

Bus Passenger Waiting Times in Huddersfield

LOCAL GOVERNMENT OPERATIONAL RESEARCH UNIT
Royal Institute of Public Administration
Report No. T 67

1977

LOCAL GOVERNMENT OPERATIONAL RESEARCH UNIT

Royal Institute of Public Administration

BUS PASSENGER WAITING TIMES
IN HUDDERSFIELD

A report of part of the
Huddersfield Bus Study

A.J.Daly
S.Zachary

PREFACE

The work described in this report formed part of the Huddersfield Urban Bus Study. This study was carried out by the Local Government Operational Research Unit on behalf of the West Yorkshire Metropolitan County Council and Passenger Transport Executive.

The main recommendations and findings of the Study are reported separately in other LGORU reports.* This report deals solely with the collection and analysis of data relating to the length of time intending passengers spend waiting at bus stops.

* These reports are summarized by Report C.229 *Huddersfield Bus Study – Summary Report*, which also lists the detailed reports of the study.

C O N T E N T S

	Page
1. INTRODUCTION	1
1.1 The Need for a Survey	1
1.2 The Scope of the Survey	2
2. COLLECTION AND VALIDATION OF DATA	5
2.1 Selection of Sites and Times for Survey	5
2.2 Recording the Data	6
2.3 Validation	9
3. ANALYSIS OF SURVEY RESULTS	11
3.1 The Pattern of Bus Movements	12
3.2 Passenger Waiting Time Statistics	12
3.3 Discussion of the Results	14
4. DEVELOPMENT OF PREDICTIVE MODELS	17
5. CONCLUSIONS	21
APPENDICES:	
1 SURVEY FORM AND SURVEYORS' INSTRUCTIONS	23
2 EFFECTS OF PRECISION IN RECORDINGS	27
3 STANDARD ERROR OF ESTIMATE OF MEAN WAITING TIME	31

1. INTRODUCTION

This report describes the collection and analysis of data on bus passenger waiting times in Huddersfield. The work was done as part of the comprehensive study of urban bus operations in the town, and was carried out by LGORU in 1974 and 1975.

1.1 The Need for a Survey

A major part of the Bus Study was concerned with assessing the impact on bus passengers of changes in the bus services offered to them. This assessment called for objective measures of the quality of service, both to permit estimates of any gains or losses in patronage, and to evaluate the overall benefits or disbenefits occasioned by the changes.

The measures of service adopted for the study were of the conventional 'generalized cost' type, which include in a linear function a number of the measurable parameters of the bus service. Among these parameters is the average time spent waiting for buses in making a given journey.

In a conventional generalized cost function as used in this study, and as recommended by the Department of the Environment, the waiting time is the sole representative of the frequency of the service. This representation is clearly valid only as an approximation, as an increase in frequency will generally not only reduce the average waiting time, but also increase the choice of times of travel, thus giving additional benefit. Little is known, however, about benefits derived by travellers from these savings in 'scheduled delay'; although one would presume that the value of time spent at ultimate origin and destination would be much less than the value of time spent waiting at bus stops. Accordingly, we believe that the use of waiting time as the sole representative of frequency is reasonable in the present state of knowledge.

Waiting time may be seen as the sum of two components: the time spent at the stop until the first acceptable bus arrives, and any further time spent at the stop when the passenger has been unable to board the first bus. The latter component is entirely beyond the control of the passenger, as it is determined by the capacities of the buses and the numbers of people who have boarded at previous stops. This latter component was estimated in the Huddersfield study by part of the TRANSEPT model.* The former component, however, is under the control of the passenger to the extent that he knows the times of buses and is prepared to rely on their time-keeping.

Clearly, the reliability of bus times will vary with local conditions from town to town, and furthermore the public perception of that reliability will also depend on local factors. Thus it was decided for the Huddersfield study to conduct a local survey of waiting times to determine the extent to which passengers could take advantage of their knowledge of bus timings and to investigate what factors influenced the times they actually spent waiting.

In designing and carrying out the survey LGORU drew on their experience gained in a previous survey during the Coventry Bus Study.**

* TRANSEPT - A Bus Model, A.Last, S.E.Leak, *Traffic Engineering and Control*, January 1976. Also available as LGORU Transportation Working Note 7.

** *Planning Urban Bus Routes*, A.J.Daly et al. LGORU Report C.149, April 1973.

1.2 The Scope of the Survey

For the Huddersfield Study, it was necessary to estimate the average waiting times likely to be experienced over a fairly wide range of circumstances. The survey on which the projections were to be based had, therefore, to cover an equally wide range of circumstances to ensure reliable estimates.

Accordingly, the survey was designed to investigate the effect of a number of factors thought to be relevant to average waiting times. These factors were:

- the headway, i.e. the interval between one bus and the next
- the reliability of the service
- the time of day.

Other factors liable to influence the distribution of waiting times, and thus complicate the assessment of those above, are:

- the particular service and bus stop — some services may be better known or have more regular users than others; in addition, the presence or absence of a shelter at any stop may influence the time at which an intending passenger aims to arrive. Also, a person travelling from home may have more opportunity to time his arrival at the bus-stop than he has on his return journey. As most of the former journeys are inbound, i.e. towards the town centre, we might expect shorter waiting times at the corresponding stops than at those associated with outbound journeys.
- the particular day — weather conditions may be important; also passengers may use Monday to gauge the pattern of bus arrivals for the rest of the week. Their subsequent waiting times would then tend to be reduced.

In order to minimize the influence of these and other complications, and to assess fairly the first three factors, it was necessary to carry out a fairly extensive survey over a period of several weeks, covering a range of stops and headways, and different periods of the day.

Two things affect people's perception of the reliability of a bus service, and must be catered for in the design of a waiting times survey. The first is when a bus does not arrive at the expected time, in particular when its time of arrival varies from day to day. If the bus is constantly the same number of minutes behind schedule, passengers may well adjust their expectation of the arrival time. But a variable arrival time inevitably lengthens waiting times, for a passenger who does not aim to arrive at the earliest likely time is liable on occasions to find himself waiting for the following bus instead. The second indicator of unreliability is when a bus fails to arrive at all. The observations of waiting times in this case are more difficult to analyse for use in prediction: it is difficult to know how many of the intending passengers knew of the cancellation in advance; and it is virtually impossible to predict the future frequencies of such occurrences. Fortunately, in Huddersfield, buses are easily identifiable with the times stated in the timetable, and certainly at the time of the survey it was rare for a scheduled bus to fail to run.

Analysis and prediction are also complicated when buses are overcrowded and passengers are unable to board the first otherwise suitable bus. Again this situation very rarely arose in the course of the survey, except during the evening peak. However, it was necessary to consider this effect in the design of the survey.

* * * * *

The following chapter gives details of the survey that was carried out in Huddersfield. Chapter 3 presents the results obtained from the survey, and Chapter 4 describes the predictive models that were developed.

2. COLLECTION AND VALIDATION OF DATA

This chapter describes the collection and validation of the data on which the subsequent analysis was based.

Particular care was necessary in the selection of appropriate bus stops for survey to give the range of circumstances necessary for the study. The considerations involved are described in Section 2.1. We were most grateful for the advice of West Yorkshire PTE staff in the selection of appropriate stops for survey.

The method used for the collection of data and the problems arising in the design of an appropriate form are detailed in Section 2.2. The chapter concludes with a description of the validation processes used to remove errors from the data.

2.1 Selection of Sites and Times for Survey

The survey took place during three weeks in November, 1974, and was carried out by employees of the West Yorkshire PTE temporarily diverted from other work.

Four periods of the day were covered:

- | | | |
|----|--------------------|---------------------|
| 1. | morning peak | (from 0700 to 0900) |
| 2. | morning off-peak | (from 1000 to 1130) |
| 3. | afternoon off-peak | (from 1330 to 1530) |
| 4. | evening peak | (from 1600 to 1730) |

For each period a number (between six and nine) of stops were selected and observations made on at least seven different days. An effort was made to spread observations over different days of the week, but observations were not made on Saturdays or Sundays.

Each stop selected had to satisfy certain criteria.

1. **It had to be 'single-service'**, i.e. all buses served by the stop had to run along the same route thereafter; so that any person waiting at the stop could catch any bus. This restriction was not just a matter of keeping things simple, but was necessary in order to be sure of the real service headway associated with each person waiting at the stop.
2. **The stop had to be used by a substantial number of people**; otherwise the collection of sufficient data to provide reliable information would take too long.
3. **The stop had to be located where it was possible to identify the arrival time of intending passengers**; if, for example, there were a shop doorway nearby from which people emerged on the approach of a bus, satisfactory observations could not be made.
4. **There had to be reasonable shelter for the surveyor**; this was not usually a problem as stops satisfying criteria (2) and (3) were normally provided with shelters.

In addition, the stops selected had to cover a sufficient range of headways to make prediction in likely future situations possible by interpolation. Equally, it had to be possible

to determine and calibrate reliably a model relating waiting times to headway (and other factors).

These conditions, especially (1), severely restricted the number of suitable stops, and it would in fact have been difficult usefully to increase the number included in the survey. Table 1a lists the stops selected and Table 1b gives the dates and periods for which they were surveyed.

Table 1a: Bus Stops Surveyed

Bus Stop No.	Routes	Bus Stop Location	HUBS No.
1	6, 7	Wellhouse School	926
2	8, 9	Britannia Road, bridge, Slaithwaite	624
3	17	Masons Arms, Lowerhouses	337
4	42, 43	Elland Town Hall	012
5	50, 51, 52, 53	New Street, Cowlersley	485
6	63	Shepherds Arms, Cowcliffe	868
7	30	Southfield Road, Almondbury	334
8	70	Rawthorpe Lane, top of Ridgeway	342
9	71	Long Lane, bottom of Ridgeway	364
10*	64	Lord Street	111
11*	16	Byram Street	110
12*	42, 43	John William Street	105
13*	17	Byram Street	110
14	85, 86	Beaumont Park (Drycloth Road/Woodside Road)	464

The HUBS number is the number used to describe this stop in the main Bus Study.

* Outbound stops (all in town centre) – others are inbound.

2.2 Recording the Data

The following information was recorded at each stop:

- arrival and departure time of each bus
- arrival times of passengers
- number of passengers (if any) unable to board each bus.

Each bus was also identified with its scheduled departure time from the nearest previous timing point (normally the stop itself).

All times were recorded to an accuracy of one minute, more precise recording being impractical. Although some uncertainty is thus introduced about the time waited by

Table 1b: Dates of Survey at Each Stop (all dates are in November 1974)

Bus Stop No.	Period 1		Period 2		Period 3		Period 4	
	Scheduled mean headway	Dates	Scheduled mean headway	Dates	Scheduled mean headway	Dates	Scheduled mean headway	Dates
1	28 min	19,20,21 22,25,26 27,29						
2	30 min	12,18,20 25,26,27 28,29						
3	12 min	11,12,13 14,15,19 20,21,26 27,28	30 min	11,12,13 14,15,19 20,21,26 27,28	15 min	11,12,13 14,15,19 20,21,26 28		
4	15 min	15,19,20 22,25,26 28,29	15 min	11,12,13 14,15,18 19,20,25 26,29	15 min	13,15,18 19,20,21 27		
5	8 min	11,12,13 14,15,18 19,20,21 22	20 min	11,12,13 14,15,18 19,20,21 22			11 min	11,12,13 14,15,18 19,20,21 22
6	30 min	11,12,13 14,15,18 20,25,27						
7	7½ min	13,15,18 19,20,21 22,27	15 min	19,21,22, 25,26,27 29	15 min	19,20,21 22,25,26 27,29		
8	20 min	11,12,13 14,15,19 21,25,26 28,29						
9	10 min	11,12,13 14,15,18 19,20,25 26,29						
10			15 min	11,12,13 14,15,18 20,25,27	15 min	11,12,13 14,15,18 20,25,27	15 min	11,12,13 14,15,18 20,25,27
11			30 min	12,18,20 22,25,26 27,28	30 min	12,18,20 22,25,26 27,28,29	15 min	12,18,20 22,25,26 27,28,29
12					15 min	15,19,20 22,25,26 27,28,29	15 min	11,12,13 14,15,18 19,20,25 26
13			30 min	11,12,13 14,15,21 25,28,29	15 min	11,12,14 15,21,25 28,29	13 min	13,15,18 19,20,21 27
14			30 min	11,12,13 14,15,19 21,25,26 28,29	30 min	11,12,13 14,15,19 21,25,26 28,29	11 min	11,12,14 21,25,29

Dates in bold type indicate that usable data was collected.

any individual, the effect on statistics such as mean waiting times is of the order of a second or so, and quite negligible compared with other sources of uncertainty (see Appendix 2).

The departure times of buses were simply recorded by the clock, rounding to the minute below. Rather than noting the arrival time of each passenger, it was more practical to record the number of passengers arriving at the stop in each minute. Two approaches were considered.

1. **A continuous minute by minute record throughout the entire observation period:** noting the number of passengers arriving in each minute; noting those minutes in which buses arrive and depart; and noting exactly how many passengers in a 'departure' minute actually arrive before the bus leaves.
2. **A separate record for the waiting interval associated with each bus,** in which a stopwatch is zeroed at the start of the interval (the time of departure of the previous bus), and the number of passengers arriving in each minute is noted. The record ceases when the bus departs, this time is noted (by the clock), and a new record begins. The arrival time of the bus is also recorded.

In either case the scheduled departure time of the bus (as defined above) is also noted, as well as the number of any remaining passengers.

On theoretical grounds there is little to choose between the two approaches – the second involves a slight information loss, but provides a valuable accuracy check in that the number of minutes with passenger arrivals recorded should be equal to or one greater than the difference between the departure time of the bus and the departure time of the previous bus. The practicalities of each approach were discussed with the PTE, and it was agreed that it would be easier to use the second one.

Appendix 1 includes a specimen copy of the form used to record the information associated with each bus, and also a copy of the notes for the guidance of surveyors. Each form also recorded the start of the waiting interval which should in general have corresponded with the departure time of the previous bus. Particular consideration was given to the method of recording passenger arrivals, so as to minimize the chance of counting any arrival in the wrong minute. In consequence the number of arrivals in each minute occurs twice on any form, once as originally recorded, and once as an actual number, suitable for transferring onto a punched card. This provides an additional check.

Provision was not originally made for recording the arrival time of each bus, but only its departure time, which with a one-man-operated bus could be up to a couple of minutes later. Therefore, the arrival time of the bus was recorded in the same way as passenger arrivals, i.e. an indication was made of the corresponding minute on the stopwatch. Consequently a bus would be recorded as arriving in, say, minute 25, and departing at, say, 0833. In the event this caused no difficulty, except that some surveyors forgot to record arrival times at all. Departure times were, however, used in the definition of the passenger waiting times (see Section 3.2).

2.3 Validation

Validation of the collected information was carried out in two stages:

1. All forms were examined manually to detect, and generally rectify, any obvious errors. For example, the 'start time' (departure time of the previous bus) might erroneously have been entered in the space for 'departure time', or a surveyor might have forgotten to total the number of passenger arrivals in the last minute (that of the departure of the bus), or perhaps have entered zero in the total for the following minute (which must be left blank so as to identify the departure minute of the bus). Such errors only occurred occasionally, but a manual inspection of all the forms was necessary.
2. All the recorded information was then transferred onto punched computer cards. From these a specially written computer program gave a 'graphical' printout of the data on a time scale, continuous over each observation period, against which was recorded:
 - passenger arrivals in each minute, counting forward from the 'start time' of the waiting interval for the appropriate bus, and identified with that bus;
 - bus arrival and departure times.Indication was also given of scheduled departure times, and whether any passengers were unable to board the bus, together with the identifying information of date, bus stop, and surveyor, so that the printout displayed all the information collected.

Presentation of the data was by day for each bus stop within each period.

The program thus displayed the entire pattern of passenger and bus arrivals associated with each bus stop, and enabled checks to be made for any possible errors of a subtler nature, for example:

- incorrect recording of the departure time of a bus (with the start time of the succeeding interval also wrong); this would be indicated if, for example, there were too few waiting minutes associated with that bus, and too many associated with that following;
- incorrect watch (recording actual time), which would be suggested if the pattern associated with that day was obviously displaced with respect to the other days at the same stop;
- the possible intrusion of any other bus (e.g. a school bus) on the route, which would be suggested by an anomalous pattern of passenger arrivals.

The program also printed messages in the cases of the following simpler errors:

- 'start time' for an interval was not the same as the 'departure time' of the previous bus (to within ± 1 minute, though this tolerance was eventually removed before data analysis);
- 'start time' for any interval plus the number of the minute in which the bus departs (i.e. the last minute in which the number of passenger arrivals is recorded) was not equal to the 'departure time' of the bus (again to within ± 1 minute).

Most of the data was coherent, both within each observation period and between different days at the same stop, and so reliable for analysis. Just occasionally, however, this internal consistency was poor, and such data was earmarked in order to exclude it from the subsequent analysis. In effect, this meant omitting an entire day's observations for a particular period and bus stop.

3. ANALYSIS OF SURVEY RESULTS

This chapter describes the analysis that was carried out on the data and gives details of the results that were obtained.

The analysis was greatly facilitated by the use of a further computer program, written specifically for this purpose. The program was also designed to minimize the effect of any remaining errors in recording. For each bus stop and period of the day statistics are given on:

- the pattern of bus movements, within each day and summarized over the different days on which the stop was surveyed;
- the distribution of passenger waiting times within each day and summarized over all the days on which each stop was surveyed.

Options were made available to exclude data containing particular types of error for particular analytical purposes. There were two such options:

1. For each period of the day and bus stop it was possible to exclude the observations of any given day from the analysis, thus avoiding the need physically to remove the data.
2. For any bus that failed to arrive, it was possible to exclude or include the entire period associated with that bus and the following bus.

Option (1) was used to exclude from the analysis any day where the collected data was thought to be insufficiently reliable to add further information to that collected on other days at the same stop. As previously remarked, most of the data was reliable, and for this reason it was possible to omit the occasional dubious observation period.

The second option was necessary because, if an expected bus fails to arrive, then its intending passengers can only, for the purposes of analysis, be regarded as waiting for the following bus with consequently longer waiting times. It is impossible to say with certainty just which passengers were waiting for which bus. If one simply wishes to describe the distribution of waiting time that actually occurred, then this procedure is in any case correct. But if one wishes to use the distributions to make predictions for a hypothetical service in which buses never fail to run, then it is probably best to exclude altogether from analysis the entire waiting interval associated with both buses. This option was exercised in practice, as it was intended to use the results for prediction. Fortunately, a bus only very rarely failed to run*, so that little data was excluded in using this option, and only slight distortion occurred relative to the distribution of the waiting times actually experienced, and even this only for some stops.

As explained above, observations were only made with an accuracy within one minute. This, in itself, has a negligible effect on the results of the analysis. Appendix 2 discusses the issue in detail.

* Altogether thirty buses — approximately 2 per cent — failed to appear of those expected. Nine of these may, however, have been wrongly expected because of a misunderstanding of the schedule.

Sections 3.1 and 3.2 give the methods that were used to analyse the data. The main results are presented in a single table (Table 2) and discussed in Section 3.3.

3.1 The Pattern of Bus Movements

The analysis of bus movements concentrated on the time of departure of buses from each stop. Two key statistics were used:

1. **The headway**, defined for this purpose to be the difference between the departure time of each bus and the departure time of the previous bus, i.e. the headway associated with each specific bus was normally the maximum time any person would have had to wait for it. The analysis took the headway to be the recorded departure time of the bus less the recorded start time of the associated waiting interval (ensuring, where applicable, that this agreed with the recorded departure time of the previous bus). The program calculated the sample mean and standard deviation of the headway for each day, and also over all days for that stop. Table 2 presents the latter results.
2. **The 'lag'**, which was defined for each bus to be the difference between its actual and its scheduled departure time from the stop (or from the last earlier stop for which a scheduled time existed). The actual values of this lag were not of interest but rather their variations, both within each observation period and between different days. These variations provided a good measure of the predictability of the arrival time, and hence of the reliability of the service. Consequently, it was not of vital importance that the scheduled time related to the actual stop observed – although in general this was the case. Nor was the synchronization of the surveyor's watch critical. For each stop, the program calculated both the sample mean and standard deviation of the lag for each day, the overall mean for all days at the stop, and the overall sample standard deviation. The latter was calculated however, using only the variations *within* each day, thus avoiding any spurious variation resulting from different synchronizations of watches on different days. Table 2 gives the overall standard deviation for each period and stop.

A further statistic, called the *pause*, was used to measure the difference between bus arrival and departure times. If the bus was noted as arriving in the r th minute of the waiting interval (the r th minute on the stopwatch) and m was the departure minute (the last minute in which passenger arrivals were counted), then the pause p was defined by:

$$p = m - r + \frac{1}{4} \text{ (in minutes)}$$

The reason for the inclusion of the constant $\frac{1}{4}$ is given in Appendix 2.

Again, sample means and standard deviations were calculated, both for each day and over all days when the stop was observed (see Table 2).

3.2 Passenger Waiting Time Statistics

For each bus stop within each observation period, statistics were calculated describing the distribution of passenger waiting times, both for each day and over all days on which the stop was observed.

Table 2: Results

Stop	No. of days	No. of buses	No. of people	Headway Mean	Headway S.d.	Lag S.d.	Pause Mean	Passenger waiting time Mean	Passenger waiting time error*	S.d.
Period 1: Morning Peak (0700–0900)										
1	6	16	75	27.2	4.0	0.8	0.3	4.3	(0.7)	3.5
2	7	21	119	30.1	1.7	1.1	0.3	5.1	(0.6)	3.2
3	9	68	706	12.0	0.9	0.6	0.4	2.8	(0.2)	1.9
4	6	41	334	15.7	3.3	2.6	0.3	6.5	(0.2)	4.3
5	9	78	603	7.8	5.2	0.5	1.7	3.0	(0.2)	2.7
6	9	27	482	30.0	0.9	0.7	1.0	3.3	(0.1)	2.5
7	7	66	808	7.5	0.8	0.6	0.6	2.5	(0.1)	1.9
8	11	53	815	20.0	0.8	0.5	0.4	3.2	(0.3)	2.8
9	9	86	1215	10.0	1.1	0.8	0.5	3.1	(0.1)	2.2
Period 2: Morning Off-Peak (1000–1130)										
3	10	30	296	30.1	1.1	0.8	0.5	3.9	(0.4)	3.3
4	5	30	186	15.4	4.7	3.3	0.3	6.3	(0.3)	4.8
5	9	36	284	20.0	0.5	0.3	1.8	6.2	(0.4)	4.7
7	6	36	200	15.0	1.1	0.7	0.3	3.8	(0.3)	2.6
10	9	54	703	15.0	1.1	0.8	7.9	5.6	(0.2)	3.4
11	8	24	509	30.0	0.8	0.5	2.7	5.7	(0.2)	4.0
13	6	18	369	29.9	0.5	0.0	3.5	6.9	(0.4)	5.7
14	11	33	685	30.0	0.7	0.4	5.0	8.6	(0.4)	6.2
Period 3: Afternoon Off-Peak (1330–1530)										
3	10	76	346	15.0	0.7	0.6	0.4	2.7	(0.2)	2.5
4	7	62	420	15.0	4.3	2.8	0.4	6.4	(0.4)	4.4
7	6	54	330	15.0	0.7	0.6	0.4	3.9	(0.4)	3.1
10	9	70	1170	15.0	1.3	0.9	6.9	5.5	(0.2)	3.6
11	9	36	1379	29.9	1.5	1.1	3.7	7.3	(0.2)	5.0
12	8	62	1876	15.1	1.7	1.3	6.9	5.7	(0.1)	3.6
13	6	46	650	15.0	0.0	0.0	7.2	5.5	(0.2)	3.5
14	11	44	1156	29.6	1.8	1.4	4.7	9.2	(0.2)	6.4
Period 4: Evening Peak (1600–1730)										
5	9	79	290	11.3	6.9	0.8	1.9	5.0	(0.3)	4.4
10	8	47	2405	14.9	4.3	2.9	4.9	6.2	(0.2)	4.5
11	9	43	1812	14.9	9.8	0.5	4.7	6.8	(0.4)	5.9
12	8	15	963	15.7	3.0	2.5	4.9	6.7	(0.4)	4.5
13	7	49	1377	12.9	2.3	0.7	5.7	5.3	(0.1)	3.5
14	6	44	1150	11.6	5.3	1.5	4.8	5.6	(0.4)	4.1

* See Appendix 3 for calculation of this standard error.

The waiting time w of any individual passenger was defined in the same way as the bus pause time, i.e. a passenger arriving in minute r on the stopwatch, with the bus departing in minute m , had a waiting time of:

$$w = m - r + \frac{1}{4} \text{ (in minutes)}$$

Again Appendix 2 gives the reason for the inclusion of the constant $\frac{1}{4}$, as well as discussing the relation of w to the exact waiting time experienced by that passenger.

Three basic statistics were used to describe the distribution of waiting times over all passengers at each stop:

- the sample mean $\bar{w} = \frac{1}{n} \sum w_i$
- the sample standard deviation $s = \left\{ \frac{\sum (w_i - \bar{w})^2}{n - 1} \right\}^{1/2}$
- the sample skewness $k = \frac{n^{1/2} \sum (w_i - \bar{w})^3}{\left\{ \sum (w_i - \bar{w})^2 \right\}^{3/2}}$

where n is the number of observations. The skewness is a measure of the asymmetry of the distribution, and in particular of its non-normality. These statistics were calculated for each day and over all days. Table 2 gives the main results for each period and stop.

It is of some importance to determine how good a representative is the 'over all days' estimate $\bar{w}$ of the hypothetical true mean waiting time for that bus stop in that period of the day. If we know that all observations of waiting times at a particular stop could be regarded as being independent and drawn from the same distribution, then the standard error of $\bar{w}$ might be calculated from the sample variance in the usual manner. However, observations are not independent, as people may arrive at a stop in twos and threes, and the 'true' distribution of waiting times may well vary from day to day, affected by, for example, the weather or the day of the week. Both these phenomena mean that the usual 'standard error' will in fact be a considerable underestimate of the likely variation of $\bar{w}$ about the true mean waiting time over all days. A reliable estimate of the standard error may be made by considering the variations in the value of $\bar{w}$ for different days (see Appendix 3).

3.3 Discussion of the Results

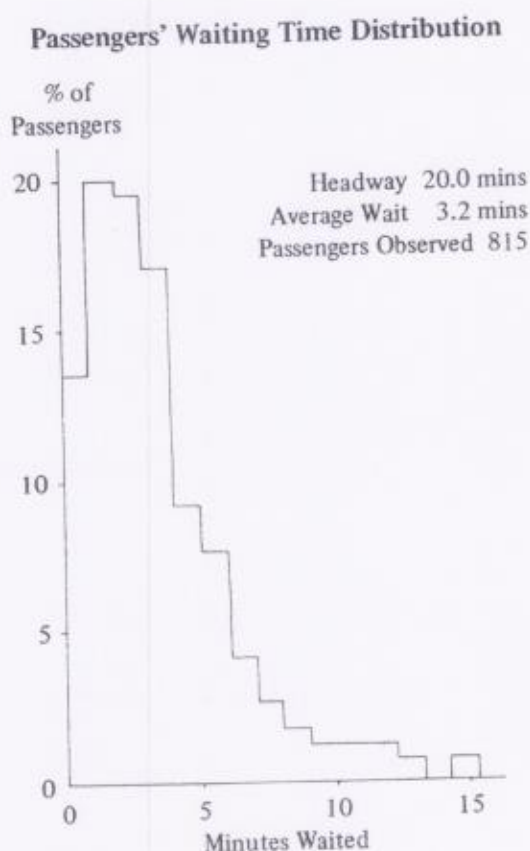
Table 2 gives the main results of the analysis. The periods are as defined in Chapter 2 and the stops are identified by the numbers used in Table 1. The column 'no. of people' means the number of passengers who arrived in those waiting periods. As stated earlier, waiting periods corresponding to more than one bus were omitted from the analysis, so that the occasional non-arrival of scheduled buses has not substantially affected the results.

It will be seen that, for each period of the day (except the evening peak), there was an almost constant headway between buses at most of the stops. Thus, it has been possible, in the analysis, to look simply at the distribution of waiting times for all buses at each stop (whether within each day or over all days), without incurring any serious loss of information.

The reliability of the buses, as measured by the standard deviation of the lag, is unusually good, and thus we would expect mean waiting times to be low, and certainly very substantially less than the mean headway (the theoretical mean waiting time for a 'random' pattern of bus arrivals). This is indeed the case, and the observation that waiting times are in general substantially less than half the mean headway, suggests that passengers time their arrival at a stop with a knowledge of the likely bus arrival time.

To obtain a sufficient sample of passengers, the stops involved in the survey had to be used by a substantial number of people, and to satisfy this condition (as well as the others outlined in Section 2.1) many of the stops had to be those at the termini of routes — both those in the suburbs and in the city centre. The arrival times of buses at those stops were frequently well in advance of the departure times and in consequence the mean 'pause' statistics do not in general represent the times taken to pick up passengers. They are of comparatively little interest, and will not be discussed further. It should, however, be pointed out that, whatever the relative merits of using bus arrival times or bus departure times in the definition of passenger waiting time, we have had no option but to use the departure time.

As might be expected, the distribution of waiting time at every stop was significantly skewed — by comparison with a random sample drawn from a normal distribution. The implication is that there are some passengers who do arrive at the stop without a knowledge of the likely bus arrival time. The measures of skewness, being difficult to interpret, are not included in Table 2. In general, however, the values obtained were very large. A typical distribution of waiting times is shown in Figure 1.



The results for bus stop 3 are somewhat anomalous, in that the waiting times are very low for all periods in which it was surveyed in spite of the high headways and considerable variability. This stop is quite near the terminus of its route, and on investigation it was found that most passengers could see the buses going up to the terminus from their houses, and could therefore time their arrival at the stop quite accurately. Since this is a rather unusual situation, this stop was excluded from subsequent investigations.

Figure 1. Histogram of Waiting Time Distribution



4. DEVELOPMENT OF PREDICTIVE MODELS

The object of the survey was to provide sufficient information to make possible the prediction of the likely waiting time of any individual at any bus stop in Huddersfield, given the characteristics of the (real or hypothetical) service available to him. The next stage was therefore to build a model relating the observed distributions of individuals' waiting times to the measured characteristics of those individuals' services, in particular the headway and also the measure of reliability described earlier.* The service available to each individual is known, as each of the stops involved was 'single service', i.e. all buses leaving it ran along the same route thereafter. This does not prevent reliable inference about individuals using multiple-service stops, provided we know their destinations and can thus deduce their effective service.

Prediction using such a model is effectively an interpolative process, and extrapolation to individuals having service characteristics outside the range of those observed or to towns other than Huddersfield will be less reliable. However, the range of service characteristics covered in the survey was such as to enable reasonable predictions of waiting times to be made in the case of all the networks and services considered in the study.

For prediction of mode and route choice, and particularly for economic evaluation, the most important measure of the waiting time distribution is almost certainly the mean, and while the standard deviation provides some additional information relevant to choice prediction, in practice there is little loss in taking this to be a simple function of the mean. Our present problem is therefore that of modelling the dependence of the mean waiting time on those characteristics given in the introduction:

- the headway of the individual's service
- its reliability, as measured by the standard deviation of the lag
- the time of day;

and possibly any other identifiable characteristic that appears relevant.

We first consider, separately for each of the four defined periods of the day, a model of the form:

$$\mu_j = a + bx_j + cy_j \quad (1)$$

where μ_j is the true mean waiting time at stop j , x_j is the mean headway, and y_j the standard deviation of the lag, at that stop.

There are some prior theoretical grounds for assuming a linear relationship between the mean waiting time at a stop and the headway, provided this is constant over time. If any variation in the headway over an interval of time is small, the same relationship will hold between the mean waiting time and the *mean* headway over the interval, but where the variation is considerable the relationship will give an underestimate of the mean waiting time. Because the observed variations in the headway *are* in general small (except in the evening peak), it should be acceptable to use simply the relation (1) to model the dependence on headway. However, the results for the evening-peak period may be less reliable.

* As stated in Chapter 3 we model the dependence of waiting time on a service in which all scheduled buses are presumed to run, and have excluded from our analysis all data for which this is not the case.

There are no such theoretical grounds for assuming a specifically linear dependence of the mean waiting time on the standard deviation of the lag, but the range of values taken by the observations of this variable is insufficient to substantiate any other relationship. By including the variable y_j in the model above we hope simply to verify its importance and obtain an idea of how much allowance to make for it in prediction.

For each observed stop we have an (unbiased) estimate $\hat{\mu}_j$ of μ_j . Equation (1) becomes:

$$\hat{\mu}_j = a + bx_j + cy_j + \epsilon_j \quad (2)$$

where ϵ_j is an approximately normally distributed random variable (by virtue of the central limit theorem) with

$$E(\epsilon_j) = 0 \quad E(\epsilon_j^2) = e_j^2 \quad (3)$$

where e_j is the standard error of ϵ_j . Because we have been able to estimate e_j independently for each observed stop we are able to assess whether equation (2) is sufficient to account for all the variation in the observed mean waiting times $\hat{\mu}_j$. A linear regression of $\hat{\mu}_j$ against x and y for each period of the day gives estimates of the residual standard error that are in general larger than the estimates e_j . The inference is that either the linear model is a less than perfect fit, or else that other sources of variation apart from the ϵ_j are present. The latter at least would appear to be the case. An inspection of the results for each period in Table 2 (excluding stop 3) suggests that these are not fully compatible with any model of the form:

$$\hat{\mu}_j = f(x_j, y_j) + \epsilon_j \quad (4)$$

where f is monotonic increasing in x and y and the ϵ_j satisfying (3) are the only source of random variation. Hence there is evidence of some random variation, for fixed x and y , in the true mean waiting times μ themselves. It was suggested in the introduction that some random variation might exist — presumably due to factors we have been unable to measure. To take account of this variation, we must modify the relation (1) to:

$$\mu_j = a + bx_j + cy_j + \epsilon_j^1 \quad (5)$$

where ϵ_j^1 is a random variable with

$$E(\epsilon_j^1) = 0 \quad E(\epsilon_j^{1^2}) = e^{1^2} \quad (6)$$

taking a value for each bus stop. Thus (2) is replaced by:

$$\hat{\mu}_j = a + bx_j + cy_j + \epsilon_j^1 + \epsilon_j \quad (7)$$

Estimates of the parameters a , b , c , their associated standard errors, and also the variance of the residuals $\epsilon_j^1 + \epsilon_j$, may be made by a suitably weighted linear regression analysis. The best choice of weights depends on the relative magnitudes of the standard errors e_j and e^1 . The magnitude of e^1 may strictly only be inferred from the regression itself, but it is possible to gain a sufficient idea of its magnitude from inspection of the data, and it would appear to be comparable with the sizes of the standard errors e_j . This implies that a simple weighting scheme should be adopted, giving equal weight, 1.0 to all observations, except where the standard error e_j of ϵ_j is usually large, when a smaller weight, say, 0.5 should be used.*

* Stop 5 in the evening peak was also given a weight of 0.5, since an unusually small number of passengers was observed there.

We have no reason to expect that there should be any real difference between waiting times in the morning and afternoon off-peak periods, and as the differences in the separate estimates of the parameters of the model (7) for the two periods are not statistically significant at even the most modest level, the results for the two periods should be combined so as to obtain increased accuracy in the estimates. Some care is then required in the analysis of the model (7) as the values of ϵ_j^1 required in the morning off-peak period and the same stop in the afternoon off-peak period will not be independent. In effect, we must use one observation for each stop observed in either or both periods.

The results of the analysis of the model (7) are given in Table 3. For the evening-peak period, the estimate of the parameter c (measuring the dependence of the mean waiting time on the standard deviation of the lag) was negative – though not significantly so. As it ought to be positive, this may be the result of using a simple model in which the headway is represented simply by its mean – as remarked earlier, considerable variation in headway exists in the evening peak. The model was simply reanalysed for this period with c constrained to be zero.

The results shown in Table 3 are discussed in the concluding chapter.

Table 3: Analysis of Linear Model

Model:

$$\mu_j = a + bx_j + cy_j + \epsilon^1 + \epsilon_j$$

where μ_j = observed mean waiting time at stop j

x_j = mean headway

y_j = standard deviation of 'lag'

ϵ^1 and ϵ_j are random variables as described in the text.

Period 1	(morning peak)	regression weights: stops 1–2 : 1 stops 4–9 : 2		
		variable	estimate	standard error
				significance level
		a	1.50	0.38
		b	0.036	0.017
		c	1.73	0.22
				1.0%
				4.5%
				0.1%
Period 2 + 3	(off-peak)	equal weight to all stops (the same stop in morning and afternoon periods being counted once)		
		variable	estimate	standard error
				significance level
		a	2.25	1.44
		b	0.166	0.060
		c	0.54	0.42
				9.0%
				1.9%
				12.9%
Period 4	(evening peak)	regression weights: stops 5 : 1 stops 10–14 : 2		
		variable	estimate	standard error
				significance level
		a	1.11	1.34
		b	0.365	0.109
		c	-0.081	0.20
				17.6%
If hold $c=0$		variable	estimate	standard error
		a	1.27	1.20
		b	0.345	0.87
		c	0.00	
				1.0%

5. CONCLUSIONS

This chapter summarizes the main conclusions to emerge from the analysis, and comments on some of their implications for the planning of bus services.

The results of the modelling analysis presented in Table 3 in the preceding chapter showed that waiting times are significantly affected by at least three factors.

1. **Time of day** – the results obtained for the three different times of day analysed showed clearly that *waiting times vary with the time of day*. Waiting times in the morning peak are less than those off-peak, which in turn are less than those in the evening peak. Moreover, the influences of headway and reliability are different at different times of day. It is reasonable to connect these effects with the different predominant purposes of trips being made in the different periods.
2. **Reliability** – even when all the scheduled buses actually run, *there is a significant dependence of waiting time on the accuracy of timekeeping*. The measure of reliability used in this study, which relates timekeeping to average rather than scheduled times, has been shown to be adequate in explaining at least some of this dependence. Reliability seems to be most important when passengers have the best chance of timing their arrivals at stops – i.e. in the morning peak.
3. **Headway** – the results indicate that although in general waiting time on low frequency routes is more than on high frequency routes, the *waiting time is much less than half the headway* for the range of headways surveyed. These low waiting times are particularly marked in the morning peak and on low frequency routes at all times.

These three factors do not explain all the variation, however, and it is clear that at certain stops particular local circumstances affected the waiting times significantly.

It must be emphasized that the results obtained are applicable only within the ranges for which observations were made. In particular, the range of headways was limited to seven and a half to thirty minutes, and the range of reliabilities by the practical considerations of selecting stops to survey. In fact, the stops selected (predominantly termini) should have better than average reliability for Huddersfield, since traffic conditions have less impact at termini than at stops further down the routes. The extent of unreliability and the form it takes will severely limit the transferability of these results to other places.

Within the limitations of the ranges to which the results are applicable, however, fairly accurate estimates of mean waiting times can be made. The residual standard errors of the regression equations were all less than one minute, so that this is an upper limit on the variation of mean waiting time due to factors not included in the modelling (e.g. the presence or otherwise of shelters). This accuracy, together with the generally high levels of significance shown in Table 3, gives considerable confidence in the results obtained.

Confidence is increased by the broad agreement between the results of this study relating headway to waiting time and those obtained elsewhere.* This agreement is that the mean waiting time increases with headway at about a quarter or a third the rate of headway increase. Other studies have not, however, measured reliability in the way used here, so no direct comparisons are possible, although it is noticeable that waiting times in large cities are higher for all headways than those in smaller cities, presumably as a result of different overall reliabilities.

It is clear from this evidence that the common assumption that waiting time is on average half the headway, is in error and should be abandoned. Instead the lower estimates derived from this and similar surveys should be substituted. Some further problems arise, however, first because a reduction in waiting time is not the only benefit given by a frequency increase, and second because the bus waiting times perceived by users of other modes may be higher than those achieved by the regular users who predominate in surveys. But errors caused by these problems are likely to be smaller and more readily corrected than those caused by inaccurate estimates of actual waiting times.

* E.g. the Coventry Bus Study (cf LGORU Report C149) and the Bradford Bus Study (R. Travers Morgan Technical Progress Papers P14 and P16) both of which quote earlier supporting surveys.

APPENDICES

SURVEY FORM AND SURVEYORS' INSTRUCTIONS

On the following pages are reproduced instructions as given to the surveyors and the form used in the survey (reduced from foolscap size).

Instructions to Surveyors

1. Each survey sheet covers the entire waiting period for one bus — starting from the moment of departure of the previous bus.
2. **Before starting**
Make sure that your name, and the month, date and day of the week are filled in on each survey sheet.
Check also that the space marked 'Bus Stop' has a code number in it.
Make sure your watch is right.
3. **At the start of the waiting period for each bus**
Enter the 'Start Time' and start the stopwatch. The 'Start Time' should be the time that the previous bus *actually* leaves the stop. Round the time to the minute below, e.g. record 0815½ as 0815.
Record also the initial number of passengers at the stop — this will normally be zero unless the previous bus was full. Record zero as 00, one as 01 etc.
If there is a known scheduled departure time from that stop for the next bus, record this in the space 'Next scheduled departure time'. Otherwise leave this space blank.
4. **In each minute of the waiting period**
Each minute on the stopwatch is covered by one row on the form.
Cross off a number in that row for each passenger who arrives in that minute, i.e. cross off '1' when the first passenger arrives, cross off '2' when the second arrives, etc.
At the end of the minute, *put a line through the rest of the row as far as number 25*, and enter the total number of passengers who arrived in that minute in the boxes on the right. This number will be the same as the last number crossed off. Record numbers less than 10 with a zero in front of them, e.g. 00, 01, 02, etc.
If *no* passengers arrive in that minute, *put a line through the entire row from 1 to 25* and enter zero (00) in the boxes on the right.
5. **When the bus arrives**
Place a vertical bar after the number corresponding to the last passenger who arrives at the stop before the bus arrives.
Continue recording passenger arrivals as before until the bus departs.
6. **When the bus departs**
Record the departure time (round to minute below as before), in the space 'Actual bus departure time' at the bottom of the form.
If the bus is full, record the number of passengers left at the stop. *If the bus is not full, leave this space blank.*

Take a new form for the next bus.

HUDDERSFIELD URBAN BUS STUDY

Passenger Waiting Time Survey

Surveyor

Month (Nov. = 1
Dec. = 2)

Date
2 3

Day of week

Bus Stop
4 8

Start Time
9 12

Initial no. of passengers
13 14

Next scheduled departure time
15 18

PASSENGER ARRIVALS

Min.	Arrivals																									Total
1	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	27
2	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
3	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
4	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
5	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
6	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
7	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
8	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
9	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
10	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
11	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
12	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
13	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
14	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
15	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
16	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
17	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
18	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
19	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
20	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
21	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
22	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
23	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
24	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
25	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
26	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
27	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	79
28	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	27
29	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
30	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
31	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
32	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
33	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
34	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
35	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	
36	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	43

Actual bus departure time
19 22

Passengers left
23 24 25

EFFECTS OF PRECISION IN RECORDINGS

All times observed in the survey — arrival and departure times of buses, and arrival times of intending passengers — were recorded to within an accuracy of one minute. Any attempt to survey with greater accuracy would, it was felt, have been much more difficult and would have provided very little extra information relevant to the subsequent analysis. In this appendix we consider in detail the effects of selecting this level of precision.

1. Effect on Headway Estimates

The headway associated with each bus was defined to be the difference between its departure time and the departure time of the previous bus. The departure time of each bus is recorded as the nearest minute *before* on the clock. If d_j is the true departure time in minutes of bus j , and d_j^1 is the recorded time then:

$$d_j = d_j^1 + \delta_j \quad (1)$$

where the δ_j are, say, uniformly distributed over $(0,1)$ and may reasonably be assumed to be independent of each other. Hence the relation between true headway h_j and calculated headway h_j^1 is:

$$h_j = h_j^1 + \epsilon_j \quad (2)$$

$$\text{where} \quad \epsilon_j = \delta_j - \delta_{j-1} \quad (3)$$

$$\text{Hence} \quad E(\epsilon_j) = 0 \quad \text{var}(\epsilon_j) = 1/6 \quad (4)$$

so that the standard error associated with the calculation of any *individual* headway is $1/\sqrt{6}$ (≈ 0.404) minutes.

However, our interest is in the mean headway $\bar{h}$ at a bus stop and in a particular period of the day. If this is calculated from n individual observations taken over r separate intervals (r is the number of days of observation if this is continuous on each day), then assuming (1) we have:

$$\bar{h} = \bar{h}^1 + \epsilon \quad (5)$$

$$\text{where} \quad E(\epsilon) = 0 \quad \text{var}(\epsilon) = \frac{r}{6n^2}$$

n is typically about 50, and r about 8, so that $\bar{h}^1$ is an unbiased estimate of $\bar{h}$ with a standard error of the order of a second! We also calculate — but do not use in our final analysis — the sample standard deviation of the headway for each stop and period of the day. This is defined to be:

$$\sqrt{\sum (h_j - \bar{h})^2 / (n-1)}$$

It may be shown that

$$E\left\{\sum (h_j - \bar{h})^2 / (n-1)\right\} = E\left\{\sum (h_j^1 - \bar{h}^1)^2 / (n-1)\right\} \leq 1/6 \quad (6)$$

2. Effect on Lag Estimates

The lag associated with each bus was defined to be the difference between its actual and its scheduled departure time. Since the latter is known exactly, we have from (1):

$$l_j = l_j^1 + \delta_j \quad (7)$$

where l_j and l_j^1 are the actual and calculated lags and δ_j is as defined above. For each stop and period of the day our interest is in the sample standard deviation of this statistic, defined to be:

$$\sqrt{\Sigma(l_j - \bar{l})^2 / (n-1)}$$

where n is the total number of buses. It may be shown that under reasonable assumptions about the distribution of the δ_j

$$|E\{\Sigma(l_j - \bar{l})^2 / (n-1)\} - E\{\Sigma(l_j^1 - \bar{l}^1)^2 / (n-1)\}| \leq 1/12 \quad (8)$$

However, there is evidence of a tendency on the part of many surveyors to round actual departure times towards those schedules, particularly when the differences were small. The effect of this is to make the calculated standard deviation a serious underestimate of the true standard deviation – the error being substantially greater than anything associated with simply rounding to the minute below. For this reason we may only be sure of our qualitative deduction that reliability is an important determinant of waiting times. Its quantitative effect is best calculated from theoretical considerations.

3. Passenger Waiting Times

For the waiting period associated with each bus, the surveyor recorded the number of passenger arrivals in each minute, having zeroed his stopwatch at the exact departure time of the previous bus. The last minute for which this was done was that in which the bus departed. Thus if there are m minutes of observations, we define the waiting time w^1 of a passenger arriving in the r th minute to be:

$$\begin{aligned} w^1 &= m - r & \text{if } r < m \\ w^1 &= 1/4 & \text{if } r = m \end{aligned} \quad (9)$$

Assuming accurate observation the relation between the true waiting time w of the individual and w^1 is:

$$w = w^1 + \delta \quad (10)$$

$$\text{where } E(\delta) = 0 \quad \text{var}(\delta) = \begin{cases} 1/6 & \text{if } w^1 = 1/4 \\ 7/144 & \text{if } w^1 \geq 1 \end{cases} \quad (11)$$

This implies that the relation between the calculated sample mean waiting time $\bar{w}^1$ and the true sample mean waiting time $\bar{w}$ of a sample of n passengers is:

$$\bar{w}^1 = \bar{w} + \epsilon \quad (12)$$

$$\text{where } E(\epsilon) = 0 \quad \text{var}(\epsilon) \leq 1/6n$$

Since the number of passengers used to estimate the mean waiting time at any stop is typically several hundred, this again means that the error associated specifically with the minute-by-minute recording procedure is only of the order of a second. This is quite negligible compared with the variation of $\bar{w}$ about any true mean waiting time, and resulting from the finiteness of the sample and variations in conditions over time.

Similarly we may derive the following relation between the calculated and true sample standard deviation of the waiting time

$$\sqrt{\Sigma(w_i - \bar{w})^2 / (n - 1)}$$

although this statistic was not used in subsequent analysis:

$$| E \{ \Sigma(w_i - \bar{w})^2 / (n-1) \} - E \{ \Sigma(w_i^1 - \bar{w})^2 / (n-1) \} | \leq 1/6 \quad (13)$$

For the waiting interval associated with each bus there was also recorded:

1. the start time of the interval (departure time of the previous bus), rounded to the minute below
2. the end time of the interval (departure time of the bus), again rounded to the minute below.

The difference between these two recorded times is h , the recorded headway, and we should have:

$$h = m - 1$$

or

$$h = m$$

and nothing else (m being the number of minutes of observations). Further, we would reasonably expect to observe $h = m-1$ about as frequently as $h=m$. This does not happen — $h=m$ is much more frequent, and we also occasionally observe $h=m+1$. This suggests the following phenomenon: if the bus departs $m+\lambda$ minutes after the start of the waiting interval when λ is small — less than a half, say — the surveyor may only record m minutes of passenger arrivals. Accordingly, we modify relation (9), increasing w^1 slightly, and use instead:

$$w^1 = m - r + 1/4 \quad r \leq m$$

This, we hope, preserves as far as possible the unbiased relation between w^1 and w , but w^1 may be in error as an estimate of the sample mean w by anything up to a quarter of a minute.

The 'pause', i.e. the difference between the arrival and departure times of each bus, is calculated in the same way as passenger waiting times (the bus arrival time being recorded in the same way as a passenger arrival time). Similar remarks as to its accuracy therefore apply.

STANDARD ERROR OF ESTIMATE OF MEAN WAITING TIME

When the sample mean:

$$\bar{w} = \frac{1}{n} \sum_{i=1}^n w_i \quad (1)$$

is used to estimate the true mean of a population, the usual calculation of the standard error of this estimate

$$\left\{ \frac{\sum (w_i - \bar{w})^2}{n(n-1)} \right\}^{1/2} \quad (2)$$

is based on the assumption that the observations w_i are statistically independent samples from the population.

For the set of all passenger waiting times $[w_i]$ observed at any given stop (in a given period of the day), this assumption does not hold, primarily because of the two reasons stated in Chapter 3:

1. passengers may occasionally arrive at the stops in groups of two or more
2. conditions influencing waiting times (e.g. weather) may vary from day to day.

Both these complications may be avoided by estimating the standard error instead from the variations in the sample means for the individual days of observation as follows: we assume, without any real loss of generality, that the relation between the sample means $\bar{w}_j$ for the individual days and the true mean μ is

$$\bar{w}_j = \mu + \epsilon_j$$

where the ϵ_j are independent random variables

$$E(\epsilon_j) = 0 \quad E(\epsilon_j^2) = \frac{\sigma_1^2}{n_j} + \sigma_2^2$$

and where n_j is the number of individual waiting times observed on day j , and σ_1^2 and σ_2^2 are constants. The best estimate $\bar{w}$ of μ and its standard error strictly depend on the relative magnitudes of σ_1^2 and σ_2^2 but provided the n_j do not differ substantially from day to day (as is the case in this survey) it makes negligible difference what ratio we take. We therefore estimate μ by

$$\bar{w} = \sum n_j \bar{w}_j / n_j$$

which is also the overall sample mean derived from the individual observations, and take its standard error to be

$$\sqrt{\sum n_j (\bar{w}_j - \bar{w})^2 / (r-1) \sum n_j}$$

where r is the number of days of observation.

This estimate is in general significantly higher for the observed waiting times than the naive estimate (2), which would have indicated a spurious degree of accuracy in the estimated mean waiting times.