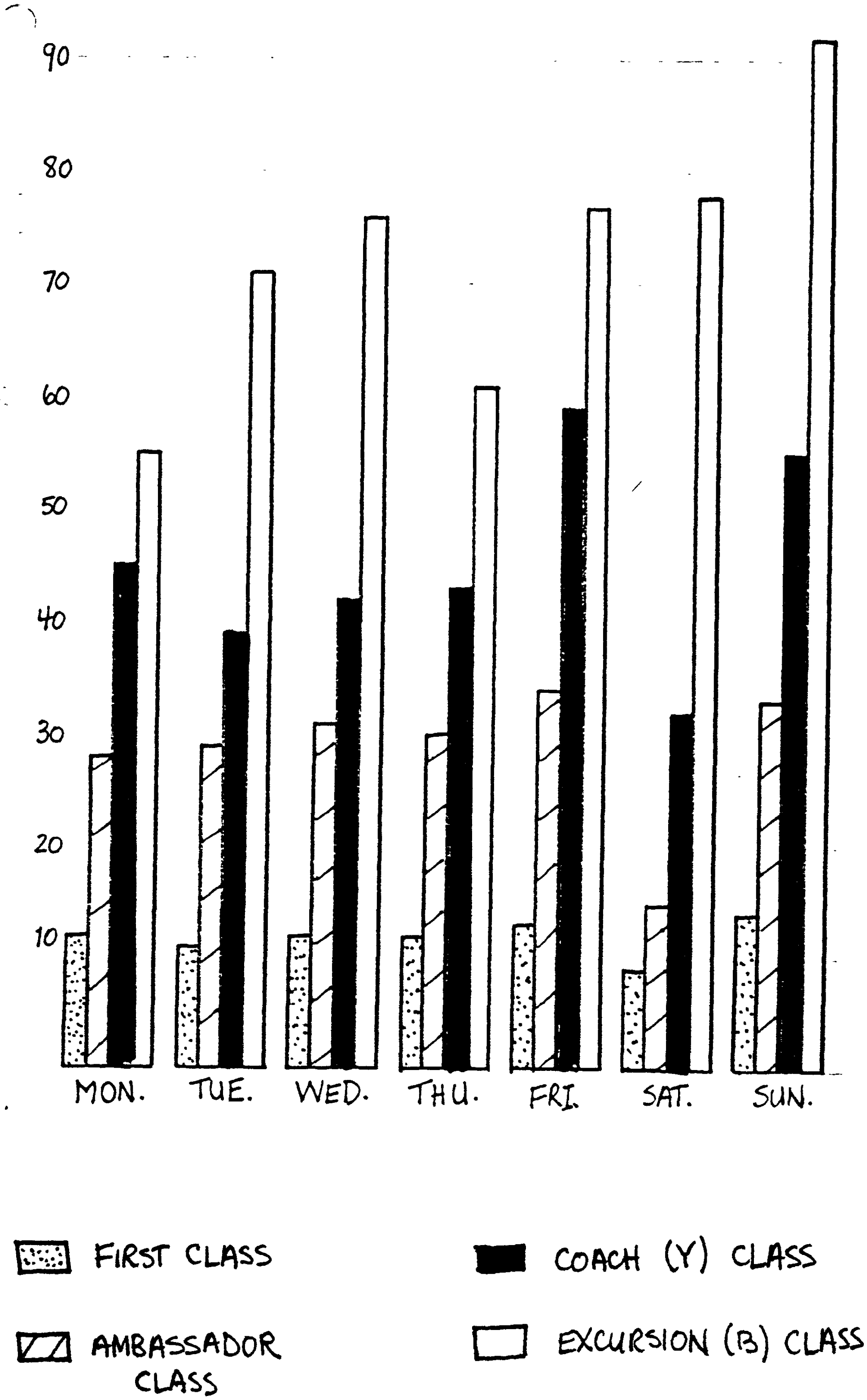
		MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY	SUNDAY
FIRST CLASS	$\bar{X}$	12	11	12	12	13	9	14
	K	.51	.45	.42	.47	.44	.64	.43
AMBASSADOR CLASS	$\bar{X}$	28	29	31	30	34	14	33
	K	.54	.45	.42	.47	.44	.64	.43
COACH (Y) CLASS	$\bar{X}$	45	39	42	43	59	32	54
	K	.46	.44	.44	.47	.44	.71	.48
EXCURSION (B) CLASS	$\bar{X}$	55	71	76	61	77	78	92
	K	.71	.62	.60	.72	.59	.53	.48
TOTAL * RESERVATIONS	$\bar{X}$	157	165	177	160	204	146	209
	K	.42	.35	.33	.39	.31	.46	.32

$\bar{X}$ = mean reservations for sample period

K = coefficient of variation = std. deviation $\div$ $\bar{X}$

* "Total" includes other fare types not listed.

FIGURE 1: DAY OF WEEK VARIATION – 8 Transcontinental Flights
August - December 1983



excursion fare (B-class) category, subject to capacity control actions in some cases and involving a different market segment of passengers, does not show as definite a peaking pattern. Sundays and Fridays are clearly days of higher demand, but booking levels for excursion class are also strong on Saturdays, Tuesdays and Wednesdays.

Saturday demand for excursion fare seats reflects the preference of most discretionary travelers for weekend departures and/or returns. The strong demand for excursion fare seats on Tuesdays and Wednesdays, on the other hand, can only be explained by the introduction of Tuesday/Wednesday "Super-Saver" fare levels lower than those applicable to travel the remainder of the week. These "BXE"-type fares were introduced in 1983 by most transcontinental carriers in August and September. These midweek fares appear to have been successful in leveling out total bookings. Monday through Thursday, leaving Saturday as the low point in total reservations. This pattern contrasts with results obtained for earlier periods of the year, particularly for individual flights

The week-to-week variation of reservations for a particular day of the week, by fare class, was measured by the coefficient of variation (k-factor) of the data points. Table 3 shows how aggregation of the fare categories into "Total Reservations" by day of week results in overall k-factors lower than those for individual fare classes. The k-factors for total reservations on a particular day of the week are in the .30 to .40 range, except for Saturdays (k=.46). Lower overall demand on Saturdays results in increased week-to-week variation in reservations, and higher k-factors are evident for the Saturday reservations levels on the full-fare categories.(F,C,Y) Otherwise, the variation in reservations for these three fare classes is remarkably similar, and is relatively constant for different days of the week.

The coefficients of variation are noticeably higher for excursion fare demand in this sample (k = .60 to .70), suggesting that B-class reservations are potentially less predictable than reservations for the full-fare categories. It should be mentioned, however, that other factors could well be responsible for the higher variation of excursion fare reservations in this aggregation of eight flights over a five-month period. Any conclusions as to the variation by fare class is best left to more detailed analysis of a more homogeneous subset of the data.

Nonetheless, we can at this point note that, even for aggregated reservations data, there appear to be significant differences in both the pattern of reservations by day-of-week for different fare categories and in the week-to-week variation in bookings by fare class. These differences should be remembered as we proceed to exploratory analysis of smaller subsets of the data and ultimately to analysis of the distributions of demand by fare category.

(b) Flights in Similar Markets, Same Direction

One type of smaller subset of the database that was chosen for exploratory analysis involved similar flights operating in the same transcontinental direction. The example to be discussed here is that of the two Boston-originating transcontinental flights in the database, examined over the traditionally low winter demand period from January through March, 1983. The objective in analyzing this type of subset of the data was to identify and compare the patterns of day-of-week variation in bookings by fare category for pairs of flights operated in the same direction of service.

The day-of-week variation by fare class for the Boston-L.A. and Boston-S.F. example of this analysis is summarized in Table 4, and portrayed graphically in

Table 4 - DAY OF WEEK VARIATION: COMPARISON OF TWO WESTBOUND FLIGHTS

Boston-L.A. and Boston-S.F., January - March 1983

			MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY	SUNDAY
FIRST CLASS	BOS/LAX	$\bar{X}$	12	19	18	16	21	8	18
		K	.31	.34	.26	.35	.27	.62	.38
	BOS/SFO	$\bar{X}$	9	9	11	12	13	5	10
		K	.42	.50	.44	.42	.37	.62	.45
AMBASSADOR CLASS	BOS/LAX	$\bar{X}$	22	21	30	30	43	13	36
		K	.49	.58	.59	.52	.40	.79	.43
	BOS/SFO	$\bar{X}$	10	9	12	15	24	4	13
		K	.55	.41	.56	.74	.65	1.14	.49
COACH (Y) CLASS	BOS/LAX	$\bar{X}$	93	82	92	79	90	61	92
		K	.66	.60	.75	.70	.70	.67	.72
	BOS/SFO	$\bar{X}$	53	39	50	54	60	28	68
		K	.64	.53	.47	.52	.69	.44	.80
EXCURSION (B) CLASS	BOS/LAX	$\bar{X}$	136	130	146	150	149	164	149
		K	.56	.57	.50	.51	.43	.48	.49
	BOS/SFO	$\bar{X}$	66	52	70	97	127	107	85
		K	.99	.84	.92	.88	.71	.86	.85

Figure 2: COMPARISON OF TWO WESTBOUND FLIGHTS — January - March 1983

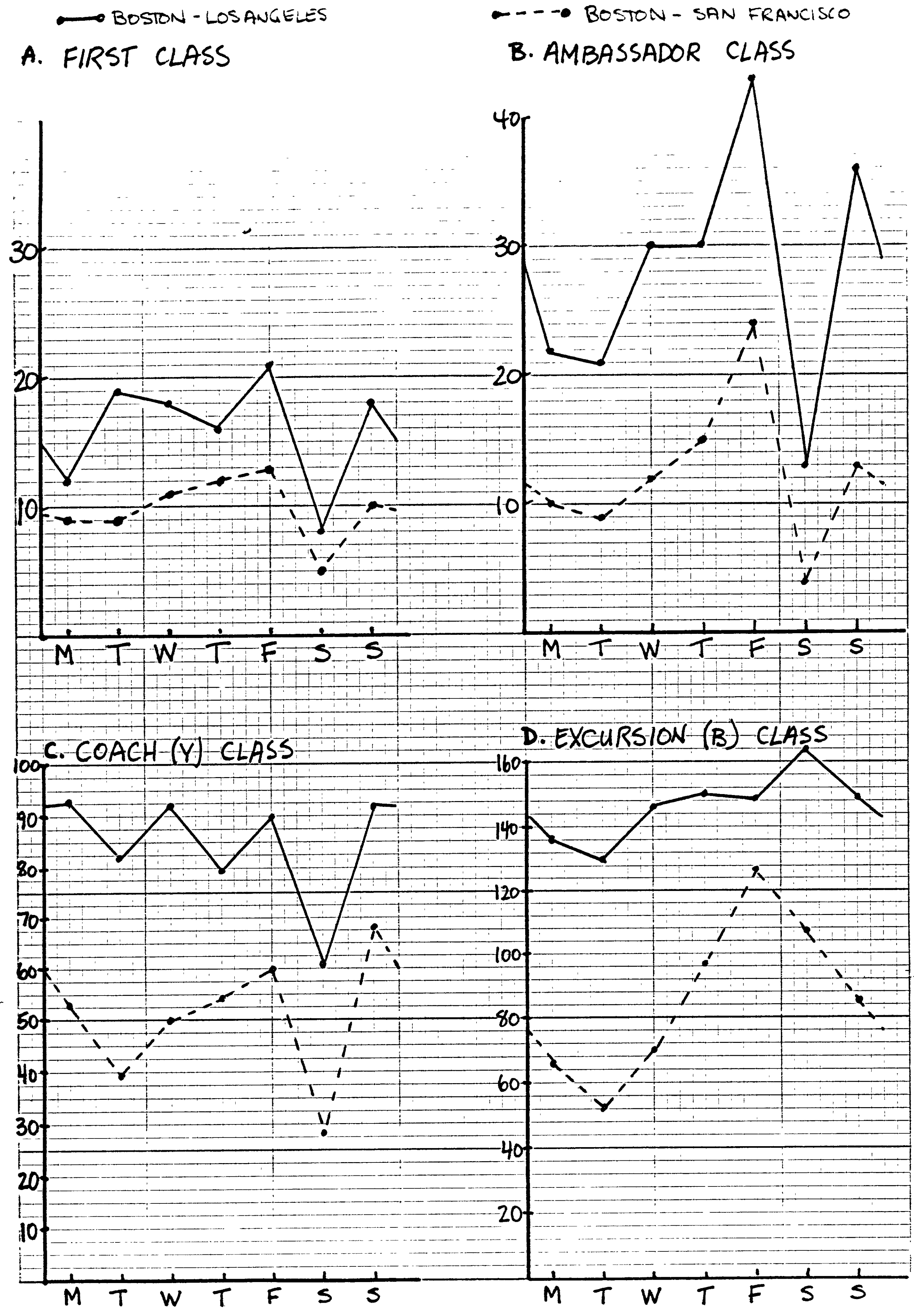


Figure 2. Overall, the demand for the Bos/LAX flight was significantly and consistently higher in all fare categories over the three-month period. Comparisons of the data from these two flights, as well as from other pairs of flights, made it clear that combining data from flights with large absolute differences in reservations levels would not be a valid method for increasing the number of data points available for assessing demand distribution patterns.

Some aspects of the day-of-week variation in reservations for these flights are worth noting, however, particularly the differences in k-factors between fare classes. For example, in the two premium fare categories (First Class and Ambassador Class), the k-factors for the more popular Bos/LAX flight are generally lower than those for the Bos/SFO flight, although the difference is not as pronounced in Ambassador Class. One reason for this result is the statistical property that, when similar data is involved, samples with higher means will tend to have lower K-factors than samples consisting of small absolute data values. Note the extremely high k-factors in First and Ambassador Class bookings in both markets for Saturday, a day with relatively low demand in the full-fare categories.

The principle of higher k-factors for lower demand markets is contradicted, however, by the reservations data for the full-fare coach category (Y). Despite much higher reservation means for each day of the week, the Bos/LAX flight exhibited higher week-to-week variation in bookings for each day of the week (see Table 4). Conversely, the coefficients of variation are higher for the Bos/SFO flight in the excursion fare category (B) while demand was lower, perhaps because B-class bookings were not restricted as often on the Bos/SFO flight due to a greater availability of seats.

The overall patterns of daily reservations totals in each fare class for the two flights, as shown in Figure 2, are similar despite a large difference in the absolute number of bookings. Day-to-day trends are most alike in the Ambassador and Coach fare categories, with mean daily bookings peaking on Fridays and Sundays, and bottoming out on Saturdays. The daily means in the Excursion category vary far less for the Bos/LAX flight by day of week, which could again be explained by a more frequent application of reservations limits on this flight.

This exploratory analysis of data from the Bos/LAX and Bos/SFO flights for the first 3 months of 1983 demonstrated that, while many similarities exist between the reservations data for the two flights, there also exist significant differences that would make sweeping comparisons of their respective demand distributions questionable. There are several characteristics of the data that would affect such comparisons:

(1) The day-of-week patterns in reservations means appear to be similar for the same direction of flight operations in similar markets, despite absolute differences in booking levels for each market.

(2) Although the data exhibits the general property that relative variation decreases as sample means increase, there are examples for which this relationship does not hold, due to unexplained factors.

(3) A comparison of the capacity controlled reservations data (B-class) for the two flights suggests that any distributional analysis should take into account the extent to which capacity controls were applied to the respective flights.

(c) <u>Flights in Same City-Pair Market, Opposite Directions</u>

Another type of data subset examined consisted of pairs of eastbound and westbound flights operated between city-pairs, once again for a relatively homogeneous period of the year. For example, the pair of flights operated in opposite directions between Philadelphia and Los Angeles were examined for the May to July period of 1983, to illustrate reservation patterns by fare class and day-of-week and to establish how the reservations data are affected by direction of service. Comparisons of the two flights in this case can be made somewhat more confidently than in the example described above, since here we are dealing with a single city-pair market.

The day of week variations in reservations means for this subset of the data are summarized in Table 5. Figure 3 shows graphically how the overall patterns of day of week variation for the two flights differ considerably among the fare classes. For First Class bookings during this period, the day-of-week variation is very similar for both directions of service. The coefficients of variation for the days of the week show no unexpected patterns, apart from displaying an inverse relationship with the mean booking levels. (Saturday First Class bookings have the lowest sample means and highest coefficients of variation for both flights).

On the other hand, directionality by day of week is more pronounced for both Ambassador Class bookings and bookings in the "Y" (Coach full-fare) category (see Figure 3 (b) and (c)). The expected pattern of Friday and Sunday peaks in reservations is evident for the <u>westbound</u> flight, while reservations for the eastbound flight are more evenly distributed throughout the week. In Y-class, the reservations peak for the eastbound flight occurs on Saturday, perhaps due to the "all-day" nature of the eastbound trip. For all three full-fare categories, the westbound flight exhibits greater day-of-week variation than its eastbound counterpart.

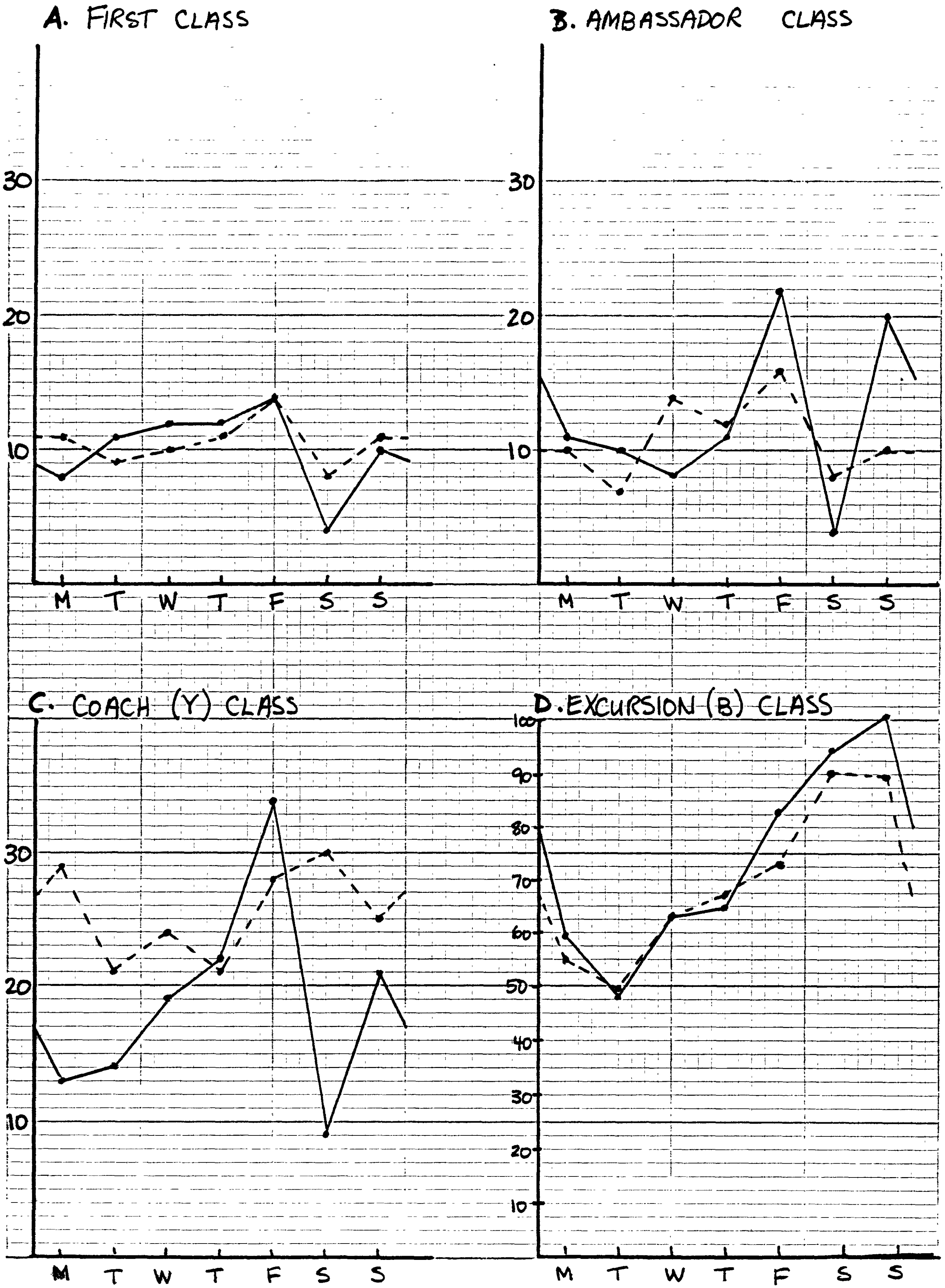
Table 5 - DAY OF WEEK VARIATION: A COMPARISON BY DIRECTION OF SERVICE

Philadelphia-Los Angeles-Philadelphia, May - July 1983

			MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY	SUN
	Westbound	$\bar{X}$	8	11	12	12	14	4	1
		K	.48	.40	.34	.40	.37	.77	.4
	Eastbound	$\bar{X}$	11	9	10	11	14	8	1
		K	.49	.54	.50	.52	.28	.63	.4
	Westbound	$\bar{X}$	11	10	8	11	22	4	2
		K	.54	.52	.55	.68	.57	1.17	.6
	Eastbound	$\bar{X}$	10	7	14	12	16	8	1
		K	.60	.70	.74	.48	.64	.57	.5
	Westbound	$\bar{X}$	13	14	19	22	34	9	2
		K	1.01	.70	.62	.62	.63	.83	.7
	Eastbound	$\bar{X}$	29	21	24	21	28	30	2
		K	.68	.68	.51	.53	.48	.99	.5
	Westbound	$\bar{X}$	59	48	63	65	83	94	10
)		K	.36	.59	.56	.44	.36	.41	.2
	Eastbound	$\bar{X}$	55	49	63	67	73	90	8
		K	.43	.45	.39	.30	.30	.29	.3

Figure 3: COMPARISON BY DIRECTION OF SERVICE — May - July 1983

●—● WESTBOUND (PHL/LAX)　　●- - -● EASTBOUND (LAX/PHL)



The k-factors for each day of the week in the Ambassador and Coach classes are high, relative to traditional estimates for total flight demand (i.e., 0.5 to 0.7 or higher in this case versus 0.3 to 0.5 in the literature). Aggregating the reservations data over all or most of the fare categories does in fact reduce the k-factors closer to the expected range. For the purposes of capacity management, however, the higher variation for the individual fare classes is extremely important.

Whether by coincidence or by design through capacity control techniques, the pattern of reservations by day of week for the two directions of service is virtually identical in the excursion fare (B-class) category. (See Figure 3(d)). The peaking of excursion class reservations on the weekend suggests that capacity controls may not be responsible for this pattern, particularly since full fare (Y) demand is not high enough to displace excursion reservations requests midweek. On the other hand, the lower coefficients of variation for each day of the week in excursion class, relative to the full-fare categories, suggest either that capacity management techniques were applied or that excursion class demand is relatively stable from week to week over a given sample period.

In summary, exploratory analysis of data from pairs of eastbound and westbound flights in individual city-pair markets generated some additional observations worth considering in examining the distributional shapes of demand of fare class:

(1) There can exist significant directionality in the day-of-week variation in reservations means for the various fare classes;

(2) This directionality of flow can differ significantly among the fare classes available for the flights involved;

(3) It is possible that the coefficients of variation for reservations on a particular day of the week in the individual fare categories may be substantially higher than the traditional values assumed for total flight demand.

(4) The patterns exhibited by the excursion fare bookings in this case (i.e. similar day of week pattern and lower coefficients of variation) again suggest the need to account for capacity control techniques that might have affected the reservations data in the first place.

V. DISTRIBUTIONAL PATTERNS BY FARE CLASS

Many of the statistical properties of aggregate as opposed to disaggregate subsets of the database that were revealed by the exploratory data analysis described above also influenced the examination of the frequency distributions of reservations by fare category. The data from larger subsets aggregated over several flights, months or days of the week produced smooth distribution shapes with identifiable characteristics, while the distributions for smaller subsets of more homogeneous data showed less consistency and were more difficult to assess. Because the information provided by distributions of reservations for disaggregate subsets is most useful in capacity management, the challenge in this phase of the analysis was to find subsets with similar characteristics which were also large enough to permit some valid conclusions to be made about demand distributions.

As part of this distributional analysis, histograms of the reservation totals in a given sample were generated as discrete approximations of reservation distributions. For many of the smaller samples examined at a disaggregate level, quantile plots and Normal probability plots were produced so that the distribution of data points could be assessed. The data samples examined first were the larger, aggregate subsets mentioned above. The focus of the analysis then shifted to very disaggregate subsets so that outliers and obvious holiday effects could be identified. Finally, some of these smallest subsets were combined to produce distributions and plots for a slightly larger, yet very homogeneous, data sample. Each of these steps will be outlined in separate sub-sections below.

(a) Distributions of Aggregate Data Subsets

Creation of what we refer to as "aggregate" data subsets for distributional analysis was based both on aggregation of reservations data by fare class from a single flight for all operations over an extended period of the year, and on aggregation of all flights in the useable database by day of the week over the same periods. Each of these aggregation procedures inevitably hides some of the variation important to capacity management, as will be discussed below. Nevertheless, the larger data subsets were plotted, and the results are outlined here because the assumption of Normally distributed requests/demand seems to fall into question no matter how the data are aggregated.

The distributions generated for several aggregate subsets for the period from January through June 1983 will be discussed here as examples of the analysis process. For these larger subsets, datapoints were simply placed into frequency histograms by fare class. These histograms were computer-generated, meaning the scales were determined by the computer as well.

The simplest categorization of the January-June data involved creation of frequency distributions by fare class for all operations of a particular flight during the period. Figure 4 shows examples of distributions of coach (Y) fare

Figure 4: Distributions of Daily Y-Class Reservations January - June 1983

FLIGHT 63 $\bar{x} = 46$ $k = .52$

FROM	TO	COUNT	SCALE=1 * PER 1.0000 OBSERVATIONS
5.0000	15.000	12	************
15.000	25.000	24	************************
25.000	35.000	32	********************************
35.000	45.000	32	********************************
45.000	55.000	20	********************
55.000	65.000	22	**********************
65.000	75.000	18	******************
75.000	85.000	7	*******
85.000	95.000	8	********
95.000	105.00	6	******
105.00	115.00	0	
115.00	125.00	1	*

FLIGHT 64 $\bar{x} = 43$ $k = .51$

FROM	TO	COUNT	SCALE=1 * PER 1.0000 OBSERVATIONS
9.0000	17.417	23	***********************
17.417	25.833	18	******************
25.833	34.250	30	******************************
34.250	42.667	29	*****************************
42.667	51.083	27	***************************
51.083	59.500	14	**************
59.500	67.917	16	****************
67.917	76.333	10	**********
76.333	84.750	5	*****
84.750	93.167	5	*****
93.167	101.58	3	***
101.58	110.00	2	**

FLIGHT 90 $\bar{x} = 55$ $k = .51$

FROM	TO	COUNT	SCALE=1 * PER 1.0000 OBSERVATIONS
10.000	22.417	17	*****************
22.417	34.833	28	****************************
34.833	47.250	34	**********************************
47.250	59.667	29	*****************************
59.667	72.083	22	**********************
72.083	84.500	15	***************
84.500	96.917	11	***********
96.917	109.33	7	*******
109.33	121.75	4	****
121.75	134.17	2	**
134.17	146.58	0	
146.58	159.00	1	*

FLIGHT 91 $\bar{x} = 48$ $k = .53$

FROM	TO	COUNT	SCALE=1 * PER 1.0000 OBSERVATIONS
6.0000	16.667	16	****************
16.667	27.333	17	*****************
27.333	38.000	30	******************************
38.000	48.667	32	********************************
48.667	59.333	27	***************************
59.333	70.000	13	*************
70.000	80.667	14	**************
80.667	91.333	10	**********
91.333	102.00	3	***
102.00	112.67	3	***
112.67	123.33	1	*
123.33	134.00	2	**

Figure 5: Distributions of Daily B-Class Reservations January-June 1983

FLIGHT 63 $\bar{x}=67$ $k=.53$

FROM	TO	COUNT	SCALE=1 * PER 1.0000 OBSERVATIONS
9.0000	23.583	19	*******************
23.583	38.167	26	**************************
38.167	52.750	32	********************************
52.750	67.333	23	***********************
67.333	81.917	25	*************************
81.917	96.500	15	***************
96.500	111.08	20	********************
111.08	125.67	10	**********
125.67	140.25	9	*********
140.25	154.83	1	*
154.83	169.42	1	*
169.42	184.00	1	*

FLIGHT 64 $\bar{x}=73$ $k=.53$

FROM	TO	COUNT	SCALE=1 * PER 1.0000 OBSERVATIONS
11.000	24.167	13	*************
24.167	37.333	24	************************
37.333	50.500	22	**********************
50.500	63.667	29	*****************************
63.667	76.833	25	*************************
76.833	90.000	11	***********
90.000	103.17	10	**********
103.17	116.33	15	***************
116.33	129.50	12	************
129.50	142.67	11	***********
142.67	155.83	6	******
155.83	169.00	4	****

FLIGHT 90 $\bar{x}=80$ $k=.51$

FROM	TO	COUNT	SCALE=1 * PER 1.0000 OBSERVATIONS
8.0000	23.250	12	************
23.250	38.500	20	********************
38.500	53.750	19	*******************
53.750	69.000	21	*********************
69.000	84.250	25	*************************
84.250	99.500	19	*******************
99.500	114.75	15	***************
114.75	130.00	14	**************
130.00	145.25	13	*************
145.25	160.50	7	*******
160.50	175.75	3	***
175.75	191.00	2	**

FLIGHT 91 $\bar{x}=77$ $k=.52$

FROM	TO	COUNT	SCALE=1 * PER 1.0000 OBSERVATIONS
2.0000	16.667	7	*******
16.667	31.333	10	**********
31.333	46.000	25	*************************
46.000	60.667	22	**********************
60.667	75.333	25	*************************
75.333	90.000	21	*********************
90.000	104.67	13	*************
104.67	119.33	15	***************
119.33	134.00	13	*************
134.00	148.67	10	**********
148.67	163.33	5	*****
163.33	178.00	2	**

reservations for flights 063 (Washington-S.F.), 064 (S.F.-Washington), 090 (Los Angeles-Washington), and 091 (Washington-L.A.). Figure 5 does the same for excursion (B) class reservations over this same period.

The most obvious characteristic common to all eight distributions in Figure 4 and 5 is a degree of positive skewness. That is, the distances from the median to data points above the median are greater than the corresponding distances to data points below the median. Visually, the skewness is apparent in the longer right-side "tail" of the distribution.

The skewness in these distributions is not particularly meaningful, since aggregation of reservations data over an extended (in this case, 6-month) time period will necessarily include the data points of high demand that create the right-side tail in the distributions. This positively skewed pattern also reflects the fact that reservations for a flight cannot be negative, meaning any extremes must by definition be at the positive end of the distribution.

Similar aggregation effects were reflected in the distributions of reservation levels by fare class by day of week, when aggregated over all flights in the useable sample for the same January-June period. Figures 6 and 7 show these distributions of reservations levels for Ambassador Class and Coach (Y) Class, respectively. The pattern of positive skewness is once again clear, and can be attributed to the statistical reasons mentioned above.

Worth noting, however, are the differences in distribution shapes among the days of the week, both within each fare class and between fare categories. In Ambassador Class, the reservations levels and distributional shapes are generally similar for Monday through Thursday, showing definite peaks and moderate right-side tails. These distributions contrast with those for Friday

MONDAY $\bar{x} = 17$ $k = .68$

FROM	TO	COUNT	SCALE=1 * PER 1.0000 OBSERVATIONS
.00000	4.6667	10	**********
4.6667	9.3333	38	**************************************
9.3333	14.000	38	**************************************
14.000	18.667	32	********************************
18.667	23.333	17	*****************
23.333	28.000	8	********
28.000	32.667	9	*********
32.667	37.333	7	*******
37.333	42.000	6	******
42.000	46.667	3	***
46.667	51.333	2	**
51.333	56.000	2	**

TUESDAY $\bar{x} = 15$ $k = .61$

FROM	TO	COUNT	SCALE=1 * PER 1.1500 OBSERVATIONS
1.0000	5.5833	17	***************
5.5833	10.167	46	**
10.167	14.750	35	*******************************
14.750	19.333	38	**********************************
19.333	23.917	12	***********
23.917	28.500	13	************
28.500	33.083	3	***
33.083	37.667	1	*
37.667	42.250	3	***
42.250	46.833	1	*
46.833	51.417	2	**
51.417	56.000	1	*

WEDNESDAY $\bar{x} = 17$ $k = .62$

FROM	TO	COUNT	SCALE=1 * PER 1.2750 OBSERVATIONS
1.0000	6.0000	13	***********
6.0000	11.000	31	*************************
11.000	16.000	51	**
16.000	21.000	28	**********************
21.000	26.000	20	****************
26.000	31.000	12	**********
31.000	36.000	5	****
36.000	41.000	3	***
41.000	46.000	6	*****
46.000	51.000	2	**
51.000	56.000	0	
56.000	61.000	2	**

THURSDAY $\bar{x} = 17$ $k = .65$

FROM	TO	COUNT	SCALE=1 * PER 1.0750 OBSERVATIONS
2.0000	6.6667	22	*********************
6.6667	11.333	43	**
11.333	16.000	29	***************************
16.000	20.667	25	************************
20.667	25.333	20	*******************
25.333	30.000	9	*********
30.000	34.667	10	**********
34.667	39.333	6	******
39.333	44.000	2	**
44.000	48.667	4	****
48.667	53.333	1	*
53.333	58.000	2	**

FRIDAY $\bar{x} = 24$ $k = .65$

FROM	TO	COUNT	SCALE=1 * PER .70000 OBSERVATIONS
1.0000	6.0833	15	**********************
6.0833	11.167	28	**
11.167	16.250	26	**************************************
16.250	21.333	23	*********************************
21.333	26.417	11	****************
26.417	31.500	14	********************
31.500	36.583	8	************
36.583	41.667	7	**********
41.667	46.750	14	********************
46.750	51.833	9	*************
51.833	56.917	9	*************
56.917	62.000	4	******

SATURDAY $\bar{x} = 9$ $k = .94$

FROM	TO	COUNT	SCALE=1 * PER 1.2500 OBSERVATIONS
.00000	4.3333	50	**
4.3333	8.6667	47	**************************************
8.6667	13.000	16	*************
13.000	17.333	19	****************
17.333	21.667	7	******
21.667	26.000	2	**
26.000	30.333	5	****
30.333	34.667	3	***
34.667	39.000	1	*
39.000	43.333	1	*
43.333	47.667	0	
47.667	52.000	1	*

SUNDAY $\bar{x} = 20$ $k = .70$

FROM	TO	COUNT	SCALE=1 * PER .87500 OBSERVATIONS
.00000	5.0833	20	***********************
5.0833	10.167	35	**
10.167	15.250	31	************************************
15.250	20.333	25	*****************************
20.333	25.417	18	*********************
25.417	30.500	12	**************
30.500	35.583	10	************
35.583	40.667	13	***************
40.667	45.750	2	***
45.750	50.833	5	******
50.833	55.917	3	****
55.917	61.000	4	*****

Figure 6: Distributions of Daily Ambassador Class Bookings By Day-of-Week - All Flights (Jan-Jun 1983)

MONDAY $\bar{x} = 49$ k= .71

FROM	TO	COUNT	SCALE=1 * PER 1.2250 OBSERVATIONS
.00000	19.083	27	***************************
19.083	38.167	49	***
38.167	57.250	48-	**
57.250	76.333	15	***************
76.333	95.417	15	***************
95.417	114.50	10	**********
114.50	133.58	6	******
133.58	152.67	0	
152.67	171.75	0	
171.75	190.83	0	
190.83	209.92	1	*
209.92	229.00	1	*

TUESDAY $\bar{x} = 39$ k= .67

FROM	TO	COUNT	SCALE=1 * PER 1.1250 OBSERVATIONS
1.0000	14.833	21	*********************
14.833	28.667	41	***
28.667	42.500	45-	***
42.500	56.333	33	*********************************
56.333	70.167	19	*******************
70.167	84.000	7	*******
84.000	97.833	1	*
97.833	111.67	1	*
111.67	125.50	1	*
125.50	139.33	1	*
139.33	153.17	0	
153.17	167.00	2	**

WEDNESDAY $\bar{x} = 44$ k= .70

FROM	TO	COUNT	SCALE=1 * PER 1.2250 OBSERVATIONS
2.0000	18.167	25	*************************
18.167	34.333	48	**
34.333	50.500	49-	***
50.500	66.667	26	**************************
66.667	82.833	11	***********
82.833	99.000	8	********
99.000	115.17	0	
115.17	131.33	2	**
131.33	147.50	0	
147.50	163.67	1	*
163.67	179.83	0	
179.83	196.00	3	***

THURSDAY $\bar{x} = 45$ k=.68

FROM	TO	COUNT	SCALE=1 * PER 1.2250 OBSERVATIONS
2.0000	18.750	25	*************************
18.750	35.500	48	**
35.500	52.250	49-	***
52.250	69.000	23	***********************
69.000	85.750	15	***************
85.750	102.50	6	******
102.50	119.25	1	*
119.25	136.00	1	*
136.00	152.75	2	**
152.75	169.50	2	**
169.50	186.25	0	
186.25	203.00	1	*

FRIDAY $\bar{x} = 56$ k= .61

FROM	TO	COUNT	SCALE=1 * PER 1.0500 OBSERVATIONS
5.0000	20.333	26	**************************
20.333	35.667	22	**********************
35.667	51.000	42	**
51.000	66.333	21-	*********************
66.333	81.667	25	*************************
81.667	97.000	13	*************
97.000	112.33	8	********
112.33	127.67	4	****
127.67	143.00	2	**
143.00	158.33	2	**
158.33	173.67	1	*
173.67	189.00	2	**

SATURDAY $\bar{x} = 38$ k= .65

FROM	TO	COUNT	SCALE=1 * PER 1.0000 OBSERVATIONS
1.0000	12.500	17	*****************
12.500	24.000	37	*************************************
24.000	35.500	29	*****************************
35.500	47.000	24-	************************
47.000	58.500	18	******************
58.500	70.000	12	************
70.000	81.500	6	******
81.500	93.000	3	***
93.000	104.50	2	**
104.50	116.00	3	***
116.00	127.50	0	
127.50	139.00	1	*

SUNDAY $\bar{x} = 54$ k= .65

FROM	TO	COUNT	SCALE=1 * PER 1.0750 OBSERVATIONS
3.0000	21.083	25	*************************
21.083	39.167	43	***
39.167	57.250	38-	**************************************
57.250	75.333	35	***********************************
75.333	93.417	21	*********************
93.417	111.50	7	*******
111.50	129.58	4	****
129.58	147.67	0	
147.67	165.75	2	**
165.75	183.83	1	*
183.83	201.92	1	*
201.92	220.00	1	*

Figure 7: Distributions of Daily Coach Class (Y) Bookings By Day-of-Week — All Flights (Jan-Jun 1983)

and Sunday, which have much broader peaks (distribution modes) and thicker right-side tails. Note that the coefficients of variation for these distributions are approximately equal to the k-factors for Monday-Thursday. The k-factors are higher, however, for the distribution of Saturday reservations in Ambassador Class. The distribution mode is at the very low end of the scale and, as a result, the right-side tail of extreme values is much longer.

In looking at these Ambassador Class distributions, we can suggest a relationship between the distributional shapes and the mean levels of reservations experienced on the different days of the week. Conveniently, the scales on the distributions in Figure 6 are approximately equal, ranging from 0 to 55, where 54 seats is the capacity of Ambassador Class on an L-1011 aircraft. It seems that, given an upper bound constraint on capacity, the distributional shapes change as mean demand increases. The distribution mode shifts to the right, and the degrees of skewness diminishes with increased mean reservations levels.

The relationship between distributional shape and mean reservations by day of week is more difficult to decipher in the case of Coach Class reservations (Figure 7). The "upper limit" for Coach Class reservations is in fact the capacity of the rear compartment of the aircraft, and this capacity is very seldom reached by the Y-fare reservations bucket alone (since there are several other fare types booked in the rear compartment). The result is extremely long right-side tails created by the few occasions on which a substantial portion of this capacity is booked in the Y-fare class, or by data point "outliers". Nevertheless, there still appears to be positive skewness in the bulk of the distribution for each day of the week.

(b) Identifying Homogeneous Subsets

The second step in the analysis of reservations distributions involved identifying systematic variation in the data so that distributions reflecting primarily stochastic variation could be generated. Unfortunately, it is extremely difficult to account for all components of systematic variation and to separate them from strictly random variations. Furthermore, the stricter the criteria used for defining homogeneous subsets of the data, the smaller the eventual size of the subset. Some generalizations about reservations data characteristics had to be made in order to create data subsets large enough to produce distributions with recognizable shapes.

The first of these generalizations involved variations in demand characteristics due to seasonality and changes in service levels or prices over the year. As mentioned previously, the period from August through December 1983 exhibited the greatest stability in terms of fare levels and service patterns for the transcontinental markets included in the dataset. For this reason, this sample period was chosen as the base from which smaller subsets of data could be drawn. The period does include a portion of the summer peak travel demand period, but any extreme values would be edited out at a later stage.

The second step in the process involved identifying the specific TWA flights to be examined, out of the total available sample of eight flights. For purposes of consistency, the analysis ultimately focussed on the three flights that did not experience changes in aircraft gauge over the sample period:

		City-Pair	Aircraft
(1)	Flight 037:	PHL/LAX	B-767
(2)	Flight 038:	LAX/PHL	B-767
(3)	Flight 061:	BOS/SFO	L-1011

While the return SFO/BOS flight was also consistent with respect to equipment used over the sample period, the data for this flight was not useable from the outset.

Given the above sample of three flights operated over a 5-month period of 1983, further categorization of data subsets was based on variations in demand over this period by day of week for the different fare classes. In this step of the process, the reservations totals by fare class for each of the three flights were examined by day of week of operation. The subsets of the data sample thus consisted of, for example, the Ambassador Class reservations totals for all Monday operations of Flight 37 during the sample period. For each subset, individual observations were plotted and descriptive statistics were generated. The results of this exploratory analysis pointed to the need to edit out extreme values experienced during recognizable holiday periods and special events, so that the remaining data in the subsets came as close as possible to reflecting the random variation in demand from week to week.

Editing the extreme values from these subsets was a subjective effort in which outliers (both high and low) associated with flights operated before, during and after national holidays were removed. Each fare class was edited independently so that any differences in holiday effects on the various reservations categories could be taken into account. Furthermore, the editing process was directed at the "full-fare" categories (First, Ambassador, and Coach) because these categories come closest to representing total reservations requests for the respective fare products. Excursion and Discount reservations totals, on the other hand, are far more affected by capacity controls and management forecasts of demand in the full fare categories. Thus, while the Excursion Class data subsets were examined for patterns in demand, they were not edited for the purposes of generating reservations distributions.

As a brief example of the editing process, we can consider the subset of Coach category reservations for Monday operations of Flight 061 (Boston-San Francisco). A plot of booking levels by week (Figure 8) shows two low "outliers" occurring on Labor Day and the day after Christmas. There is also an extremely high value that occurred on October 3rd. While the high outlier was not associated with a particular holiday, the fact that the First and Ambassador categories also showed high outliers on this day suggested that some unusual event precipitated the high booking levels. All three outliers were thus removed from the data subset.

Such a subjective editing process could not remove all outliers caused by exogenous factors, but effectively accounted for the most obvious non-random demand variations within each data subset. The net effect of the editing process on the sample mean and coefficient of variation of each data subset edited is summarized in Tables 6-8. As one would expect, removal of extreme outliers served to reduce the coefficient of variation in virtually every subset edited. Further, high and low outliers were deleted with approximately equal frequency, so that the sample means of each subset did not change substantially in most cases.

The consistent exception to this pattern involves the Ambassador (C) Class subsets for all 3 flights. TWA reduced Ambassador Class fares relative to Coach (Y) fares in August in order to stimulate C-class traffic. As a result, the

Figure 8: BOOKINGS BY WEEK – BOS/SFO "Y" CLASS; MONDAYS, AUG-DEC 1983

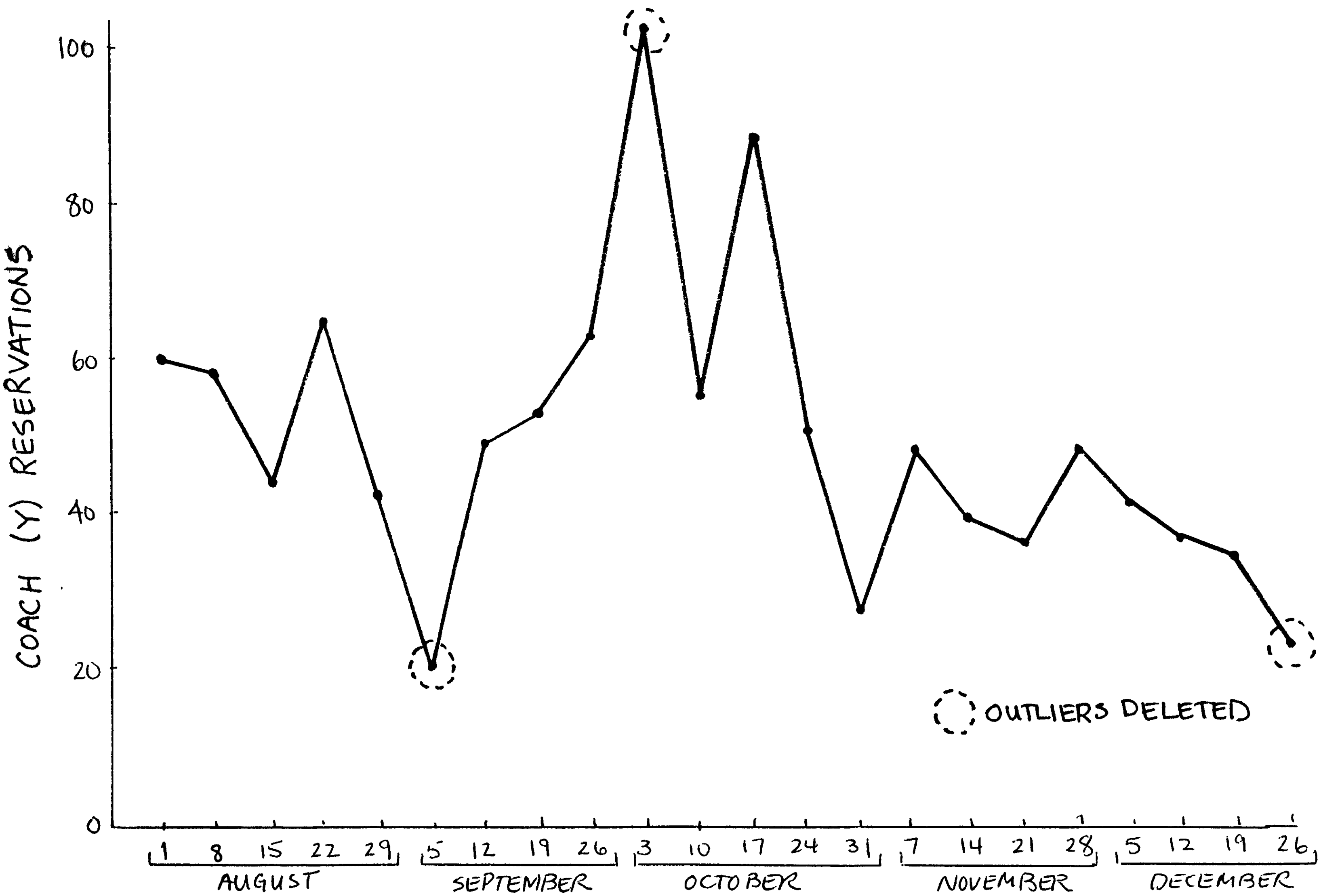


Table 6: Effect of Editing on Reservations Data Subsets
Flight 037 PHL/LAX August - December 1983

	Monday		Tuesday		Wednesday		Thursday		Friday		Saturday		Sunday	
	$\overline{X}$	K	$\overline{X}$	K	$\overline{X}$	K	$\overline{X}$	K	$\overline{X}$	K	$\overline{X}$	K	$\overline{X}$	K
FIRST CLASS														
Original	10	.44	11	.41	11	.47	11	.42	13	.46	5	.79	13	.34
Edited	10	.33	11	.36	12	.43	11	.40	14	.33	4	.68	14	.30
AMBASSADOR														
Original	20	.37	23	.38	24	.42	24	.43	29	.44	8	.63	30	.45
Edited	23	.26	26	.27	28	.30	30	.26	38	.21	11	.42	36	.23
COACH (Y)														
Original	28	.24	28	.35	35	.33	30	.28	45	.35	8	.36	36	.32
Edited	28	.24	30	.29	39	.19	31	.22	47	.29	8	.30	37	.24

Table 7: Effect of Editing on Reservations Data Subsets
Flight 038 LAX/PHL August - December 1983

	Monday		Tuesday		Wednesday		Thursday		Friday		Saturday		Sunday	
	$\overline{X}$	K	$\overline{X}$	K	$\overline{X}$	K	$\overline{X}$	K	$\overline{X}$	K	$\overline{X}$	K	$\overline{X}$	K
FIRST CLASS														
Original	11	.46	11	.48	10	.40	13	.39	12	.52	10	.55	12	.41
Edited	13	.27	10	.43	11	.33	13	.36	12	.45	8	.39	11	.40
AMBASSADOR														
Original	26	.54	22	.33	24	.52	25	.54	28	.47	15	.61	25	.39
Edited	34	.26	26	.20	29	.20	32	.29	36	.26	20	.38	29	.32
COACH (Y)														
Original	43	.33	28	.36	34	.41	37	.44	48	.33	23	.52	39	.42
Edited	45	.22	29	.31	33	.27	37	.22	53	.22	23	.38	37	.38

Table 8: Effects of Editing on Reservations Data Subsets
Flight 061 BOS/SFO August - December 1983

	Monday		Tuesday		Wednesday		Thursday		Friday		Saturday		Sunday	
	$\overline{X}$	K	$\overline{X}$	K	$\overline{X}$	K	$\overline{X}$	K	$\overline{X}$	K	$\overline{X}$	K	$\overline{X}$	K
FIRST CLASS														
Original	11	.51	12	.41	11	.30	9	.49	10	.42	5	.82	14	.55
Edited	11	.46	12	.36	11	.25	10	.43	11	.33	6	.71	15	.50
AMBASSADOR														
Original	19	.35	19	.41	22	.35	22	.45	27	.39	7	.74	27	37
Edited	19	.27	20	.30	23	.26	25	.33	31	.20	6	.52	27	.17
COACH (Y)														
Original	50	.40	50	.28	52	.28	53	.29	72	.27	36	.52	63	.34
Edited	49	.29	51	.27	56	.19	56	.20	76	.20	34	.44	65	.28

booking totals for August proved to be significantly lower than those for September through December for all days of the week, and were deleted from each of the subsets. The net effect on the Ambassador Class subsets was thus an increase in sample means and a large decrease in coefficients of variation.

With extreme values deleted, the number of data points in each sample subset was reduced to an average of approximately 16-20 observations. Distributional plots of each subset were generated in the form of frequency histograms. It was clear from these histograms, however, that the number of observations within each subset was generally too small to produce any identifiable or consistent pattern in the distributional plots. A sample of 18 points spread across 8-12 histogram ranges simply did not permit any generalizations about distribution characteristics to be made.

It was possible, however, to make use of other exploratory data analysis tools to gain additional insight into the distribution of data points within these smallest sample subsets. Box plots of each of the 63 subsets summarized in Tables 6, 7 and 8 (3 flights x 3 fare classes x 7 days of the week), were used to identify distributional differences and similarities among subsets, and Normal probability plots were generated so that the validity of the assumption of Normally distributed reservations totals could be assessed.

Figures 9, 10 and 11 show the boxplots of the edited day-of-week subsets for the three classes and each of the three flights analyzed. These boxplots graphically represent the median, upper quartile, lower quartile, and overall range of each data subset.

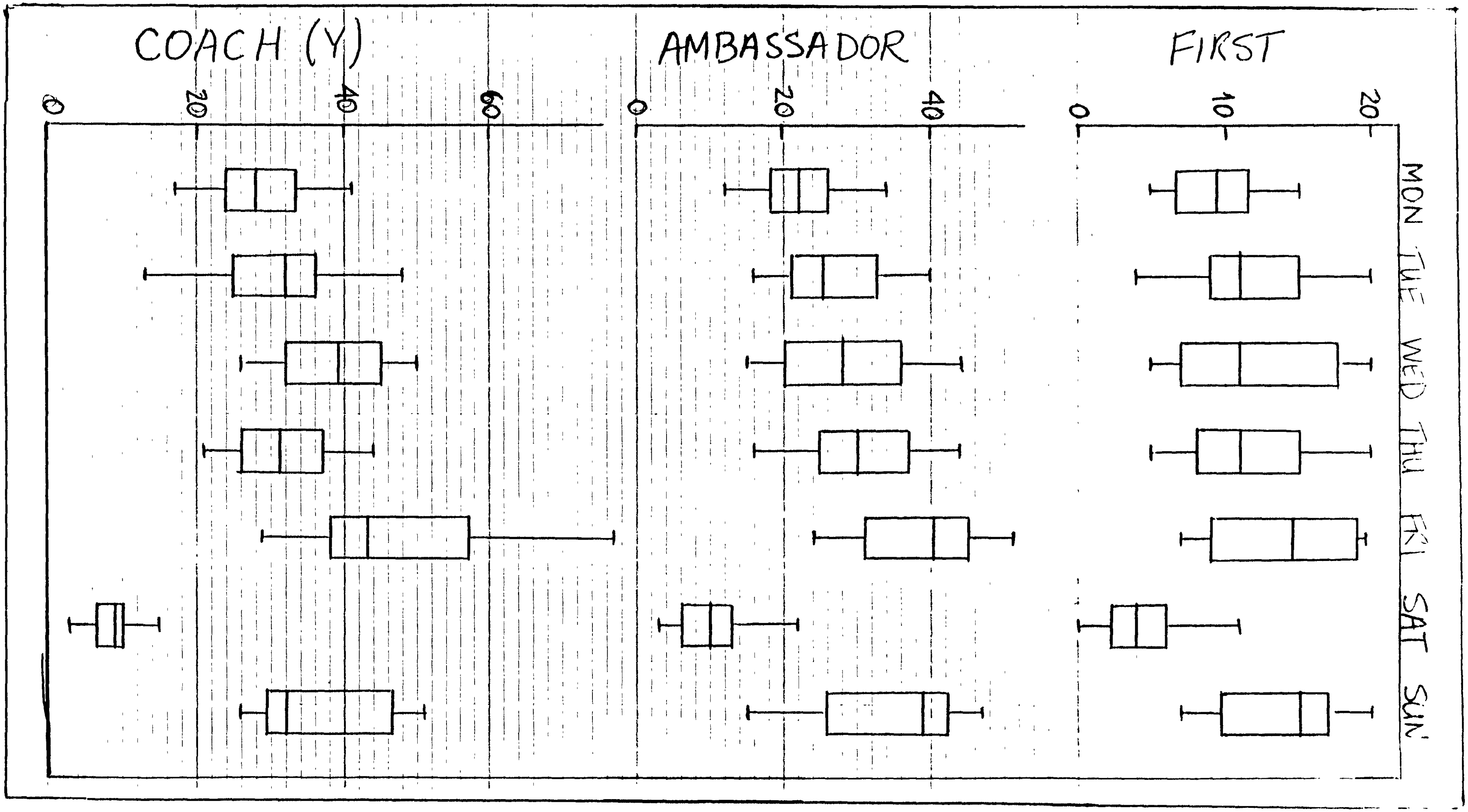


Figure 9: BOXPLOTS OF RESERVATIONS DATA BY DAY OF WEEK
FLIGHT 37 PHL/LAX AUG-DEC 1983

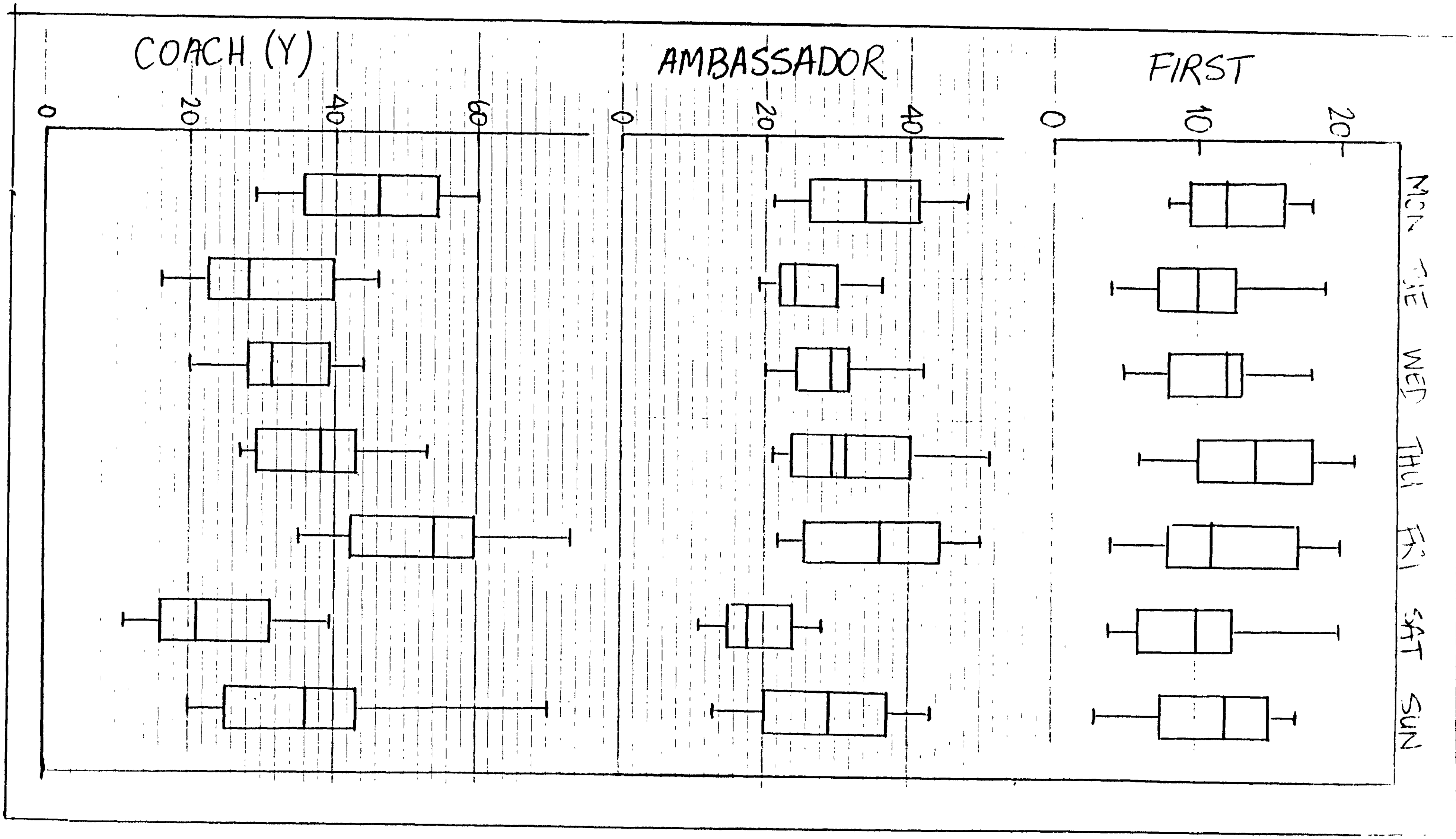


Figure 10: BOXPLOTS OF RESERVATIONS DATA BY DAY OF WEEK
FLIGHT 38 LAX/PHL AUG-DEC 1983

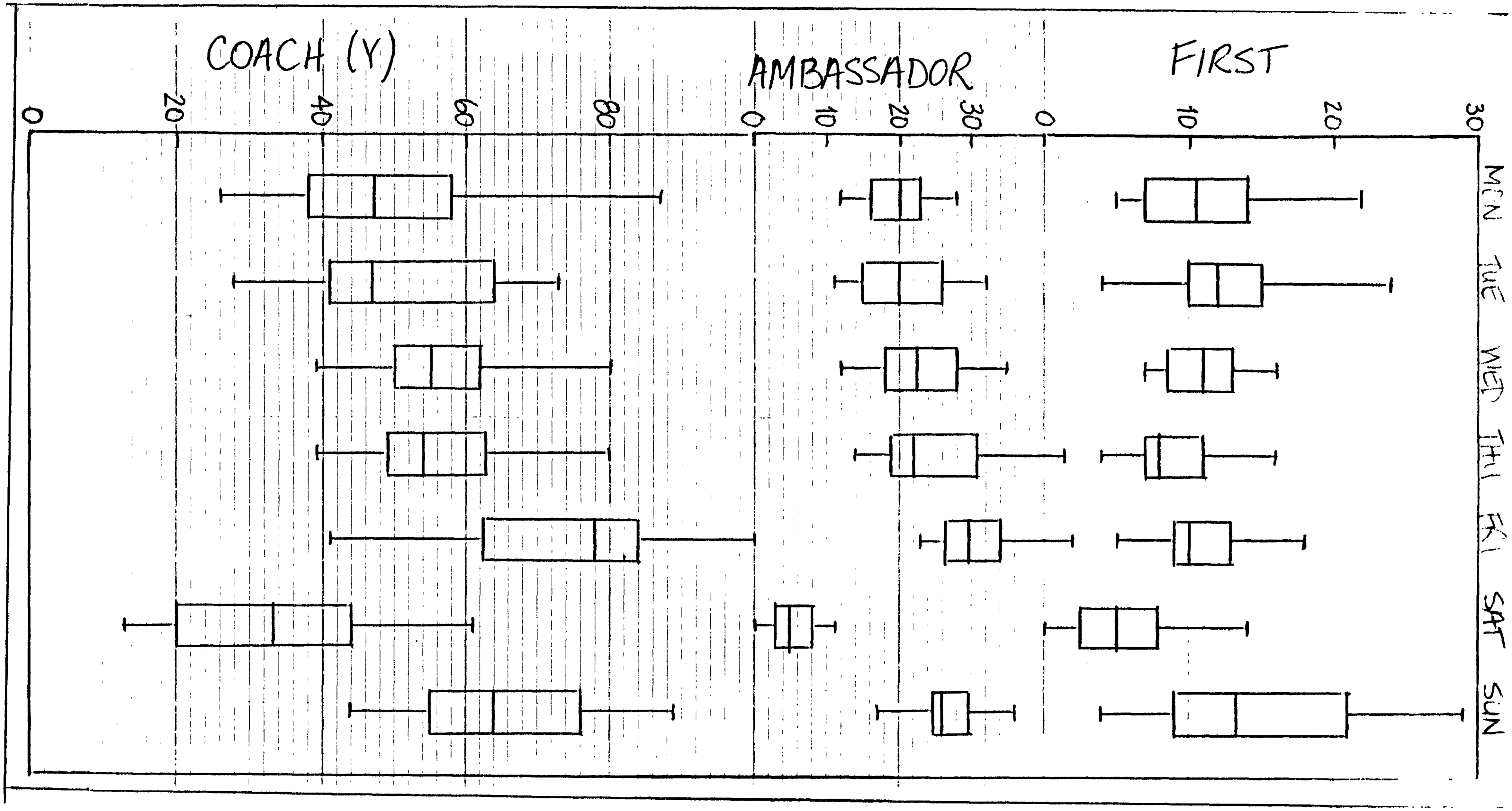


Figure 11 : BOXPLOTS OF RESERVATIONS DATA BY DAY OF WEEK
FLIGHT 61 BOS/SFO AUG - DEC 1983

Comparisions of the boxplots across days of the week for any particular flight and class of fare highlight the extent to which the distributions of data points within these smallest subsets differ. Overall, the peaking patterns discussed earlier are evident in the location parameters (medians) of each group. The distribution of each group, however, is determined by the spread of the data sample about the median, specifically by the relative sizes of the quartiles as well as the lengths of the upper and lower end tails. Focussing first on Figure 9 (Flight 37), we can make several general observations about the distributions. Relative spread of the data, as measured by the absolute size of the inter-quartile range (IQR) is lowest on Saturdays, when demand is also lowest. Among the remaining days of the week, relative spread is lowest on Mondays, Tuesdays and Thursdays, in all three fare classes. The IQR is generally largest on Fridays and Sundays, days of higher demand.

With respect to symmetry of the data subsets, we can say that the distributions appear to be relatively symmetrical for most of the days with lower mean reservations levels. The picture is substantially different, however, for the peak demand days, Friday and Sunday. It appears that, for Flight 37 at least, the "Y" reservations, which do not approach the coach compartment limit even on peak days, are positively skewed. On the other hand, the First and Ambassador Class reservations distributions are centered much closer to their respective capacity limits and reflect a negative skewness.

The corresponding boxplots for the eastbound Flight 38 (Figure 10) do not suggest as clear a pattern in terms of spread and skewness as that described for Flight 37, above. The peak days in the eastbound direction include Mondays in addition to Fridays and Sundays, and the Saturday trough is not nearly as pronounced as it is for westbound flights. In general, the distribution spreads (in terms of IQR length) are smallest on Tuesdays, Wednesdays and Saturdays, largest on peak demand days. A pattern in distribution skewness is not readily apparent from the Flight 38 boxplots, although positive skewness seems to be most prevalent in the "unconstrained" Y-fare category. However, the broader relationships between skewness and fare class capacity postulated for Flight 37 do not appear to hold for Flight 38.

The Flight 61 boxplots in Figure 11 are more comparable to the Flight 37 boxplots, at least in terms of directional day-of-week patterns. Differences between the spreads of the daily distributions are not as pronounced as for the previously discussed flights, particularly in the "Y" fare category. Overall, there is no clear relationship between distribution spreads and locations for the day-of-week subsets of the Flight 61 data.

There is a pattern, however, in distribution shapes, as measured by degree of symmetry. The majority of the Flight 61 subsets in all three fare categories appear to be positively skewed to some extent. Only the peak day (Friday) in the coach category shows any negative skewness. This pattern of positive skewness is generated by reservations data from a low load factor flight on which capacity limits were seldom reached.(recall Figure 2)

The boxplot comparisons thus suggest the possibility of a relationship between reservations distribution spreads and demand day-of-week, as measured by the median number daily reservations accepted. This relationship is by no means clear and in fact was not reflected in all the data sample subsets. Similar contradictions with respect to distribution symmetry were evident, although a general pattern of positive skewness in cases of low demand relative to fare class capacity and possible negative skewness for high demand cases could be postulated.

Normal probability plots of each subset were used to assess a more specific characteristic of the data distributions: the degree to which the data within each subset are Normally distributed. Normal probability plots provide a visual representation of how the actual data sample compares to a Normally distributed set of data of the same size and with the same parameters (mean and variance). The data conform exactly to the Gaussian model when the Normal probability plot of points is a perfectly straight line. In practice, we expect to see random deviations from a straight line, particularly for smaller samples. Systematic deviations from a straight line, however, suggest a poor fit of the Gaussian distribution to the data. For example, a positively skewed data distribution will generate points on the Normal probability plot that curve away from the straight line at the upper (top right) end of the plot.

Normal probability plots of the reservations data subsets, when examined as a group, proved to be inconclusive with respect to the consistency of fit between the Gaussian distribution and the data subsets. Figures 12, 13 and 14 show a sample of Normal probability plots for Flights 37, 38 and 61, respectively. The plots included in these figures were selected to illustrate both good fits and poor fits, and were also paired by day of week to permit comparisions between Ambassador Class and Coach reservations distributions.

The plots in Figure 12 (Flight 37) illustrate cases in which the fit of the Gaussian model to the data proved to be similar for the two fare classes for the same days of the week. The Monday data subsets seem to fit the model well for both classes, while the Sunday subsets show a pattern of deviation from the Normal model. The Sunday coach class subset shows substantial positive skewness, as

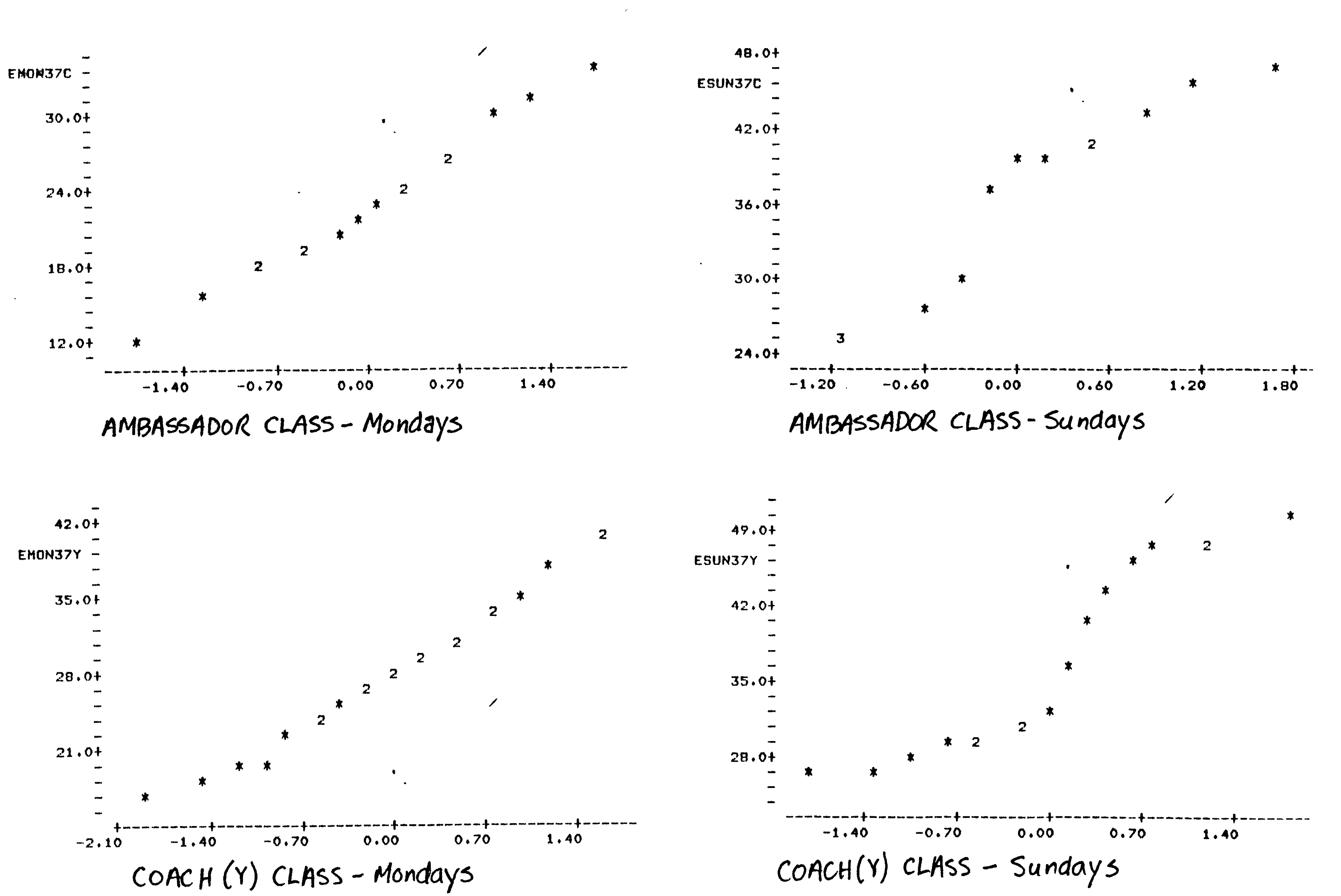
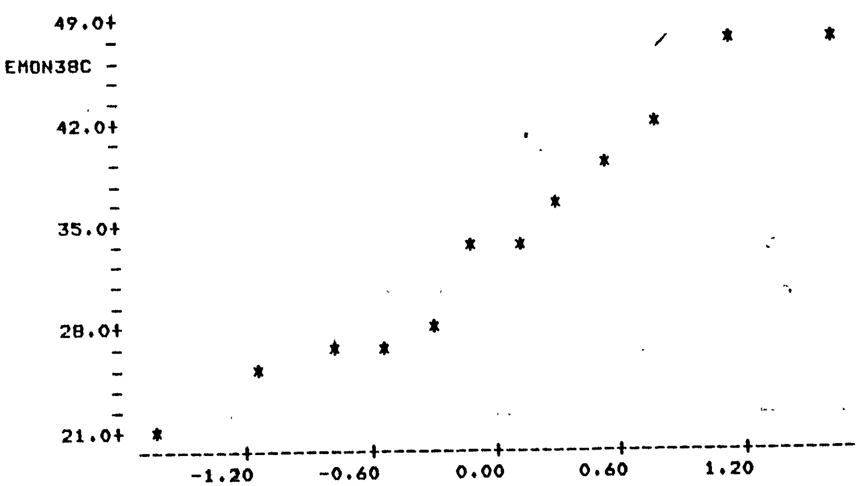
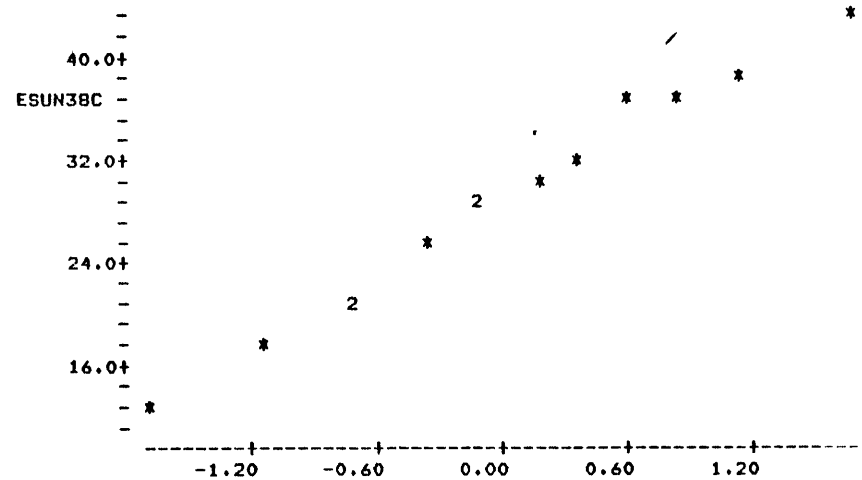


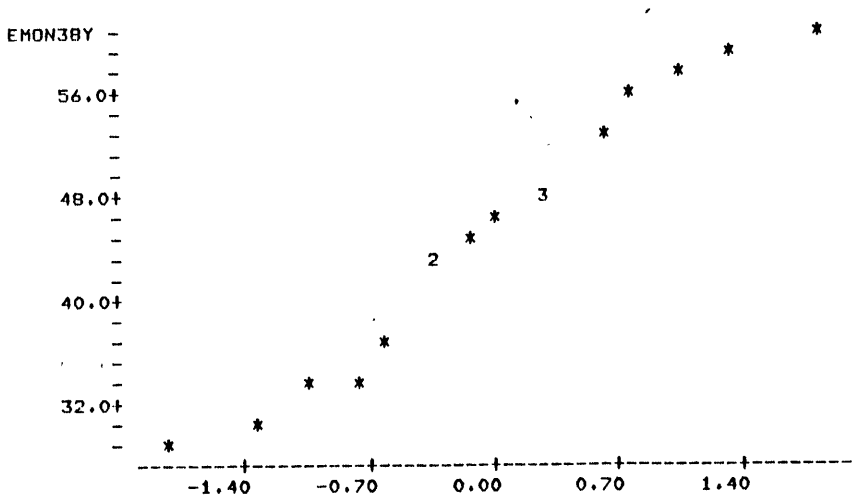
Figure 12: NORMAL PROBABILITY PLOTS - FLIGHT 37 PAL/LAX AUG-DEC 1983



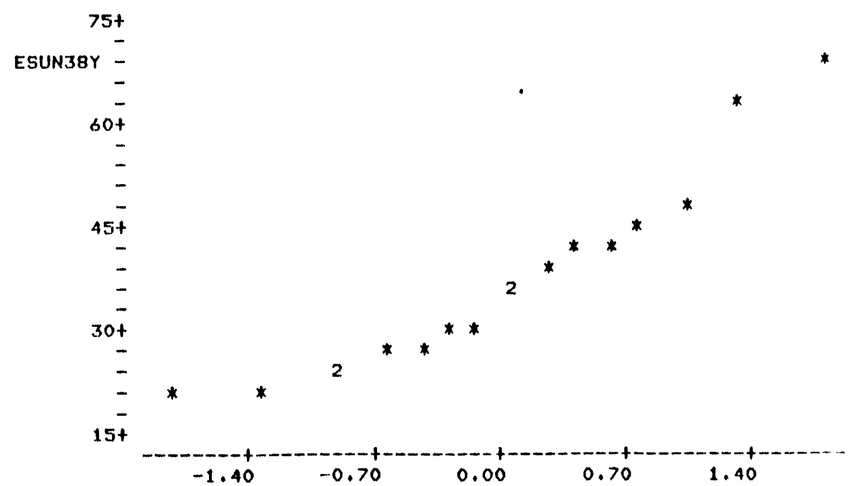
AMBASSADOR CLASS - Mondays



AMBASSADOR CLASS - Sundays

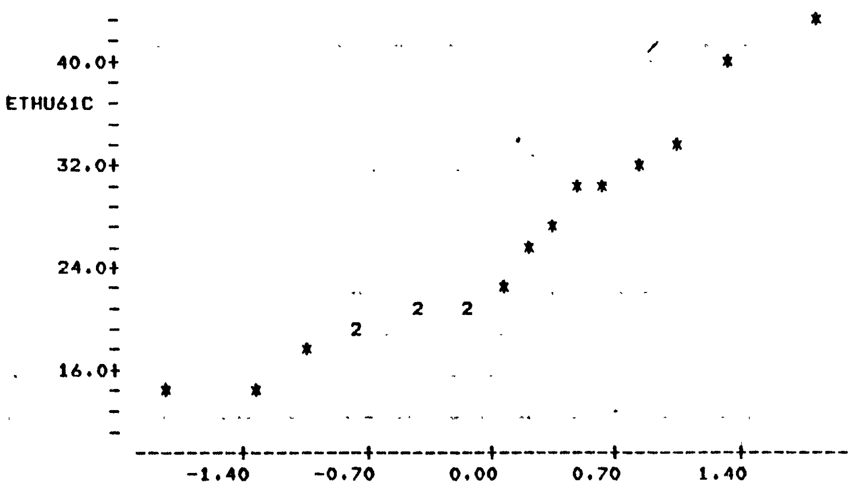


COACH (Y) CLASS - Mondays

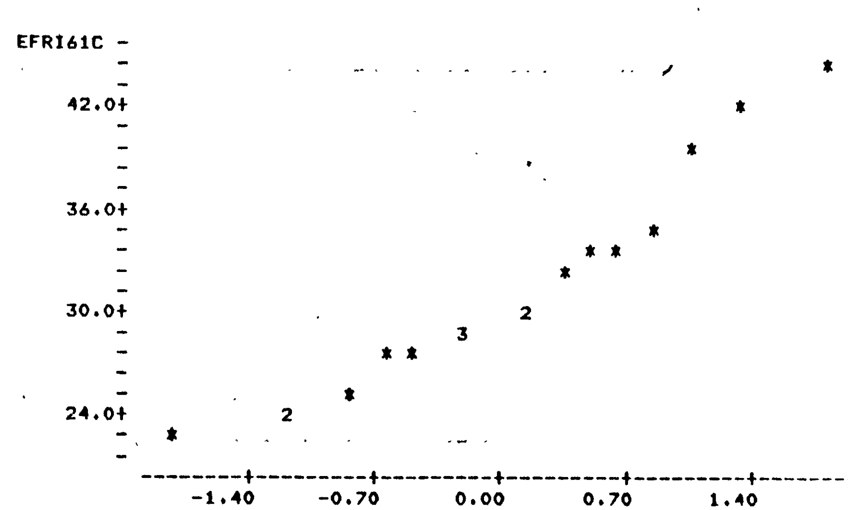


COACH (Y) CLASS - Sundays

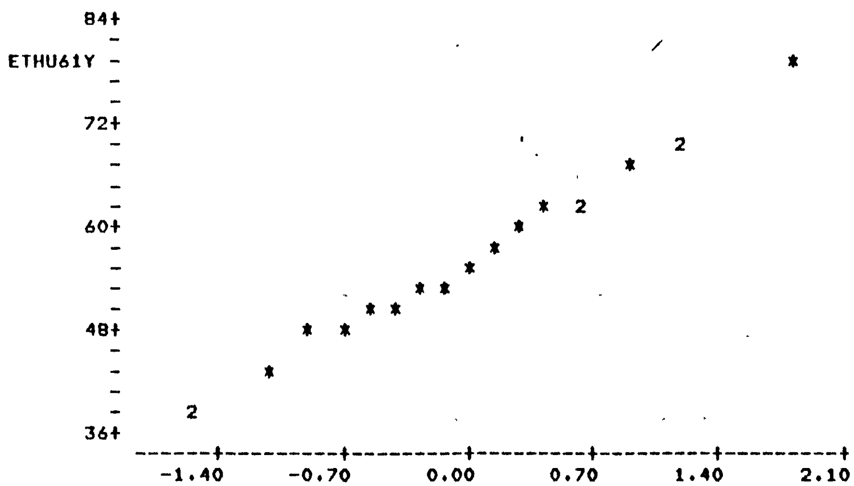
Figure 13: NORMAL PROBABILITY PLOTS — FLIGHT 38 LAX/PHL AUG-DEC 1983



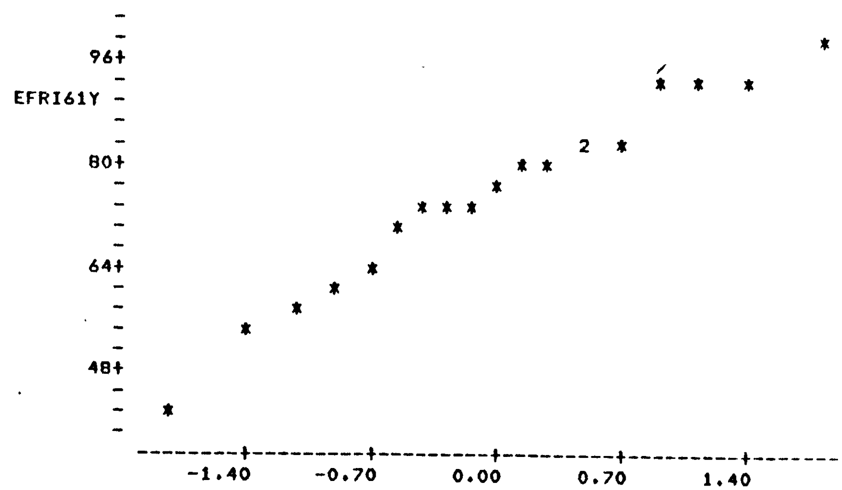
AMBASSADOR CLASS - Thursdays



AMBASSADOR CLASS - Fridays



COACH (Y) CLASS - Thursdays



COACH (Y) CLASS - Fridays

Figure 14: NORMAL PROBABILITY PLOTS — FLIGHT 61 BOS/SFO AUG-DEC 1983

the probability plot curves away from a straight line above the mean. The Sunday Ambassador Class subset, on the other hand, demonstrates a pattern of negative skewness, due likely to high demand and a limited capacity, as discussed previously.

The plots from Flight 38 (Figure 13) depict cases in which the fit of the data to the Gaussian model differed between Ambassador and Coach classes for the same days of the week. The Monday Coach class data fit the model reasonably well, whereas the Ambassador class data showed a far more pronounced S-shape, which suggests a more uniform distribution than the Gaussian model. Conversely, the Coach class data for the Sunday subset showed substantial skewness in comparison to the fit of the Ambassador Class data.

Figure 14 (Flight 61) shows a similar variety of fits with the Normal distribution. Worth noting is the pronounced positive skewness for Ambassador class bookings on Thursday and Fridays, while Coach class bookings for the same days fit the model more closely.

Examination of the Normal probability plots for all of the data subsets did not reveal an overall consistent pattern among the subsets. It is possible, however, to postulate a relationship between data distributions and the combination of level of demand and aircraft capacity by day of week. There is some indication that significant departures from a Gaussian distribution of reservations is more likely to occur on days of extreme (high or low) demand in a particular fare class. On days of the week with very low demand, positive skewing of the reservations distribution seems to occur. On days of extremely high demand, negatively skewed distributions appear in the fare classes for which average demand approaches seating capacity. And, in the Coach fare category, where capacity limits were seldom threatened for the flights examined, high demand days of the week produced positively skewed distributions. In general, the remainder of the cases seemed to fit the Normal model reasonably well.

Application of more formal goodness-of-fit tests to the data samples did not clarify these patterns of fit. It was not possible to reject the assumption of normality at an acceptable level of significance for the majority of samples because of the small sample sizes. In the cases of the most extreme deviations indicated by the normal probability plots, it was possible to reject the assumption of a Gaussian distribution of the data at well above a 90 % level of significance. However, there were numerous other samples for which the normal probability plots showed substantial deviation from the model, but which were not large enough for valid application of goodness-of-fit tests.

It was therefore necessary to consider ways of combining some of these smallest data subsets into larger, yet still homogeneous groups. While the smallest sample subsets clearly indicated departures from the Gaussian model of reservations distributions under certain conditions, larger samples were required so that some statistically valid assessment could be made. This process of combining subsets and analyzing the resulting distributions is discussed below.

(c) Assessment of Combined Reservations Distributions

The final step in the exploratory data analysis performed on the reservations data subsets involved combining two or more of the individual edited subsets with similar characteristics and then once again assessing the patterns exhibited by the resulting distributions. The process by which the subsets were combined and, more importantly, the results of the subsequent distributional analysis are described in this section.

In brief, the individual data subsets from each flight and within each fare class were combined according to similarities in their sample distributions and parameters. Comparisions were made with respect to the sample means, variances, k-factors and extreme points, and were reinforced by further comparisons of both the boxplots and Normal probability plots of the individual subsets. Two or more of these subsets were grouped together for this distributional analysis when all or most of these comparisons showed little difference between the subsets.

Table 9 provides some examples of these comparisions and lists the combined data samples that were ultimately examined for distributional patterns. In the majority of instances, the groupings stem from similarities in the distributions of reservations demand on peak days of the week as opposed to low demand days, as one would expect. Further, the patterns of directionality discussed previously are also reflected in the different groupings for the different flights. For example, Friday and Sunday Ambassador Class data are paired for Flight 37 (westbound), whereas Fridays and Mondays are the most similar peak days for Flight 38 (eastbound).

The distribution of each grouped data sample was examined with the help of boxplots and Normal probability plots, as before. Overall, the relationships between distribution shapes on the one hand and demand levels and capacity limits on the other were again evident in this assessment of the combined distributions. That is, the grouped data for midweek low demand periods as a rule showed a much better fit to the Gaussian model than the grouped data for higher demand days of the week. Examples of plots that reinforce this notion of a relationship between distribution shape and mean reservations levels are provided in Figures 15 and 16, which include plots of grouped data from Flights 37 and 38, respectively.

In Figure 15, the differences between the grouped subsets are apparent for the pairs of plots shown for both Ambassador Class and Coach Class reservations. The Wednesday/Thursday Ambassador Class plot fits the Guassian model well, while the higher demand Friday/Sunday plot clearly does not. A formal goodness-of-fit

Table 9 - CHARACTERISTICS OF DATA SUBSETS COMBINED FOR DISTRIBUTIONAL ANALYSIS (August-December 1983)

FLIGHT	CLASS	DAYS COMBINED	$\overline{X}$	High	Low	K
037	Ambassador	Wednesday	28	44	15	.30
		Thursday	30	44	16	.26
		Friday	38	51	24	.21
		Sunday	36	47	25	.23
	Coach	Monday	28	41	17	.24
		Tuesday	30	48	13	.29
		Wednesday	39	51	26	.19
		Sunday	37	51	26	.24
038	Ambassador	Tuesday	26	36	19	.20
		Wednesday	29	42	20	.20
		Monday	34	48	21	.26
		Friday	36	50	22	.26
	Coach	Tuesday	29	46	16	.31
		Wednesday	33	58	20	.27
		Thursday	37	53	27	.22
		Sunday	37	70	20	.38

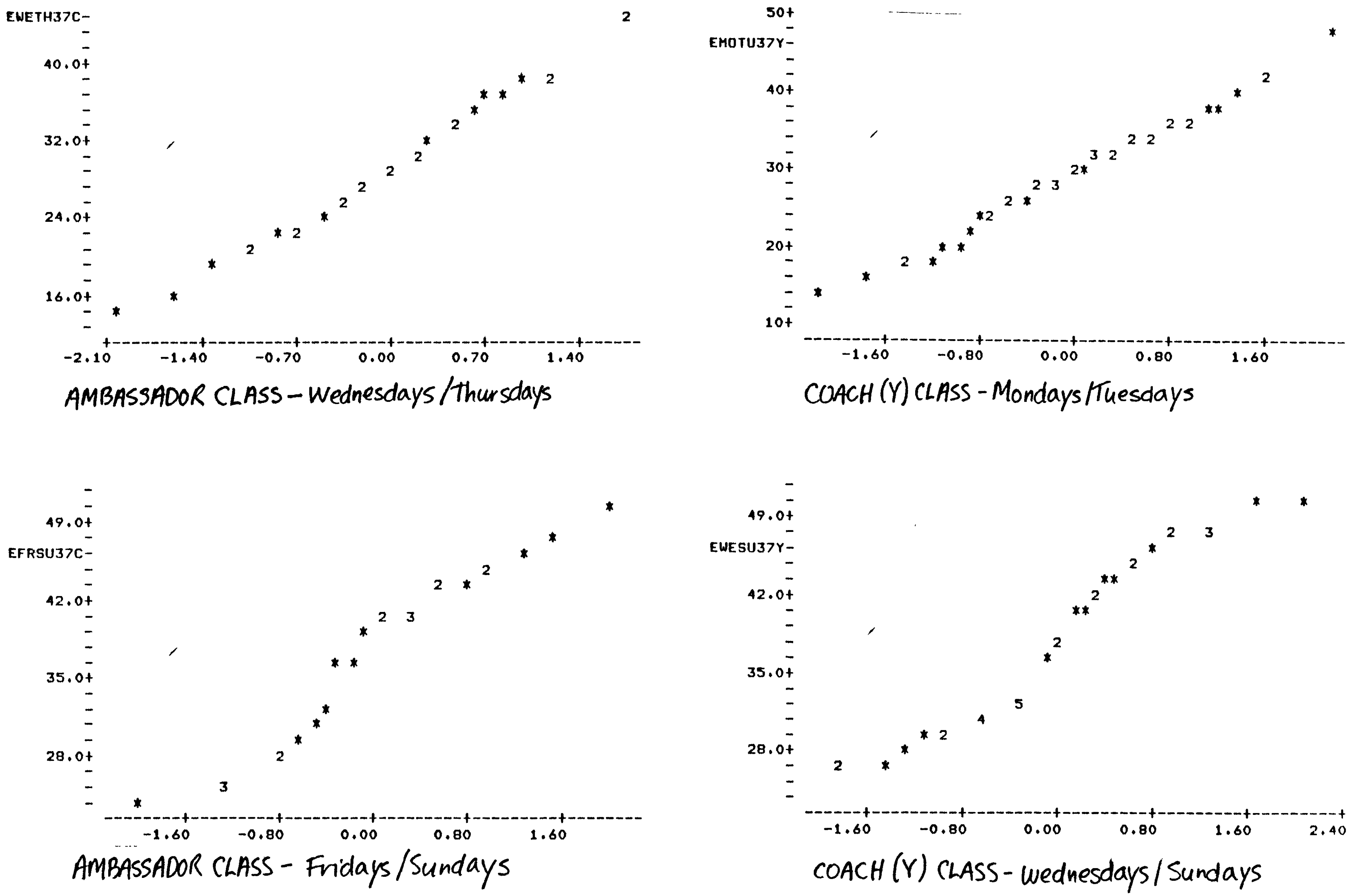
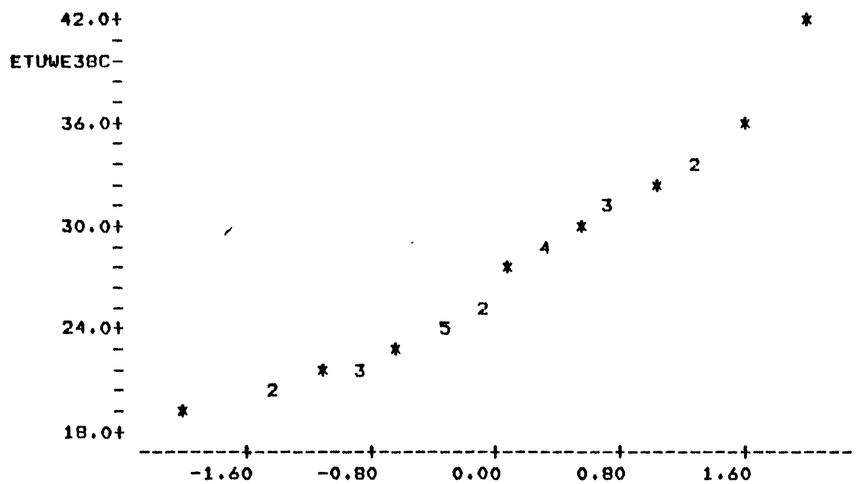
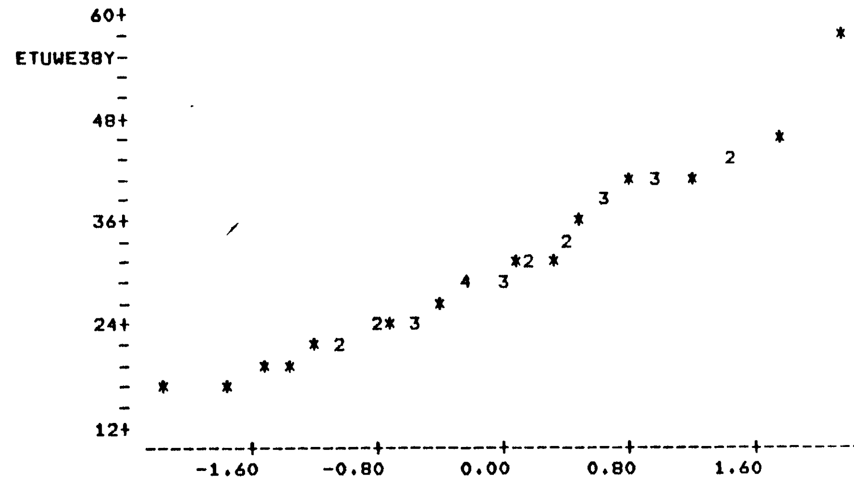


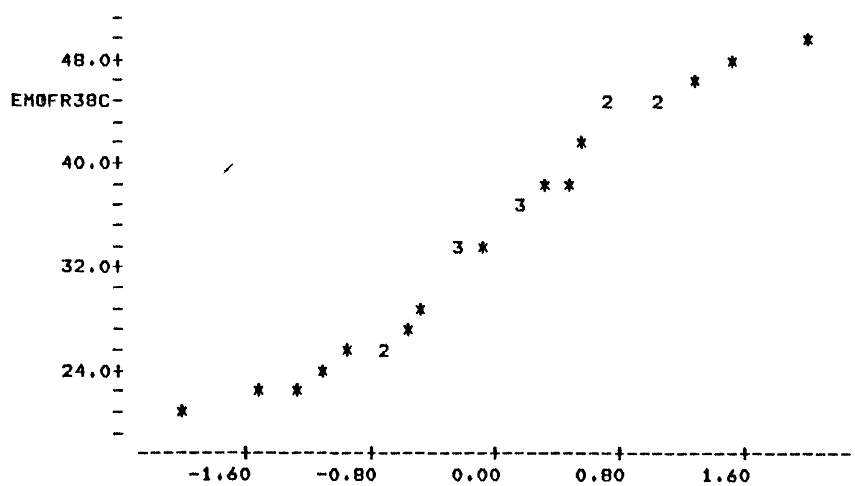
Figure 15: NORMAL PROBABILITY PLOTS — FLIGHT 37 GROUPED SAMPLES



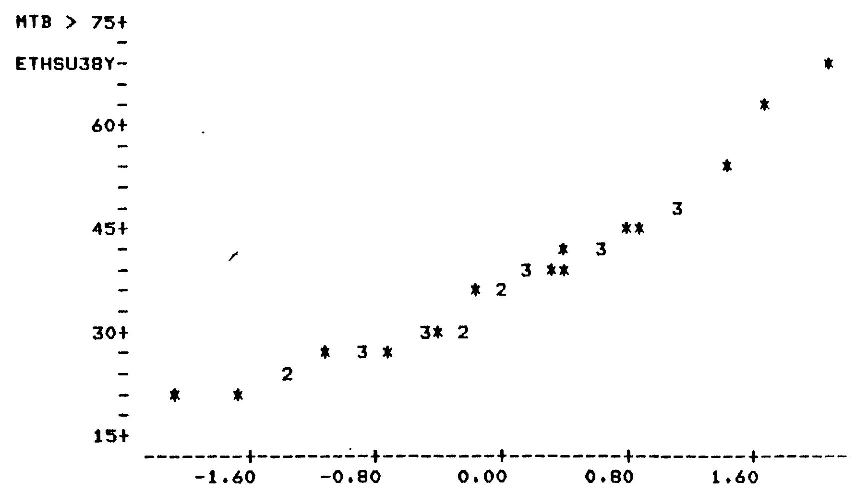
AMBASSADOR CLASS - Tuesdays/Wednesdays



COACH (Y) CLASS - Tuesdays/Wednesdays



AMBASSADOR CLASS - Mondays/Fridays



COACH (Y) CLASS - Thursdays/Sundays

Figure 16: NORMAL PROBABILITY PLOTS — FLIGHT 38 GROUPED SAMPLES

test on the latter grouped distribution confirms that we can reject the assumption of normality at greater than a 90 % level of significance. The two Coach Class plots in Figure 15 show a similar pattern, although the Wednesday/ Sunday combination represents a more moderate peaking of demand. Nonetheless, the assumption of normality in this case can be rejected statistically at a 95 % level of significance.

The plots in Figure 16 for the two groups of Ambassador Class data show virtually identical results to those of Flight 37, except that the peak demand grouping for the eastbound Flight 38 includes Mondays and Fridays, as discussed. The Gaussian model for the grouped reservations distributions on these peak demand days of the week can be rejected at a 95 % level of significance. The Flight 38 Coach Class plots included in Figure 16 were selected to refute the notion that our postulated relationship between demand levels and reservations distributions holds true in all cases. For this flight, the lower demand group (Tuesday/Wednesday) deviated from the Gaussian distribution to a greater extent than the higher demand (Thursday/Sunday) group, although both plots show signs of positive skewness.

The effects of combining the edited day-of-week data subsets were twofold. On the one hand, the larger sample sizes permitted formal goodness-of-fit tests to be applied with greater confidence, and allowed us to reject the Gaussian model in many of the extreme demand cases at significance levels of approximately 90 - 95 %. On the other hand, because none of the original subset distributions were identical, any combination of subsets inevitably obscured some of the within-sample variation. As a result, the normal probability plots, while showing smoother curves, were more difficult to interpret.

VI. CONCLUSIONS AND FURTHER RESEARCH

Our research thus far has involved exploratory data analysis of the 1983 reservations data for selected TWA transcontinental flights. The analysis has focussed exclusively on day-of-departure reservations total by fare category, with the objective of identifying differences in reservations patterns among the various fare categories that might be of relevance to capacity management practices. An examination of the systematic patterns in the data provided the basis for disaggregating the dataset into homogeneous subsets for a more detailed analysis of the distributional characteristics of reservations by flight and fare category.

The results of the exploratory phase of the analysis described in this paper confirmed many of the demand patterns found in our initial analysis of only two transcontinental flights. For the aggregate reservations data, significant differences among fare categories in both the pattern of reservations by day-of-week and the degree of week-to-week variation were evident. Comparison of reservations data for flights operated in similar markets over the same period suggested the following additional characteristics:

(1) Day-of-week reservations patterns appear to be similar for same-direction flights in similar markets, in spite of absolute differences in mean reservations levels;

(2) As expected, relative variation (k-factors) of reservations tends to decrease as sample means increase;

(3) Reservations data for capacity-controlled fare classes must be interpreted carefully, particularly when capacity controls were likely to have been applied.

A comparison of data from pairs of eastbound and westbound flights operated in the same city-pair markets prompted further observations:

(4) Substantial directionality in the day-of-week variation in reservations levels exists for the various fare classes of the transcontinental flights examined;

(5) This directionality can also differ significantly among fare classes;

(6) The coefficients of variation (k-factors) for reservations on a particular day of the week in the individual fare categories were substantially higher than the values traditionally assumed for total flight demand.

Plotting the distributions of reservations levels for the aggregate data groups examined in the exploratory phase of the analysis continued to bring the assumption of Normally distributed reservations/demand by fare category into question. Furthermore, the distribution shapes differed noticeably among days of the week, both within each fare class and between fare categories. The aggregate distributions suggested a relationship between distribution shape, particularly degree of skewness, and the mean reservations level for the data sample under consideration.

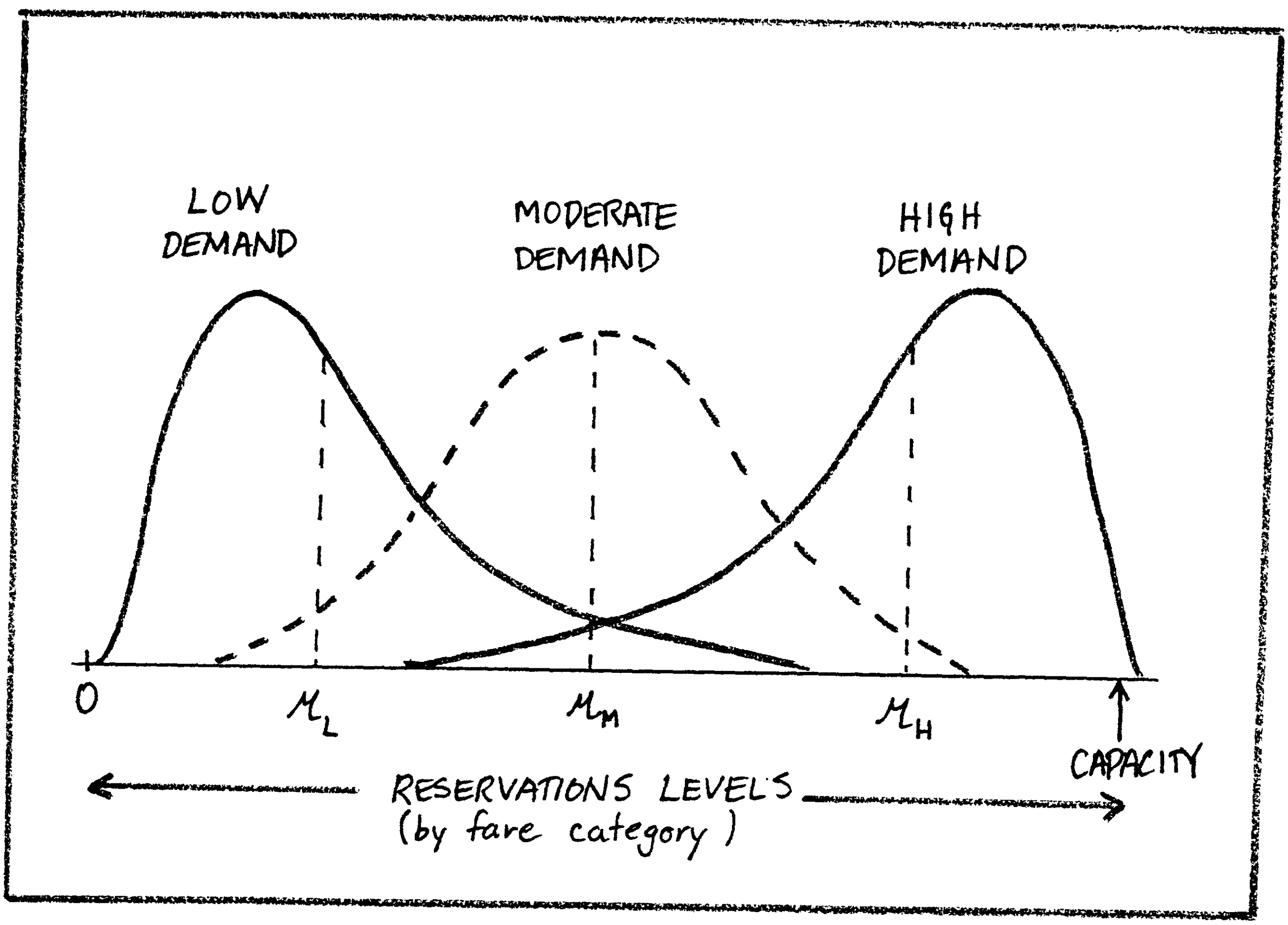
While the bulk of the aggregate distributions showed positive skewness, some of this skewness could be dismissed as resulting from aggregation of data with substantial systematic variation. To permit assessment of the distributional characteristics of reservations with much of this systematic variation removed, the dataset was disaggregated and outliers were edited, creating much smaller homogeneous data subsets. More detailed evaluations of the distribution of reservations data within each of these subsets were then performed.

The pattern of positive skewness for these distributions was not consistent, as many of the edited subsets conformed reasonably well to the Gaussian model, or at least did not deviate enough from the model to permit statistical rejection of the normality assumptions. However, the subsets that deviated most from the Normal model were generally those for days of the week with extreme (high or low) mean reservations levels, reinforcing the notion of a relationship between distribution shapes and demand.

Finally, data subsets with similar characteristics were re-aggregated to create samples with a larger number of data points. The larger sample sizes allowed goodness-of-fit tests to be applied to the sample distributions. The relationship between distribution shapes and the combined factors of reservations levels and fare class capacity appeared to hold in the majority of cases examined. Reservations data for low-to-moderate demand days of the week generally fit the Normal model well. On the other hand, the sample distributions of data from higher demand days showed significant departures from the Gaussian model, to the point that the assumption of normality could be formally rejected at significance levels of 90% or more.

The results of this analysis suggest a potential model of the relationship between reservations distributions for a flight and fare category on particular days of the week, and the historical demand levels on those days. As illustrated in Figure 17, we can hypothesize that, as the locations of reservations distributions shift in a positive direction, their shapes change as well. For cases in which extremely low levels of demand relative to fare class capacity

Figure 17: CONCEPTUAL MODEL OF DISTRIBUTION PATTERNS



occur, significant positive skewness of the reservations distribution might be expected. The assumption of Normally distributed reservations data seems to be most valid for moderate levels of demand, where capacity limits are generally not reached. And, at the upper end of the demand scale, it seems that capacity limits might induce negative skewness in the distribution.

This type of movement in reservations distributions for particular flights clearly would have implications for capacity management. However, before such implications are considered, further empirical analysis is required to confirm or reject this hypothesis and to better define the relationships. Furthermore, the degree to which such distribution behavior can be applied in general and the extent to which this behavior might differ among fare classes should be examined. Continuation of this aspect of our research into the capacity management problem is planned for 1985.

Another, somewhat different, analysis of the available TWA reservations data is also about to commence. Given 35-day booking histories for each of the flights examined, we are interested in the cumulative reservations process for a flight. An examination of booking trends by fare class and under differing conditions of ticket purchase and/or ticket price will be undertaken to identify characteristics of relevance to capacity management. Conclusions about reservations behavior, together with conclusions about reservations distributions, will ultimately be used to develop descriptive models of air transportation demand that will provide practical input into the capacity management process.