INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or “target” for pages apparently lacking from the document photographed is “Missing Page(s)”. If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.

2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of “sectioning” the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.

University Microfilms International
300 N. Zeeb Road
Ann Arbor, MI 48106
Matthews, Timothy Charles

THE ACCURACY OF TIME SAMPLING PROCEDURES FOR ESTIMATING
BEHAVIOURAL FREQUENCY

University of Hawaii

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106

Copyright 1985 by Matthews, Timothy Charles All Rights Reserved
THE ACCURACY OF TIME SAMPLING PROCEDURES
FOR ESTIMATING BEHAVIOURAL FREQUENCY

A DISSERTATION SUBMITTED TO THE GRADUATE DIVISION OF THE
UNIVERSITY OF HAWAII IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN EDUCATIONAL PSYCHOLOGY
MAY 1985

By
Timothy C. Matthews

Dissertation Committee:
Harold I. Ayabe, Chairman
Daniel D. Blaine
Mary E. Brandt
Charles A. Glisson
Roland G. Tharp
ACKNOWLEDGEMENTS

I owe a deep debt of gratitude to many people who have provided the support and encouragement which enabled me to complete this study. Foremost among these people is my wife, Noelle, who provided the initial incentive for me to undertake graduate study, and who has provided unwavering and unquestioning support ever since, often making personal sacrifices in order to maintain that support.

Equally strong support was provided by Professor K. Brian Start of the University of Melbourne, Australia, with whom I worked for several years, who encouraged my graduate studies and brought research opportunities to my attention, and who has become a valued friend and colleague.

I am equally indebted to my parents, Harold and Betty, who provided a home environment which valued an inquiry approach and the endeavour to always do one's best, and who have provided invaluable support for my educational pursuits for the last thirty years.
I sincerely thank all the members of my committee for their constructive and insightful comments during the execution of this study. My thanks are due also to the other faculty members of the Department of Educational Psychology at the University of Hawaii from whom I have learned so much.

Acknowledgement is due to the Education Department of Victoria, Australia, which helped financially to make this study possible by awarding me a Fellowship for Advanced Study.

Acknowledgement is also due to the East-West Center, Honolulu, which provided financial support by granting me a Degree Participant Award to undertake doctoral studies.
ABSTRACT

The term time sampling covers a broad range of techniques whose common aim is to quantify observed behaviour in a sample of time during which the behaviour is observed. Researchers using time sampling aim to infer the true pattern of the behaviour of interest from data on the observed pattern of behaviour. Inaccuracy in data obtained from time sampling procedures was classified into two sources: (1) the observers and subjects, and (2) time sampling procedures themselves. This paper focused on the inaccuracy due to the procedures themselves. Inaccuracy due to time sampling procedures was classified into two sets of factors: (1) the pattern of true behaviour, and (2) the design decisions made by the researcher. Recent studies have suggested that accuracy of time sampling is affected by the interaction of these two sets of factors. Since the researcher cannot manipulate the pattern of true behaviour, this paper focused on the design decisions made by the researcher in an attempt to ascertain the relationship between design decisions and accuracy. The study employed a Monte Carlo approach, using a specially written computer program to generate 1000 "true" behaviours and then to simulate time sampling studies based on these data under a wide range of values of time
sampling design factors. The simulated time sampling data was analysed by multiple linear regression techniques to ascertain factors related to accuracy of the data. For all three simulated time sampling procedures the regression analyses produced an $R^2$ value in excess of .99. The results of the study show that accuracy of time sampling is a linear function of certain design factors. Two conclusions were drawn from the findings of this study: (1) accuracy of time sampling procedures can be perfectly predicted from mean duration of behaviour, length of time interval, and length of gap between intervals, and (2) most previous studies using time sampling procedures have produced inaccurate data. Three applications of the findings of the study to time sampling studies were noted: (1) a formula was derived to enable values of design factors to be set such that the data collected is free from inaccuracy produced by time sampling procedures, (2) a formula was derived which can be used as a "correction factor" for data previously collected by time sampling, and (3) the usual method of using time sampling in settings where the true behaviour pattern changes over the period of the study will result in data of differential accuracy over the period of the study, but this problem can be overcome by altering values of design variables over the period of the study.
TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................ iv

ABSTRACT .................................................. vi

LIST OF TABLES ........................................... xi

CHAPTER I. OVERVIEW OF BEHAVIOURAL OBSERVATION

1.1 Why Observe Behaviour? ................................ 1
1.2 Classification of Behavioural Observation Methods .... 3
1.3 The Shift from Qualitative to Quantitative Methods .... 5
1.4 Objectivity through Quantitative Methods ............. 6
1.5 Qualitative Methods
   1.5.1 Diary Descriptions ................................ 7
   1.5.2 Specimen Description .............................. 9
1.6 Quantitative Methods
   1.6.1 Time Sampling ................................... 10
   1.6.2 Event Sampling .................................. 13
   1.6.3 Trait Rating ................................... 15

CHAPTER II. METHODOLOGICAL ISSUES IN BEHAVIOURAL OBSERVATION

2.1 Systematic and Random Errors of Measurement ......... 17
2.2 Major Sources of Systematic Error in Behavioural Observation ........................................ 19
2.3 Systematic Error: The Subjects
   2.3.1 Subject Reactivity ................................ 20
   2.3.2 Response Sets ................................... 24
2.4 Systematic Error: The Observer
   2.4.1 Researcher Expectancy ............................ 26
   2.4.2 Knowledge of Reliability Assessment .......... 31
   2.4.3 Observer Drift .................................. 34
2.5 Systematic Error: Design and Analysis
   2.5.1 Calculation of Reliability Measures ............ 36
   2.5.2 Design Considerations in Time Sampling ....... 45
2.6 Scope and Purpose of the Present Study
   2.6.1 Scope of the Study .............................. 55
   2.6.2 Aim of the Study ............................... 58
### CHAPTER III. METHOD

3.1 Overview ........................................ 61
3.2 Independent Variables .......................... 62
   3.2.1 Independent Variables: 
      True Behaviour Pattern ....................... 62
   3.2.2 Generating the True 
      Behaviour Pattern ............................ 67
   3.2.3 Independent Variables: 
      Design Decisions .............................. 73
   3.2.4 Simulating Time Sampling 
      Studies ........................................ 76
3.3 The Dependent Variable ......................... 78
3.4 Method of Analysis .............................. 81
   3.4.1 The Model ................................. 81
   3.4.2 The Unit of Analysis ....................... 82
3.5 Summary .......................................... 83

### CHAPTER IV. RESULTS: TIME SAMPLING PARAMETERS

4.1 Overview .......................................... 85
4.2 Nominal Values of Parameters .................. 87
   4.2.1 Distributions of Nominal 
      Values of Parameters ........................ 87
   4.2.2 Intercorrelations of Nominal 
      Values of Parameters ........................ 93
4.3 Actual Values of Parameters .................... 98
   4.3.1 Distribution of Actual Values 
      of Parameters ............................... 98
   4.3.2 Intercorrelations of Actual 
      Values of Parameters ....................... 108
4.4 Distribution of Proportion 
    of Occurrence .................................. 115
4.5 Summary of Characteristics of 
    Generated Data .................................. 116

### CHAPTER V. RESULTS: REGRESSION ANALYSES

5.1 Model One ........................................ 118
5.2 Model Two ........................................ 123
5.3 Model Three ...................................... 132
CHAPTER VI.

6.1 Summary of Findings 135
6.2 Relation of Findings to Previous Research 136
6.3 Application of Findings to Time Sampling Studies 139
6.4 Implications of Findings of Present Study 143

APPENDIX A 148

BIBLIOGRAPHY 160
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Hypothetical Data Showing Observer Reliability</td>
<td>39</td>
</tr>
<tr>
<td>2</td>
<td>Frequency Distribution of Values of Nominal Mean Duration of Behaviour</td>
<td>88</td>
</tr>
<tr>
<td>3</td>
<td>Frequency Distribution of Values of Nominal Standard Deviation of Duration of Behaviour</td>
<td>88</td>
</tr>
<tr>
<td>4</td>
<td>Frequency Distribution of Values of Nominal Mean Duration of Non-Behaviour</td>
<td>89</td>
</tr>
<tr>
<td>5</td>
<td>Frequency Distribution of Values of Nominal Standard Deviation of Duration of Non-Behaviour</td>
<td>89</td>
</tr>
<tr>
<td>6</td>
<td>Frequency Distribution of Values of Time Interval in Seconds</td>
<td>90</td>
</tr>
<tr>
<td>7</td>
<td>Frequency Distribution of Values of Gap between Time Intervals in Seconds</td>
<td>90</td>
</tr>
<tr>
<td>8</td>
<td>Frequency Distribution of Number of True Occurrences of Behaviour</td>
<td>91</td>
</tr>
<tr>
<td>9</td>
<td>Chi-square Goodness-of-fit Tests on Distribution of Values of Nominal Parameters</td>
<td>92</td>
</tr>
<tr>
<td>10</td>
<td>Intercorrelations between Values of Nominal Parameters</td>
<td>94</td>
</tr>
<tr>
<td>11</td>
<td>Frequency Distribution of Values of Actual Mean Duration of Behaviour</td>
<td>100</td>
</tr>
<tr>
<td>12</td>
<td>Frequency Distribution of Values of Actual Standard Deviation of Duration of Behaviour</td>
<td>101</td>
</tr>
<tr>
<td>13</td>
<td>Frequency Distribution of Values of Actual Mean Duration of Non-behaviour</td>
<td>102</td>
</tr>
<tr>
<td>14</td>
<td>Frequency Distribution of Values of Actual Standard Deviation of Duration of Nonbehaviour</td>
<td>103</td>
</tr>
<tr>
<td>TABLE</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>15</td>
<td>Intercorrelations between Values of Actual Parameters</td>
<td>109</td>
</tr>
<tr>
<td>16</td>
<td>Frequency Distribution of Values of Actual Proportion of Occurrence</td>
<td>115</td>
</tr>
<tr>
<td>17</td>
<td>Summary of Stepwise Regression Solution for Whole Interval Time Sampling: Model 1</td>
<td>120</td>
</tr>
<tr>
<td>18</td>
<td>Summary of Stepwise Regression Solution for Partial Interval Time Sampling: Model 1</td>
<td>120</td>
</tr>
<tr>
<td>19</td>
<td>Summary of Stepwise Regression Solution for Momentary Time Sampling: Model 1</td>
<td>121</td>
</tr>
<tr>
<td>20</td>
<td>Summary of Stepwise Regression Solution for Whole Interval Time Sampling: Model 2</td>
<td>129</td>
</tr>
<tr>
<td>21</td>
<td>Summary of Stepwise Regression Solution for Partial Interval Time Sampling: Model 2</td>
<td>130</td>
</tr>
<tr>
<td>22</td>
<td>Summary of Stepwise Regression Solution for Whole Interval Time Sampling: Model 2</td>
<td>131</td>
</tr>
<tr>
<td>23</td>
<td>Summary of Regression Solutions for three Time Sampling Procedures: Model 3</td>
<td>133</td>
</tr>
<tr>
<td>24</td>
<td>Parameter Estimates from Regression Solutions for three Time Sampling Procedures: Model 3</td>
<td>133</td>
</tr>
</tbody>
</table>
CHAPTER I
OVERVIEW OF BEHAVIOURAL OBSERVATION

1.1 Why Observe Behaviour?

It is almost trite to say that people constantly wonder about human behaviour. "Of all the most human tendencies none is more human than wondering about ourselves and other human beings... It was this basic human tendency that gave rise to the field we call psychology" (Sprinthall and Sprinthall, 1981, p.3). We often note a person's actions, predispositions, or abilities and wonder why he did whatever it was, why he seems to prefer a particular thing, or how he became so capable in an area. Usually the trigger which starts this line of thought is a casual observation of some action or behaviour of a person. This casual observation may lead to wondering about what process was working to produce the action observed. However, we cannot observe the process directly. "Only behaviours can be observed; the processes and characteristics which caused the behaviour are out of sight. They can be inferred, but they cannot be observed" (Cartwright and Cartwright, 1974, p. 7).
Herein lies the fundamental purpose of behavioural observation; that is, to systematically collect information about observed behaviour in order to draw inferences about the processes which explain that behaviour. Irwin and Bushnell (1980) present several more specific purposes for behavioural observation, including: to generate hypotheses or ideas, to answer specific questions, to provide a more realistic picture of behaviour and events than other methods, to help understand behaviour, and to evaluate people.

The notion of behavioural observation is extremely straightforward. As Wright (1960) notes, "... this must be the simplest way of all to study child behaviour. One gets within seeing and hearing distance of a child, observes and records something of his behaviour or situation or both, and then scores, classifies, summarizes, freely interprets, or otherwise does something with the recorded observations" (p. 71).

Over the last eighty years behavioural observation as a general approach has been applied through a number of methods to the study of innumerable questions relating to human behaviour and the processes behind it. Developing the point made by Cartwright and Cartwright
(1974) that only behaviours can be observed and not the processes behind them, it can be argued that the only way to collect information relating to the processes of the mind is through behavioural observation methods. This information is then used as base data for inferences about these processes. Accepting this notion leads to the suggestion that the field of psychology rests entirely on behavioural observation in the broad sense for its base data.

1.2 Classification of Behavioural Observation Methods

This discussion will focus on those observational methods termed by Wright (1960) as direct observations. The two criteria Wright uses to distinguish direct observation methods are: "(1) No planned arrangements stand between the observer and his target phenomena, and (2) neither does appreciable time: recording closely follows observing" (Wright, 1960, p. 71). Consequently such methods as tests, interviews, questionnaires, and various other self-report measures are excluded, as too are observations collected from secondary sources, retrospective descriptions, and physiological measures such as those produced by electro-cardiogram machines and similar devices. It should be noted however that the
areas excluded from this discussion are still behavioural observation methods in the broader sense.

Direct observational methods have been classified by different authors according to different criteria. Wright (1960) classifies direct observation methods on the basis of selection of the content and timing of the behaviour to be observed, the method of recording the data, and the method of analysing the data. Achenbach (1978) distinguishes 'open' methods where the behaviour is recorded without prior categorization from 'closed' methods where the behaviour is recorded in terms of pre-defined categories. Irwin and Bushnell (1980) distinguish 'narrative' procedures from 'sampling' procedures. Kerlinger (1973) distinguishes procedures by their degree of objectivity. Phinney (1982) points out that "In all observation, a compromise must be made between the qualitative aspects - the richness, the context, the sequence of events; and the quantitative aspects - the rate or frequency of events and the possibility for statistical evaluation" (p. 21). The methods will be differentiated in this discussion on the basis of the type of data generally collected - qualitative versus quantitative. The qualitative methods discussed here are the diary description and the specimen
description. Both these methods collect the data in the form of a narrative description, and analyses tend to be qualitative in nature. Quantitative methods presented here are time sampling, event sampling, and rating scales, where data is usually (but not necessarily) collected in a quantifiable manner, and analyses tend to be quantitative.

1.3 The Shift from Qualitative to Quantitative Methods

The earliest approaches to behavioural observation were of the qualitative type. However these methods rely heavily on the ability of the observer to accurately record what he sees in narrative form, and tend to produce a large amount of information which is not easy to manipulate for specific purposes. They have been criticised for being too costly in observation time, too limited in sample size, and too biased (Irwin and Bushnell, 1980).

The quantitative approaches have been developed in an attempt to provide objectivity, control, and efficiency (Irwin and Bushnell, 1980). Objectivity, control, and efficiency are produced by careful and detailed specification of all aspects of the data collection which includes operationally defining the behaviour(s) of
interest, training observers to apply coding schemes correctly, selecting a small number of behaviours to observe, specifying particular time constraints for the observations, carefully defining what constitutes an occurrence of the behaviour of interest, and a number of similar techniques.

1.4 Objectivity through Quantitative Methods

The techniques for enhancing objectivity essentially provide a set of rules within which the observations are made, on the assumption that a prescription of rules will provide objective data. "Objective methods of observation are those in which anyone following the prescribed rules will assign the same numerals to objects and sets of objects as anyone else. An objective procedure is one in which agreement among observers is at a maximum" (Kerlinger, 1973, p. 491).

This section of the paper provides a brief overview of several of the most common methods of direct observation. Wright (1960) notes that direct observation is "... a scientific practice that includes observing and associated recording and analysis of naturally occurring things and events" (p. 71). The methods are subdivided into qualitative and quantitative methods as
discussed above. Quantitative methods are presented here with the assumption that they will provide the objectivity to which Kerlinger (1973) refers. This assumption of objectivity is discussed in the following section.

1.5 Qualitative Methods
1.5.1 Diary Descriptions

Perhaps the simplest method of behaviour observation is that known as diary descriptions. As the name suggests, the diary description involves collecting observations over a period of time in the form of a narrative record or diary. The behaviours noted are usually limited to new behaviours. The diary may be limited to behaviours in particular areas, referred to as a topical diary, or may attempt to cover all areas of behaviour, referred to as a comprehensive diary (Wright, 1960). The diary of observations, then, provides a sequential record of growth changes and behaviour episodes. Sometimes referred to as baby biographies, diary descriptions have been mainly applied to observation of children, and usually by their parents or other close relatives.
The keeping of a diary is probably the oldest of behaviour observation techniques. Certainly the oldest extant studies of behaviour employed this method. Wright (1960) refers to a diary description published by Teidmann in 1787. Irwin and Bushnell (1980) claim that the first published diary description of an infant was Pestalozzi's work "A Father's Diary" published in 1774. Achenbach (1978) lists among practitioners of this method such people as Charles Darwin and Jean Piaget.

The diary description has the obvious advantage of simplicity. However, its usefulness is highly dependent on the observer's skill in accurately recording observed behaviours. The further complication, even if the records are complete and accurate, lies in the nature of the data. Data recorded in diaries are voluminous and are qualitative, thus requiring considerable effort and subjectivity to analyse them. Furthermore, as Irwin and Bushnell (1980) point out, the data were usually based on a biased selection of subjects, were recorded by biased observers, and provided too few cases for generalization.
1.5.2 Specimen Description

The specimen description technique is very similar to the diary description in that it is a narrative description of behaviour or events as is the diary description. The specimen description method differs from the diary description method in that it is based on a scheduled and continuous observation period during which the observer records everything about the behaviour and the setting of the child. The choice of the child and the setting (time and place) are made in advance to suit the particular interest of the research. The length of the observation period used has varied from a few minutes up to a whole day of continuous observing and recording. The resulting descriptive narrative is intended to provide a specimen of the behaviours and immediate settings of the children observed, hence the name specimen description (Wright, 1960).

As with diary descriptions, specimen descriptions have the advantage of simplicity. Specimen descriptions have an added advantage in that an attempt is made to record all behaviours and therefore can be reviewed repeatedly for different purposes. However, like the diary descriptions the data produced are voluminous and qualitative. Achenbach (1978) points out that a specimen
description of one boy's behaviour during one day filled 435 published pages. The data needs to be reduced in some way to a more manageable amount. Wright (1960) suggests a number of ways in which this may be done, all of which require judgements and interpretations on the part of the analyst, and require immense amounts of time to accomplish. However, Irwin and Bushnell (1980) suggest that the classroom teacher may usefully employ the method without the need to develop coding systems for reducing the data after it is collected.

1.6 Quantitative Methods

1.6.1 Time Sampling

Time sampling aims to collect data only on specific predetermined behaviours over a finite predefined observation period. The data collected usually relates to duration of occurrence or frequency of occurrence of a behaviour. The method used to collect the observational data is to break the duration of the study up into a number of time intervals, and collect data only within some intervals. The observations would be made at regular predefined intervals and for regular predefined lengths of time within the observational period. Wright (1960) states that the purpose of this method is "...
Several features of this method are worthy of note. Firstly, the behaviour(s) to be observed must be selected in advance of the data collection. The behaviour must be operationally defined so that the observers consistently record the same behaviour. Operational definitions need to be precise and unambiguous. As an example, consider a study aiming to look at aggressiveness of children. The researcher may define punching another child as behaviour which indicates aggression. However, the operational definition of 'punching another child' needs to be precise enough that this behaviour can be easily distinguished by the observer(s) from similar behaviours such as slapping, pushing, prodding or patting another child.

Secondly, the data collected are based on overt behaviour. The hypothetical study of aggression aims to examine a trait which can be observed directly. Thus, the data collected are of behaviours considered by the researcher to relate to aggression. From the data on the child's behaviour, the researcher will make inferences about aggression. Thus, to the degree that the behaviour
to be observed is not isomorphic to the trait under study, the data collected will not be a pure reflection of the trait the researcher wishes to make inferences about and may lead him to erroneous conclusions.

Thirdly, the length of each time interval and the space between time intervals (and therefore the total number of intervals in a given period of time) must be set in advance by the researcher. It is intuitively obvious that the length of time interval and gap between time intervals will affect the accuracy and volume of the data generated. If the researcher makes inappropriate choices for the length of time interval or gap between time intervals, the accuracy of the data may be seriously diminished. As an extreme example, if a researcher decided to study the behaviour of a child over a three hour period by observing the child for one five-second interval during that period, then his inference would be either that the behaviour never occurs, or that the behaviour always occurs, depending on whether it was or was not observed during the time sample in which the observation data was collected.

Finally, since the researcher may be collecting data on several behaviours simultaneously, and since the
behaviours have been operationally defined in advance, it is customary to draw up some form of coding sheet, typically with one row for each behaviour and one column for each time period. The coding sheet enables the observer to simply place a mark against each behaviour which occurs during the time period, thus providing a simple and uniform recording scheme.

1.6.2 Event Sampling

As the name of the technique suggests, the purpose of this approach is to sample 'events' in the behaviour of the subjects. The researcher defines in advance which events will be observed, and provides operational definitions of these. The researcher also defines in advance what information will be collected. McLaughlin (1975) distinguishes 'event sampling', which involves taking a frequency count of the number of times a discrete behaviour occurs, from 'duration recording', which is the process of measuring the duration for which a specific behaviour occurs. Other types of information which may be collected include the nature of the behaviour; the setting in which the behaviour occurred; the sex, age and other biographical information for each participant in the event; and the role of each
participant in the event. The information is usually entered onto a recording sheet prepared in advance.

The type of data collected by event sampling is not limited to either qualitative or quantitative alone. Depending on the nature of the research question, the researcher may collect quantitative data such as the duration of the event, qualitative data such as a narrative description of the setting, or both types of data simultaneously. Event sampling also allows the use of checklists and other similar coding schemes to be employed in conjunction with narrative or quantitative data.

The method has the advantage of observing behaviour in naturally-occurring integral units. In other words the event is the unit of analysis and "... each event is a sample of its own class in the behaviour streams of classified children in selected life settings" (Wright, 1960, p.104). Thus the information gained for each event observed is considered as a sample of the behaviour, and generalizations are made from the events which were sampled to the population of events from which this sample was drawn.
Having defined the behaviour in which he is interested and the information to be collected, the researcher "... waits for the selected behaviour to occur and then records it" (Irwin and Bushnell, 1980, p.178). Waiting for the behaviour to occur would appear to make the data collection an expensive exercise since the observer must wait for the behaviour to occur. However, the observer is free to do other things when the behaviour of interest is not occurring. Depending on the situation this may prove very useful. For instance if the setting was a hospital and the observer was a nurse employed at the hospital, then she could carry out her normal duties whilst not observing, thus providing a very economical and relatively unobtrusive means of data collection.

1.6.3 Trait Rating

Like time sampling and event sampling, trait rating looks at predetermined types of behaviour, this time with the intention of making an estimate of the strength of a particular trait or traits in the subject(s). The observer watches the subject for a period of time, perhaps taking notes as he observes, and then "... on the basis of his accumulated observations, he rates the subject on prespecified traits" (Achenbach, 1978, p.169).
The ratings would be made on some sort of predetermined scales, although the format of trait rating scales varies widely. The method is very simple, and is perhaps best summed up by Irwin and Bushnell (1980): "Rating scales are simply measures designed to quantify impressions gained from observation" (p. 204).

The observer who makes the ratings does so on the basis of his own judgement, which is both the strength and the weakness of trait rating. Achenbach (1978) notes that allowing the observer to make ratings on the basis of his own judgement enables the observer to "... take account of the cumulative context and intensity of behaviour more flexibly than he could if recording narrowly defined behaviours by means of time or event sampling" (p. 189). However allowing the observer to exercise his own judgement leads Kerlinger (1973) to conclude that "The intrinsic defect of rating scales is their proneness to constant or biased error" (p. 548).
CHAPTER II
METHODOLOGICAL ISSUES IN BEHAVIOURAL OBSERVATION

2.1 Systematic and Random Errors of Measurement

Consideration of direct behaviour observation methods raises a wide range of methodological issues. Many of the issues raised are those which are generally of concern to empirical and experimental studies. General methodological issues have been discussed in detail by a number of authors, especially Campbell and Stanley (1963).

However, many of the methodological issues are relatively specific to those direct behaviour observation methods categorized previously as quantitative methods. This section discusses some of the methodological issues which are of particular relevance to, or perhaps even specific to, quantitative behavioural observation methods. The issues presented here are discussed with particular reference to the distinction between systematic and random errors of measurement.

Random errors of measurement are "... the sum or product of a number of causes: the ordinary random or
chance elements present in all measures due to unknown causes, temporary or momentary fatigue, fortuitous conditions at a particular time that temporarily affect the object measured or the measuring instrument, fluctuations of memory or mood, and other factors that are temporary or shifting" (Kerlinger, 1973, p. 443).

Random errors of measurement in behavioural observation data are produced by a wide range of factors: inadequate observer training, ambiguous definitions of the behaviour to be observed, variability between the subjects or the observers, and so on. Random errors can be minimized by careful thought and planning on the part of the researcher. Some of the techniques which may be employed to reduce random error were noted in section 1.3 above.

Random errors of measurement are by definition self-compensating; that is they do not produce a bias in the set of measurements obtained, but tend to cancel out the effects of each other. The effect of random error essentially is to reduce the reliability of the data without distorting it.
Systematic errors on the other hand are those which tend to lean in one direction, thus producing a bias in the data. Systematic error is far more insidious and difficult to manage than is random error. It is often very difficult to detect in actual studies. Systematic error will frequently tend to increase the reliability of the data (and hence, the confidence the researcher places in the data), but it will reduce the validity of the data by introducing biases that are often not recognized.

2.2 Major Sources of Systematic Error in Behavioural Observation

The three fundamental ingredients of behavioural observation studies are the subject, the observer, and the design of the study. All three of these facets can introduce systematic biases into the data collected by direct observation methods. This section briefly discusses some of the ways each of the ingredients in a behavioural observation study can reduce the validity of the data by introducing systematic errors of measurement.
2.3 Systematic Error: The Subjects

2.3.1 Subject Reactivity

The objective of behavioural observation is to collect data on the way people behave. The data should reflect the 'typical' behaviour of the subjects in 'natural' settings. However, placing an observer in the situation means by definition that the situation is no longer 'natural' in that it is no longer the exact setting in which the behaviour under investigation normally occurs. The fact that the situation is no longer exactly the same is not a criticism in itself. However, if the placing of an observer in a situation affects the way the subjects behave, that is if their behaviour in the presence of an observer is not 'typical' of their behaviour, then the data collected does not relate to the research aim and may lead to erroneous conclusions.

The crux of the question of subject reactivity is whether the presence of an observer, or more precisely the subject's awareness that he is being observed, will lead to the subject behaving differently from the way he would behave if he were not aware that he was being observed. If in fact the knowledge that he is being observed changes the person's behaviour, then the
conclusions drawn from the study cannot be extended to people in general, but remain specific to people who are aware that they are being observed.

The classic study showing the effects of observation on subject behaviour, discussed by Homans (1958), was conducted at the Hawthorne plant of the Western Electric Company, and has given the label Hawthorne Effect to this phenomenon. The study discussed by Homans was designed to assess the relationship between worker productivity and working conditions. A small group of employees was selected to work in a separate room under systematically varied working conditions. The results showed that almost every change in working conditions increased productivity, and furthermore productivity continued to increase even when the experimental alterations in conditions stopped. Follow-up interviews with the subjects suggested that their awareness of being involved in an experiment and being observed influenced their behaviour. The influence here was so great that it swamped the effects of the experimental manipulations, thus rendering the observations useless as estimates of the subjects' behaviour under 'normal' conditions.
Achenbach (1978) notes that a similar phenomenon has been observed in studies of the effects of drugs. In these studies subjects are typically randomly assigned to try separate drugs, with some subjects given a substance that has no physical effects (known as a placebo). Often the subjects who receive the placebo report that they feel better. Since the placebo has no physical effects, the change reported by the subject is obviously due to the subject being aware that he is part of an experiment and is being observed.

Wildman and Erickson (1977) cite a number of studies in the area of subject reactivity, investigating such aspects as the effect of the conspicuousness of observers, the characteristics of the observer, and the use of inconspicuous recording devices. These studies have to date produced mixed results, with no clear weight of evidence to either support or deny the possibility that subject behaviour changes when being observed.

Similarly researchers examining the possibility that subjects will get used to (habituate to) being observed over time and that thus reactivity may only be a problem in the initial stages, have reported equally unclear results. Wildman and Erickson (1977) conclude
that "We cannot assume that behaviour during observation, even after it has stabilized, represents 'typical' behaviour" (p. 271).

Kent and Foster (1977) state rather more strongly their feelings on the effect of observers on subject behaviour: "There seems little reason to doubt that the presence of an observer may, in fact, affect the behaviour of those he observes" (p. 289).

However Kent and Foster (1977) are less sure of the type of effect and degree of influence the observer may have: "The number of factors determining the magnitude and direction of behaviour change may be so great that manifest reactivity is scattered and almost completely unpredictable" (p. 289).

The issue of subject reactivity is still largely unresolved. Perhaps as Kent and Foster (1977) suggest, there are too many factors at work in this phenomenon to allow researchers to adequately explore the area, thus implying that a broader and more general understanding may not be possible. Perhaps on the other hand, researchers armed with complex multivariate techniques which have become practical with the advent of large
computers and packaged programs, may be able to design studies encompassing the multiplicity of variables at work. Whichever way the issue is resolved in the future, the most concise statement of the current situation comes from Wildman and Erickson (1977) who conclude that "The behavioural effects of being observed remain largely unknown" (p. 271).

2.3.2 Response Sets

Achenbach (1978) describes response sets as a subject's "... dispositions to respond in certain stereotyped ways regardless of what they are asked to respond to" (p. 182). It is immediately apparent that observations of this type of behaviour will not accurately describe the response of a subject to a particular stimulus or situation, since the subject's response is made without reference to the situation. In other words response sets are a biased set of behaviours which may therefore lead to erroneous conclusions.

Achenbach (1978) identifies several different types of response sets which are common in relation to self-report measures. These include the social desirability set (the tendency to give responses which conform to conventional standards of desirable behaviour), yeasaying
and naysaying (the tendency to say yes or no consistently regardless of the content of the questions), and individual tendencies to use either the extreme or the intermediate categories only on multiple response items.

For an experimenter who is interested in noting observed behaviours, these factors are even more likely to present problems than they are for the experimenter using self-report scales. For the latter, the data can be examined to determine the presence or absence of response sets, and it is possible to design instruments to control the effects of this phenomenon. The observer who is collecting data of a purely observational nature may be unaware that such factors are at work, or at the very least he may be aware of the factors but unable to determine at which time these factors are influencing observed behaviour.

Subjects may be acting in a manner which they feel conforms with desirable behaviour (either deliberately or subconsciously), or they may in fact be behaving quite 'naturally' in a socially desirable manner. The behaviour observed in both instances is the same. However without manipulating the situation, the experimenter is unable to determine whether there is a
bias in the behaviours observed. Similarly subjects may in fact agree or disagree with questions put to them, or they may be responding with a yeasaying or naysaying response set. Again the observed behaviour is the same in both instances.

As with the problem of subject reactivity, the researcher again is in the situation where the observed behaviour may or may not be an artifact produced by the subject. However in this instance it is an artifact produced regardless of the situation, where subject reactivity is an artifact produced in response to being aware of being observed.

2.4 Systematic Error: The Observer

2.4.1 Researcher Expectancy

Behavioural data collected by observational methods by definition requires an observer, and it is a truism that a human observer will not be absolutely accurate in his recording of what he observes. One source of inaccuracy in the observer relates to what the observer expects to see, or more specifically which behaviours the observer expects the subject(s) to exhibit.
The observer is frequently provided with a list of operationally defined behaviours of interest to the researcher, and is given training in the use and interpretation of this observation schedule. The researcher may or may not explain to the observer the hypotheses of the study. If he does not, the observer may well generate his own hypotheses. Whether the researcher gives the observer his hypotheses implicitly or explicitly, or not at all, there is a strong possibility that the observer will have some sort of expectation of the behaviour of the subjects. The observer's expectation may well lead to systematic (and probably unconscious) biases in recording of observations.

Researcher expectancy is referred to as expectation biases by Achenbach (1978), who credits Robert Rosenthal with the initial work in the early 1960's on the effects of observer expectation on observations. Rosenthal and his associates demonstrated that informing observers of the expected findings resulted in biased observational recordings, with these recordings biased towards supporting the hypotheses of the study. In simple terms, he noted that observers recorded more of or higher levels of the behaviours they expected to see. The reason for
this is not necessarily that observers deliberately change or screen their observation data. Wildman and Erickson (1977) point out that "The informed observer might either have been biased or more sensitive to the relevant behaviours than the uninformed observers were" (p. 262).

A number of studies conducted in the area during the 1970's (Kent (1972), Kent, O'Leary, Diament and Dietz (1974), Kass and O'Leary (1970), and Skindrud (1972)) have shed little light on the phenomenon, with some researchers finding an effect which they ascribe to the experimenter giving specific expectations to the observers, and others using the same general design failing to find such an effect.

More recently researchers have attempted to link several factors together, rather than simply looking at the effect of observer expectancy alone. Kent and Foster (1977) posit that "It is possible that the knowledge of predicted results may combine with other factors in field experimental settings to produce biased results" (p. 284).
A study by O'Leary, Kent and Kanowitz (1975) combined the previous approach of giving separate groups of observers different expectations in relation to the same observation period with experimenter feedback to the observers after each observation session. When an observational recording was made in line with the expectation given initially to the observer, the experimenter provided positive feedback. When an observer's data sheet failed to reflect the expectation given, the experimenter would give a negative reaction, such as "You don't seem to be picking up the treatment effect of X". The analysis of the observers' data showed a significant effect due to the combination of experimenter feedback and the provision of expected results to the observer.

Harris and Ciminero (1978) designed an ingenious experiment to investigate a similar type of interactive hypothesis. In this study, the observer was given an expectation of the behaviour by the experimenter, and this expectation was reinforced indirectly through the behaviour of the 'subjects'. Observers were asked to record the frequency of eye contact and face touching in a videotape of an interview between a psychologist and his patient. The videotape was made specifically for
this experiment, with the psychologist reinforcing the patient for eye contact. In the tape the psychologist also provided reinforcement at random times. Face touching and eye contact occurred with equal frequency, and the observers were told that reinforcement would increase the frequency of eye contact, but were given no expectation about the frequency of face touching.

The analysis showed that the observed number of eye contacts increased when the observers were given the expectancy that this would happen. The authors suggest from this that observers will note more occurrences of a behaviour if they are given the expectation that the behaviour will increase in conjunction with a situation in which the attention of the observer is drawn to the behaviour (eye contact) by another behaviour (face touching) in the observation period. The authors generalize that perhaps extraneous cues in the situation affect observer's recordings.

Studies in the area of researcher expectancy have been conducted under controlled experimental conditions where the researcher knew the true level of the behaviour being observed, and was in a position to use this information as a yardstick against which to assess the
observational data. In natural settings the true level of the behaviour is not known, and the experimenter must rely on the observational data as a substitute for the 'true level' of the behaviour. Research into the interactive effects of expectation and other factors suggests that treating observational data as "true" data may lead to erroneous conclusions, since the researcher will usually not be aware of what factors are impinging on his observers, let alone what effect these may have. As Kent and Foster (1977) point out, "It is clear that circumstances conducive to shaping of behavioural recordings may actually occur in a variety of applied settings" (p. 285). It is equally clear that the researcher will often be oblivious to this phenomenon, or at the very least unaware of its effects.

2.4.2 Knowledge of Reliability Assessment

In almost all observational studies the researcher will attempt to estimate how reliably or consistently his observers are recording the behaviours of interest. The assumption is that if observers are consistent then they are accurately recording the behaviours of interest (although it should be pointed out immediately that this assumption is fallacious).
Kent and Foster (1977) and Wildman and Erickson (1977) both credit Reid (1970) with first investigating the notion that observers may record more reliably when they are aware that their recordings are being assessed for reliability. Reid trained a group of observers in the application of a particular observation code to video-tapes of mother-child interactions and told the observers that during the training period their recordings were being compared to a criterion sheet. After the training period the observers were told that they would be viewing unscored videotapes and that they would be the only ones viewing these tapes. In reality criterion sheets had been prepared for all tapes (both the training session and the 'real' tapes) and observer reliability was calculated for each tape. Reid reports that there was an average decrease of 25% in observer agreement with the criterion when observers were not aware that the reliability of their observations was being assessed.

The findings reported by Reid (1970) were supported in a study by Romanczyk, Kent, Diament and O'Leary (1973) who assessed reliabilities overtly and covertly within each session, and again found that reliability was lower during covert than during overt assessment. The authors
also reported that the observers recorded 25% less behaviour during the covert assessments compared with the overt assessments.

Taplin and Reid (1973) conducted a study to investigate the effect of varying the information given to the observers about assessment of their recordings. After the initial training period one group of observers was told that its reliabilities would not be checked, one group was told that its reliabilities would be checked randomly without their knowledge, and one group was told that its reliabilities would be checked occasionally but only with advance warning that this would occur. In fact reliability was assessed constantly using a criterion protocol as a reference point. The authors report an immediate drop in reliability after the overt reliability in the training sessions for all groups. Of even more importance perhaps was the finding that the reliability of the third group was significantly higher during overt than during covert assessment. In other words the reliability of this group varied up and down as a function of knowledge of reliability assessment.

The finding reported by Taplin and Reid (1973) is a classic example of what Campbell and Stanley (1963) refer
to as reactive arrangements which in this case refers to
the observer's "... knowledge that he is participating
in an experiment" (p. 20) and is another manifestation of
what was referred to earlier in this paper as the
Hawthorne Effect.

2.4.3 Observer Drift

A phenomenon of observational data collection which
has grave implications for researchers assessing the
change of behaviour over time is the tendency for
agreement between raters who work together to be higher
than between raters who do not work together. Several
studies have shown that when raters are trained as a
group initially and then broken into smaller groups or
pairs to carry out data collection, the agreement between
the members of the smaller groups is higher than the
agreement between observers of different groups.

Johnson and Bolstad (1973) described an unpublished
study by DeMaster and Reid in which the observers were
divided into 14 pairs and assigned to three separate
feedback conditions for discussing the accuracy of their
observations. DeMaster and Reid found that the within-
pair reliability scores were higher than reliability
scores between observers not in the same pair and were,
also, higher than agreement scores calculated with the protocol (that is, the 'true' behaviour pattern).

Kent, O'Leary, Diament and Dietz (1974) also report finding reliability measures were higher within pairs of observers than between pairs of observers, and further reported that the differences between pairs of observers produced significantly different recordings in several of the behaviour categories being observed.

Phenomena such as observer drift were recognized by Campbell and Stanley (1963), who refer to such phenomena as instrument decay, the general term meaning that the calibration of an instrument (in this case observers) may change over time, thus producing spurious results. Kent and Foster (1977) conclude that "reliability estimates calculated from the data of observers who have trained and rated as a team provide inadequate and inflated measures" (p. 302). Furthermore, as the study by Kent et al. (1974) demonstrates, the data recorded by such observers may be quite inaccurate. The unwary researcher would be faced with apparently high reliability estimates which could conceal significantly different recordings of the behaviour observed.
2.5 Systematic Error: Design and Analysis

2.5.1 Calculation of Reliability Measures

As applied to behavioural observation data, the notion of reliability is essentially a question of consistency; the researcher is interested in how consistently the observer scores the behaviour. Irwin and Bushnell (1980) state that "Reliability refers to the extent of observer agreement or consistency in recording observational information" (p. 182). Irwin and Bushnell go on to infer that high reliability is synonymous with high accuracy. In this context accuracy is defined as the degree to which what is recorded by the observer is congruent with what actually occurred. Kerlinger (1973) discusses reliability in terms of error of measurement, stating that "Reliability can be defined as the relative absence of errors of measurement in a measuring instrument", and further that "Errors of measurement are random errors" (p.443). The preceding discussion of the types of error introduced by the observer showed clear empirical evidence that agreement between observers (that is consistency or reliability) could be influenced by a number of factors and that high inter-observer agreement does not necessarily indicate accurate recording of observations. In Kerlinger's approach this amounts to saying that measurement errors due to observers are not
always random. Thus, researchers reporting high reliability (either between observers or with the same observer across a number of observations) have not necessarily demonstrated accuracy of data recording. Reliability is a prerequisite for, but not a sufficient condition for the obtaining of accurate data.

Aside from the issue of interpreting the meaning of reliability measures in terms of accuracy, which is in itself a serious problem, researchers reporting observer reliability have used a wide variety of computational methods, leading to a wide variety of incompatible (but often similar-sized) figures which are actually measuring different things (that is the 'reliability' figures have different meanings).

The simplest measure of reliability is that which examines the "... degree of agreement between two or more people in their scoring of a variable" (Achenbach, 1978, p. 77). It seems intuitively obvious that if two observers independently recording the same incident agree strongly in what they recorded, then the recordings are reliable indicators of what actually transpired.
Calculation of agreement reliability is a simple matter. The number of intervals during which both observers agree that the target behaviour occurred or did not occur are summed, divided by the total number of observation intervals, and converted to a percentage figure. Agreement reliability, then, is the percentage of intervals in which observers agreed in what they recorded.

The calculation of various reliability figures can best be presented through a numerical example. This example is based on a fictitious study which divided the observation period into 100 observation intervals, and used two observers to record for each interval whether the target behaviour did or did not occur. Table 1 below presents the data from these two observers. Each cell in the table has been labelled with a letter from A to D to permit reference in the following discussion to particular cells in the table.
TABLE 1
Hypothetical Data showing Observer Reliability

<table>
<thead>
<tr>
<th>OBSERVER NUMBER 1</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Occurrence</td>
<td>(A)</td>
<td>(C)</td>
</tr>
<tr>
<td>Non-occurrence</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

OBSERVER NUMBER 2

<table>
<thead>
<tr>
<th>Occurrence</th>
<th>(B)</th>
<th>(D)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Non-occurrence</td>
<td>8</td>
<td>82</td>
</tr>
</tbody>
</table>

In this example each observer recorded 10 occurrences of the target behaviour, but only two of these occurrences were recorded in the same interval by both observers. Agreement reliability between these observers is calculated by summing cells A and D, dividing this by the sum of all cells (i.e. the total number of observational intervals), and multiplying by 100 to produce a percentage figure. For the data in Table 1 this figure evaluates to 84%.

This numerical example immediately highlights a major weakness in this approach; the observers only agreed on the occurrence of the behaviour in 2 intervals out of 100, and disagreed on the occurrence in 16 instances out of 100. Yet the reliability estimate given by this procedure is a rather high 84%. Most of the
cause for this high reliability figure is the agreement between the observers on the non-occurrence of the behaviour. As Kent and Foster (1977) point out, if the focus of the study is to examine the frequency with which the behaviour occurs then "... high reliability in recording the absence of that behaviour is not particularly relevant" (p. 310).

This line of thought gave rise to a measure of reliability based only on the occurrence of the behaviour, and hence is termed occurrence reliability. Occurrence reliability is based only on the instances where at least one observer recorded the behaviour, and is computed as $A / (A + B + C)$ converted to a percentage. For the data in Table 1 the occurrence reliability figure would be calculated to be 11%. This figure contrasts starkly with the 84% agreement reliability computed from the same data. An analogy to occurrence reliability where non-occurrence of the behaviour is the primary interest is computed as $D / (B + C + D)$ expressed as a percentage, which on the present data would evaluate to 83.7%.

One solution proposed to the problem of disparate reliability estimates is to report both the occurrence
and the non-occurrence reliability (Hawkins and Dotson, 1975). However this suggestion results in unnecessary confusion and leaves the reader to wonder which figure should be given more weight in assessing the 'true' reliability of the observational data.

A further weakness with methods of estimating reliability lies in the lack of criteria for establishing acceptable levels of observer agreement (Wildman and Erickson, 1977). Harris and Lahey (1978) suggest that 80% is conventionally used as a level for acceptable agreement. However they point out that this level is purely arbitrary.

Kent and Foster (1977) point out that all of the commonly used reliability measures presented above are "... inflated to some extent by chance agreements among observers on the occurrence and/or nonoccurrence of behaviours" (p. 310). Chance agreement on occurrence is calculated as \( \frac{(A + B)(A + C)}{(A + B + C + D)} \) which for the data in Table 1 is 1 agreement (or 1% agreement expected by chance). Similarly the chance agreement on non-occurrence is calculated by \( \frac{(B + D)(C + D)}{(A + B + C + D)} \) which on the present data is 81 agreements or 81% agreement expected by chance alone.
Thus the combined chance agreement for occurrences and non-occurrences on this data is 82%. Comparison of the 84% agreement actually achieved by the observers in this example shows that while the reliability figure reported may be numerically large, and above the conventional 80% acceptability level, it may in fact not be very far different from the agreement which would be obtained by chance (in this case only 2% better).

Kent and Foster (1977) further point out that the level of chance agreement (in agreement reliability) is lowest when both observers each record the target behaviour half the time, and increases as recording frequency deviates from 50%. In contrast occurrence reliability measures are increasingly inflated as the frequency of occurrence rises, and non-occurrence measures are increasingly inflated as frequency of occurrence falls. Thus, reliability measures for different experimental conditions, where the frequency of recorded behaviour changes, are not directly comparable. Kent and Foster suggest the use of a formula for correction for chance agreement which they attribute to Cohen (1960).
A number of authors recently have proposed alternatives to the generally used agreement reliability measures. Some of these (for example Hopkins and Herman, 1977) have proposed methods based on the comparison of the observed agreements to the agreement level expected by chance. Comparison of observed agreement level to the agreement level expected by chance is extremely similar to the notion of chi-square, but does not yield figures which can be interpreted consistently as indicating the same level of reliability.

An alternative approach to the problem was proposed by Hartmann (1977), who bases his method on a correlation-type approach and derives a figure in the range of -1.0 to +1.0 as the estimate of reliability. Again this figure is not readily interpretable to the lay reader.

More recently a strong case was made by Harris and Lahey (1978) for a method which combines the agreement information for occurrences and non-occurrences and weights the reliability estimate for the frequency of these as well as adjusting for chance agreements. Whilst the technique proposed by Harris and Lahey seems to cover the theoretical objections raised against other methods,
it does not provide a figure comparable to those generally reported. The method applied to the data presented here would yield a reliability estimate of 9.4%.

It is clear from the preceding discussion that the method of calculating observer reliability exerts a profound effect on the figure which the researcher reports. With such a plethora of different methods currently in use, the reader of research articles is left in the position of having virtually no idea of the "true" reliability of the data. Progress towards resolution or standardization in reporting the reliability of observational data can best be demonstrated by contrasting the position almost twenty-five years ago with the position less than five years ago. In 1960 Wright stated that "Reported measures of observer agreement in different studies do not clearly favour any principal method over any other" (p. 119). In reviewing recent work in the area Harris and Lahey (1978) conclude that "No current method of calculating interobserver agreement has been widely accepted" (p. 523).
2.5.2 Design Considerations in Time Sampling

As noted earlier, time sampling is a term covering a broad range of techniques whose common aim is to quantify observed behaviour in a sample of time during which the behaviour occurs. The technique derives from the need for efficient and parsimonious data collection and its objective is to produce maximum accuracy with minimum data. Time sampling is analogous to sampling subjects from a universe in order to conduct a study which can be generalized to the universe sampled. In time sampling the universe is the continuum of time and the "subjects" are small units of time. Typically the observation session is divided into a number of time intervals in which data on specified behaviours of interest is collected. The data collected in each interval are typically whether the behaviour occurred in that interval. Time intervals may be contiguous or may be separated by time gaps during which recording of observed behaviour is made.

The first consideration in the design of a time sampling study is the rule for deciding whether to record an occurrence of the behaviour of interest in a particular interval. Powell, Martindale and Kulp (1975) distinguish between whole interval time sampling where
the behaviour is recorded as occurring if it occurs throughout the whole interval, partial interval time sampling where the behaviour is recorded as occurring if it occurs at any time during the interval, and momentary time sampling where the behaviour is recorded as occurring if it occurs at the end point of the interval.

The second design consideration is the length of the time interval and the length of the gap between time intervals to be used for recording observations. Decisions on the length of time interval and the length of gap between intervals are typically based on the researcher's knowledge of the pattern of the behaviour of interest, or on decisions made by researchers in previous studies and rules of thumb.

In 1939, a rule of thumb was proposed for the appropriate length of the time interval (Arrington cited in Irwin and Bushnell, 1980). The suggestion made was that the time interval should be approximately the length of a single instance of behaviour. The spacing of the time intervals should be based on the length of the time interval selected, the number of subjects to be observed within the interval, and the amount of detail to be recorded. On the latter point, Hutt and Hutt (1970)
suggested that observers collecting data on three or more categories of behaviour require ten seconds to one minute to record the data.

Researchers using time sampling techniques aim to infer the true pattern of the behaviour of interest from the data on the observed pattern of behaviour. Thus, the validity of the findings of such studies rests initially on the degree of accuracy with which the observed data represents the true behaviour pattern. A number of researchers recently have examined the relationship between the true behaviour pattern and the estimates given by time sampling using various combinations of values for the design parameters.

Powell, Martindale and Kulp (1975) conducted a study which aimed to examine the bias of different time sampling procedures for estimating proportion of occurrence of behaviour. Powell et al. (1975) identified three types of time sampling procedures in common usage. These procedures differ in the rules they use for recording behaviour. Powell, Martindale and Kulp proposed the terms (1) whole interval time sampling, (2) partial interval time sampling, and (3) momentary time sampling be used when, to be scored as an interval in
which the behaviour occurred, the behaviour must occur throughout the whole interval, in any part of the interval, or at the end point of the interval respectively. They compared these three different types of time sampling procedures to a criterion of continuous observation to examine the degree of bias produced by each procedure in estimating proportion of occurrence of behaviour. Comparison of time sampling procedures to the criterion was made with different lengths of observational intervals (varying between ten seconds and five minutes). They conclude that both whole interval and partial interval time sampling lost accuracy as the length of the intervals increased. They showed that the whole interval time sampling under-estimated and the partial interval time sampling over-estimated the behaviour increasingly as the length of the time interval was increased. Momentary time sampling showed no systematic bias in estimating proportion of occurrence of behaviour.

A similar study by Green and Alverson (1978), based on computer-generated data which varied the mean length and the variance of duration of behaviour and time between occurrences of behaviour, compared the estimates obtained by whole interval, partial interval, and
momentary time sampling (as did Powell, Martindale and Kulp, 1975, although the Powell et al. study was based on only one set of data). Green and Alverson substantiate the findings of Powell et al. that both whole interval and partial interval methods produced biased estimates. Green and Alverson extended this finding to suggest that the degree of bias was a function of the ratio of interval length to the sum of the mean duration of behaviour and the mean time between behaviours.

Murphy and Goodall (1980) compared the accuracy of whole interval, partial interval, and momentary time sampling for estimating the proportion of occurrence of behaviours of differing frequency and mean duration. Murphy and Goodall's study was based on a series of five-minute videotapes of nine profoundly retarded children which were analysed by continuous recording to provide a criterion or 'true' proportion of behaviour. Frequency of behaviour varied from 2 occurrences to 33 occurrences during the recording session. Mean duration of behaviour varied from 0.3 seconds to 123 seconds. Murphy and Goodall report that whole interval time sampling consistently underestimated and partial interval time sampling consistently overestimated the true proportion of occurrence of behaviour. Momentary time sampling
provided consistently more accurate estimates than the other two methods. Murphy and Goodall suggest that whole interval and partial interval methods may provide accurate estimates if the minimum duration of behaviour is much larger than the observation interval, but unless this can be guaranteed momentary time sampling should be used.

Green, McCoy, Burns and Smith (1982) conducted an ingenious study on the bias of various time sampling procedures for estimating proportion of occurrence of the behaviour. Green et al. distinguished between bias due to observers and bias due to the method of time sampling. The study involved observers using the three different time sampling procedures for viewing a series of videotaped sessions of a subject who deliberately manipulated her behaviour to provide differing proportions of time for which the behaviour occurred for the different observation sessions. The researchers compared the degree of bias across whole interval, partial interval and momentary time sampling which could be ascribed to the observer and also the degree of bias which could be ascribed to the time sampling procedure employed. Green et al. found that partial interval time sampling overestimates and whole interval time sampling
underestimates true proportion of occurrence, while momentary time sampling does not produce biased estimates. Green et al. further found that the degree of bias was primarily due to the time sampling procedure rather than to the observers.

Repp, Roberts, Slack, Repp and Berkler (1976) mechanically generated records of behaviour patterns which differed in frequency (every 6 seconds, every 60 seconds, and every 5 minutes on average) and differed in pattern (constant across the observation session, or in a burst where 75% of the behaviour occurred in the first half of the observation session). Repp et al. then used partial interval time sampling to estimate proportion of occurrence of the behaviour and frequency of occurrence of the behaviour. In relation to proportion of occurrence they conclude that partial interval time sampling produces a bias in data collected which is related to the true rate of occurrence of the behaviour. In relation to estimates of frequency of occurrence they report that partial interval time sampling underestimates the true frequency of occurrence. The study by Repp et al. implies that underestimation of frequency of occurrence by partial interval time sampling is related to the length of an instance of the behaviour and the
length of the time between occurrences of the behaviour. Repp et al. recommend that continuous observation counting the number of occurrences of the behaviour be used if a measure of frequency of occurrence is required. Repp et al. conclude that there is an interaction between the true pattern of behaviour and the choice of length of time interval and gap between intervals which affects the observational data collected.

Researchers examining the issue of degree of bias of partial interval, whole interval and momentary time sampling for estimating proportion of occurrence have reported remarkably consistent results. The general finding is that, relative to the true proportion of occurrence, partial interval overestimates, whole interval underestimates, and momentary time sampling shows no appreciable bias. Ary (1984) has provided an elegant mathematical explanation for the consistency of findings in this area. Ary demonstrated mathematically the extent of the overestimation and underestimation to be expected through partial interval and whole interval time sampling respectively. Ary further demonstrated mathematically that the expected mean of sampling errors for momentary time sampling is zero, thus indicating that momentary time sampling will produce an unbiased estimate.
of true proportion of occurrence. Ary was further able to demonstrate, by mathematically arriving at an estimate of the standard error of estimate, that sampling error using momentary time sampling cannot exceed error under either whole or partial interval time sampling. Thus, not only does momentary time sampling produce an unbiased estimate of true proportion, it will also be more accurate than other methods.

Powell and Rockinson (1978) examined the accuracy of partial interval time sampling as a method for estimating frequency of occurrence of behaviour. Their study was based on 11 sets of data generated by computer to simulate behaviour of varying frequencies and durations. Powell and Rockinson analysed these data using partial interval time sampling with intervals of 5, 10, 20, 30, 60, and 120 seconds, comparing the estimates obtained to the true value of the data sets. Powell and Rockinson's results suggest that partial interval time sampling underestimated frequency of occurrence under most combinations of frequency and duration of the true behaviour pattern. Powell and Rockinson also report that the accuracy of the estimates increased, peaked, and then declined as the length of the time interval increased. Powell and Rockinson conclude that there are many
situations in which partial interval time sampling cannot yield valid measurements (by which they mean estimates) of frequency of occurrence of behaviour.

Tyler (1979) conducted two studies in which he compared the accuracy of partial interval and momentary time sampling for estimating frequency of, and proportion of time spent on the target behaviour. The first study was based on observation of six autistic children for 20 sessions of ten minutes each. Accuracy of the time sampling method was determined by reference to a criterion produced by continuous observation. Tyler reported that in this first study both time sampling methods produced poor estimates of frequency, and that partial interval time sampling overestimated that true proportion of occurrence while momentary time sampling provided an accurate estimate of proportion of occurrence. The second study was based on a large number of computer generated 'true' behaviour patterns which were again sampled (in this case by a computer program) using partial interval and momentary time sampling. Tyler reported that in this second study both time sampling methods produced acceptable accuracy when the time interval was short and also when the true behaviour occurred less than 10% of the time. Again in estimating
proportion of occurrence Tyler reported that partial interval time sampling produced overestimates and momentary time sampling produced acceptable accuracy.

From the foregoing it is apparent that length of time interval for observation, length of gap between time intervals for recording, type of time sampling, and the true pattern of the behaviour interact to produce substantially different results. Unfortunately, as Green and Alverson (1978) note, the researcher usually does not know the true pattern of behaviour (since this is usually the question under study) and hence is not in a position to make the appropriate choices of design factors.

2.6 Scope and Purpose of the Present Study

2.6.1 Scope of the Study

The preceding discussion provided an overview of direct observation techniques. Several authors (Johnson and Bolstad, 1973, Jones, Reid and Patterson, 1975) have suggested that the development of direct observation techniques may well be the single most important contribution to the discipline of psychology. Sanson-Fisher, Poole, and Dunn (1980) point out that such techniques are widely used by behavioural scientists. Direct observation techniques were differentiated on the
basis of the type of data collected - qualitative or quantitative. Quantitative methods include event sampling, time sampling and trait rating. Reviewing studies published in the Journal of Applied Behaviour Analysis between 1968 and 1975, Kelly (1977) found that 76% of the studies employed direct observation procedures. Of these, 41% employed some form of time sampling. Phinney (1982) suggests that time sampling is the technique generally used for observational research with children. It seems therefore that time sampling is one of the major data collection methods used in behavioural studies. The present study will be limited to an examination of time sampling as a quantitative method for behavioural observation.

Error in behavioural observation was classified as either random error or systematic error. Random error reduces the reliability of observational data but does not produce a bias in the data. Systematic error, however, tends to lean in one direction, thus producing a bias in the observation data. This study will exclude consideration of random error and will focus only on systematic error in behavioural observation.
Potential bias in time sampling studies was classified into three major sources. The first two sources are the observers and subjects, who can introduce bias into the observational data in a number of ways. Since a number of studies have been published on the various ways in which the observers and the subjects can introduce bias, this area now seems to be fairly well understood.

The third major source of bias relates to the time sampling procedures themselves. Recently some researchers have conducted studies in this area, and have suggested factors with the potential to introduce bias into the observed data as a direct result of the use of time sampling techniques. Bias produced by the time sampling procedures themselves is the focus of the present study.

Recent studies have suggested that bias due to time sampling techniques themselves is associated with two separate sets of factors. The first set of factors covers the decisions made by the researcher in designing the study, such as (1) the length of time interval to be used, (2) the length of the gaps between the intervals (usually provided to allow the observer time to record
the observations), and (3) the criterion used to decide whether or not to record an occurrence of the behaviour for a given time interval.

The second set of factors associated with bias in observational data recording relates to the true pattern of behaviour. Factors included in this category are (1) the mean duration of the behaviour of interest, (2) the rate at which the behaviour occurs, and (3) the variability of the true behaviour pattern. Since researchers do not have control over the pattern of behaviour of their subjects, it follows that the set of biasing factors which the researcher can control is the set relating to the design of the study. Consequently this study will focus on design factors as a source of bias in time sampling studies.

2.6.2 Aim of the Study
Several researchers (Powell, Martindale and Kulp, 1975, Green and Alverson, 1978) have examined the effect of design factors on the accuracy of time sampling for estimating proportion of occurrence of behaviour. The general finding of such studies is that partial interval time sampling overestimates and whole interval time sampling underestimates the proportion of occurrence,
while momentary time sampling shows no bias in estimating proportion of occurrence.

Research into the effect of design factors on the accuracy of time sampling for estimating frequency of occurrence of behaviour has been less conclusive. Researchers in this area have been unable to discover a relationship between design factors and bias, and have concluded that time sampling is not a suitable technique for estimating frequency of occurrence of behaviour.

In summary, research in the area to date suggests that momentary time sampling produces unbiased estimates of proportion of occurrence, and that time sampling procedures are not appropriate for estimating frequency of occurrence. McDowell (1973) recommends against the use of time sampling procedures because of the problem of estimating frequency of occurrence through time sampling and suggests the use of continuous recording as the method providing the greatest versatility. In addition to versatility, continuous observation will yield inherently unbiased estimates of proportion of occurrence and frequency of occurrence since there is no sampling and therefore no sampling error involved. However this
method is extremely labour-intensive and loses several of the advantages of time sampling.

The aim of the present study is to investigate the effect of design factors and true behaviour pattern on the accuracy of time sampling for estimating frequency of occurrence of behaviour. The design factors to be considered are the type of time sampling (whole interval, partial interval, and momentary), the length of the time interval, and the length of the gap between time intervals. If it is possible to establish a relationship between design factors and accuracy, it will then be possible to utilize this relationship to select values of the design factors in order to maximise accuracy. Producing a satisfactory level of accuracy of time sampling for estimating frequency of occurrence would allow the use of time sampling for simultaneously estimating both proportion of occurrence and frequency of occurrence of behaviour, thus maintaining the previous advantages of time sampling and also adding a new dimension to the type of data which can be extracted by time sampling.
CHAPTER III

METHOD

3.1 Overview

The aim of the present study was to investigate the effect of design factors on the accuracy of time sampling for estimating frequency of occurrence of behaviour, where accuracy is the degree to which the number of occurrences of a behaviour recorded by a perfect observer matches the true number of occurrences of behaviour. This aim was achieved by examining, concurrently, the statistical effect of all factors involved in time sampling studies in a manner which provided experimental control over all factors using a Monte Carlo type study based on a large number of computer generated behaviour patterns. The computer generated behaviour patterns represent the actual behaviour of the subject (which in field studies of course is not known). The computer generated behaviour patterns were examined by a computer program which simulated time sampling procedures under varying values of the factors (length of time interval sampled, etc.) to produce estimates of the number of occurrences of behaviour. These estimates represent the data normally recorded by observers in time sampling studies. The analysis then proceeded through the use of multiple regression techniques to build a model which
accounted for the effect of design factors on the accuracy of the estimates of frequency of occurrence of behaviour.

3.2 Independent Variables

Two sets of independent variables were considered in this study. The first set of variables related to the pattern of the true behaviour. The second set of variables related to the design decisions made by the researcher in setting up a study using Time Sampling.

3.2.1 Independent Variables: True Behaviour Pattern

True behaviour pattern consists of a number of occurrences of the target behaviour spread across time and interspersed with a number of non-occurrences of the target behaviour. Five variables were used to describe true behaviour pattern. True behaviour pattern can vary in the duration of each occurrence of the behaviour. That is, across a period of time the behaviour may occur, for example, five times, with durations of 5, 8, 13, 20 and 21 seconds. Duration of occurrence of the true behaviour can be described in terms of its mean and standard deviation. Similarly the periods of time when the behaviour does not occur may vary in duration, and this non-occurrence of behaviour can similarly be
described in terms of its mean duration and its standard deviation.

The final variable used in describing the true behaviour pattern is the number of occurrences of the behaviour. Thus, a complete description of a true behaviour would be given by the mean and standard deviation of the duration of occurrence, the mean and standard deviation of the duration of non-occurrence, and the number of occurrences of the behaviour of interest. The number of non-occurrences of behaviour in the general case will be in the range of the number of occurrences of behaviour plus or minus one. Throughout this study it will be assumed that the number of occurrences of non-behaviour is equal to the number of occurrences of behaviour. This assumption will simplify much of the algebraic manipulation which follows, but will not affect the logic of the model developed.
The following symbols will be used to refer to these variables:

\[ X_b = \text{Mean Duration of Behaviour} \]
\[ X_g = \text{Mean Duration of Non-behaviour} \]
\[ s_b = \text{Standard deviation of Duration of Behaviour} \]
\[ s_g = \text{Standard deviation of Duration of Non-behaviour} \]
\[ T = \text{True Number of Occurrences} \]

The five variables listed above are the "simple" or fundamental variables involved in the description of the true behaviour pattern. A number of researchers have used variables not included in the above list. Variables not included in the above list can be seen as composites of or derivatives of the variables listed above.

One variable commonly used is the proportion of time for which the behaviour occurs. Proportion of time for which the behaviour occurs can be derived from a knowledge of the mean duration of occurrence and the mean duration of non-occurrence of the behaviour.
Specifically,

Proportion of time for which behaviour occurs = \[ \frac{\sum X_b}{\sum X_b + \sum X_g} \]

= \[ \frac{TX_b}{TX_b + TX_g} \]

= \[ \frac{TX_b}{T(X_b + X_g)} \]

= \[ \frac{\bar{x}_b}{\bar{x}_b + \bar{x}_g} \]

Another variable which is commonly used is rate of occurrence of the behaviour, which can be calculated from the number of occurrences and the total duration of the observation period.

Specifically,

Rate of Occurrence = \[ \frac{T}{\sum X_b + \sum X_g} \]

= \[ \frac{T}{TX_b + TX_g} \]

= \[ \frac{T}{T(X_b + X_g)} \]

= \[ \frac{1}{\frac{X_b}{X_b + X_g}} \]
Since proportion of time for which behaviour occurs and rate of occurrence are actually composites of the five variables listed above it would be possible to use the composite variables as independent variables describing true behaviour pattern. That is, one could consider proportion of occurrence as the independent variable and manipulate proportion of occurrence to alter true behaviour pattern. The decision to select the five variables listed above as the independent variables was based on the fact that each of the five variables can be manipulated independently, without affecting the values of the other four variables. This independence is not true of the "composite" variables, since manipulation of composite variables requires simultaneous manipulation of other variables to maintain the same behaviour pattern, or the removal of constraints (such as the requirement to maintain the same behaviour pattern).

Stated another way, specifying the proportion of occurrence or rate of occurrence in effect specifies the relationship between mean duration of occurrence and mean duration of non-occurrence. However, specification of proportion of occurrence or rate of occurrence does not uniquely describe the behaviour pattern. For example, a proportion of occurrence of 0.5 (i.e. the behaviour
occurs for 50% of the time) can occur in an infinite number of different behaviour patterns in which the mean duration of occurrence is equal to mean duration of non-occurrence. Thus, specifying proportion of occurrence does not adequately describe the behaviour pattern since duration of behaviour itself is not fixed. However specifying a mean duration of behaviour of, for example, 30 seconds, and a mean duration of non-occurrence of behaviour of 30 seconds not only describes duration of occurrence of behaviour, but also specifies proportion of time for which the behaviour occurs.

3.2.2 Generating the True Behaviour Pattern

The true behaviour pattern was generated in two stages by a fortran computer program called BEHGEN which was specially written for this purpose. A listing of the BEHGEN program appears in Appendix A. The first stage of this process involved random selection of values for the five parameters of true behaviour pattern described above. The selection of values for the five parameters was made by generating five random numbers from a specified type of distribution to be used as the values of the five parameters. At each iteration of the program BEHGEN a new set of values for the parameters of "true"
behaviour was generated randomly. Each value in this set was randomly generated independently of every other value in the set. The values were generated by reference to the International Mathematical and Statistical Library (IMSL) function called GGUBFS described in IMSL Inc. (1980). Reference to function GGUBFS returns a random number drawn from a rectangular distribution with a range of 0.0 to 1.0. The value returned by function GGUBFS was then adjusted to provide values for the five parameters of true behaviour selected randomly from the following distributions:

\[
\begin{align*}
\bar{X}_b &: \text{Rectangular Distribution; } \\
&\quad \text{Range 1 - 1200 seconds} \\
\bar{X}_g &: \text{Rectangular Distribution; } \\
&\quad \text{Range 1 - 1200 seconds} \\
S_b &: \text{Rectangular Distribution; } \\
&\quad \text{Range 0 - 1200 seconds} \\
S_g &: \text{Rectangular Distribution; } \\
&\quad \text{Range 0 - 1200 seconds} \\
T &: \text{Rectangular Distribution; } \\
&\quad \text{Range 1 - 30 Occurrences}
\end{align*}
\]

The range of values for mean duration of behaviour and non-behaviour were selected to provide behaviour patterns varying from very short duration behaviour occurring for a very low proportion of time (where \(X_b\) was 1 second and \(X_g\) was 1200 seconds, resulting in a
behaviour which occurred for less than 1% of the time) to very long duration behaviour occurring for a very high proportion of time (where X_b was 1200 seconds and X_g was 1 second, resulting in a behaviour which occurred for almost 100% of the time).

The range of values for standard deviation of duration of behaviour and non-behaviour were selected to provide behaviour patterns varying from absolutely regular (where both s_b and s_g were zero) to extremely variable (where s_b and s_g were 1200).

The range of values for number of behaviours was selected to provide a maximum observation period of 20 hours, which is longer than any real-life observation period is likely to be, thus ensuring that the majority of observation periods generated would reflect the length of observation periods normally used in practice.

In the second stage of the process the program proceeded to generate a pattern of occurrences and non-occurrences of behaviour under the randomly generated values of the parameters of "true" behaviour. The "true" behaviour pattern was stored in a vector where the odd-numbered elements contained randomly generated values of
the duration of occurrences (in seconds) and the even-numbered elements contained the values of the duration of non-occurrences. The number of elements in the true behaviour pattern vector was twice the value of \( T \), providing for \( T \) occurrences of the behaviour. The duration of each separate occurrence and non-occurrence of behaviour was randomly generated by reference to the IMSL (1980) function GGNQF. Reference to the function GGNQF returns a random normal deviate drawn from a normal distribution with a mean of 0.0 and a standard deviation of 1.0. The value returned by the function GGNQF was then adjusted to provide a distribution of occurrence durations and non-occurrence durations with the following characteristics:

Duration of occurrence:
Normal Distribution with Mean = \( X_b \) and Var = \( S_b \)
under the constraint that every value is greater than zero

Duration of non-occurrence:
Normal Distribution with Mean = \( X_q \) and Var = \( S_q \)
under the constraint that every value is greater than zero

In generating values for duration of occurrence and non-occurrence of behaviour a constraint was placed on the values so that all values were greater than zero. This constraint was necessary to ensure that the program did not generate negative numbers, which would represent
negative values for duration of occurrence and non-occurrence of behaviour. Negative values for duration of occurrence and non-occurrence of behaviour would obviously be meaningless in terms of true behaviour patterns, since it is not possible for behaviour to occur for zero seconds or less.

At this stage, the program has generated one true behaviour pattern consisting of a number of occurrences and non-occurrences of behaviour. The duration of occurrences and non-occurrences of behaviour were drawn from normal distributions with means and standard deviations selected at random. The number of occurrences was also set randomly.

After the first true behaviour pattern had been generated, the program then calculated the actual mean and standard deviation of duration of occurrence of behaviour and the actual mean and standard deviation of duration of non-occurrence of behaviour, using the values of duration of occurrence and non-occurrence of the first true behaviour pattern. Calculation of actual mean and standard deviation of duration of occurrence and non-occurrence was necessary because of the effect of the constraint placed on the generation of values of duration
to prevent negative numbers being generated. The effect of the constraint that negative numbers were not to be generated was that some values drawn from the normal distributions with means $X_b$ and $X_g$ and standard deviations $s_b$ and $s_g$ were not used as values for duration of occurrence and non-occurrence, respectively. Since values from only the side of the distribution below the mean were rejected, the effect of the constraint to reject negative numbers was to raise the actual mean and lower the actual standard deviation relative to the values for the parameters of true behaviour pattern nominally being used. Analysis was based on the actual values for mean and standard deviation of duration of occurrence and non-occurrence calculated from the true behaviour pattern.

The ranges of actual values for mean and standard deviation of duration of occurrence and non-occurrence of behaviour were subjected to a restriction similar to that placed on the ranges of the nominal values for mean and standard deviation of duration of occurrence and non-occurrence. Behaviour patterns with actual values outside the allowed range were rejected as being too extreme to represent real-life values, and were replaced by another randomly generated true behaviour pattern with
newly generated nominal values for the five parameters of true behaviour pattern. The ranges of the actual values for the five parameters of true behaviour pattern were:

\[ X_b \]: Range 1 - 1500 seconds  
\[ X_g \]: Range 1 - 1500 seconds  
\[ S_b \]: Range 0 - 1200 seconds  
\[ S_g \]: Range 0 - 1200 seconds  
\[ T \]: Range 1 - 30 Occurrences

This process of generating behaviour patterns was repeated 1000 times giving data on 1000 "true" behaviour patterns, with the duration of occurrence and of non-occurrence being normally distributed within each behaviour pattern, and with the mean and standard deviation of duration of occurrence and non-occurrence and the number of occurrences randomly selected independently for each of the 1000 behaviour patterns.

3.2.3 Independent Variables: Design Decisions

There are three variables involved in designing a time sampling study (a) the time sampling procedure to be used, (b) the time interval to be used, and (c) the gap between time intervals.
There are several different time sampling procedures described in the literature. This study will follow the distinction made by Powell, Martindale and Kulp (1975) between whole interval, partial interval and momentary time sampling. Under whole interval time sampling the observer records an occurrence of the behaviour of interest in an interval only if the behaviour occurs throughout the whole of the time interval. Under partial interval time sampling the observer records an occurrence for the interval if the behaviour occurs during any part of the interval. Under momentary time sampling a behaviour is recorded as occurring only if it is actually occurring at a specified point in the interval (usually the end of the interval). The alternative to a time sampling approach is continuous recording which can be viewed as the limiting case for the first three approaches when the interval sampled becomes infinitely small. Continuous recording records exactly the behaviour observed, but can not be used easily in practice since continuous recording requires specialized recording equipment which is expensive and obtrusive, requires specially trained observers, and can only be used to observe one subject at a time. However, continuous recording can be used in computer-simulated
studies to provide a criterion against which to assess the degree of inaccuracy in time sampling procedures.

The other two design variables relate to the length of the time interval to be sampled and the length of the gap between time intervals. The length of the time interval is expressed in seconds or minutes, and is held constant across a study, or at least across an observation period. For example, if a time interval of five seconds is chosen, then the observer records a check mark for each five-second interval in which the behaviour of interest occurred.

In many studies, the researcher provides a time gap between observation intervals in which the observer records the data. The time gap may be of any length, although it is typically short (10 to 60 seconds) and its length is usually determined by the amount of data the observer must record for each time interval. Logically this gap may be set between zero and a very large value.

Thus, in order to describe fully the time sampling strategy used in a particular study one needs to know (a) the time sampling procedure used, (b) the length of the time interval, and (c) the length of the gap between time
intervals. These three variables constitute the design of the study which is under the direct control of the researcher, and were used in this study as independent variables.

The following symbols will be used to refer to these variables:

\[ D = \text{Type of Time Sampling} \quad (1 = \text{whole interval}, \]
\[ 2 = \text{partial interval}, \]
\[ 3 = \text{momentary}) \]

\[ I = \text{Length of Time Interval Observed (in seconds)} \]
\[ G = \text{Length of Gap between I's (in seconds)} \]

3.2.4 Simulating Time Sampling Studies

The simulation of a time sampling study was also carried out in two stages by a computer program specially written for the purpose. The BEHGEN program operated on the data previously generated to represent the true behaviour pattern.

The first stage of the simulation of a time sampling study involved randomly generating the values of time interval \( I \) and gap between intervals \( G \). The values of \( I \) and \( G \) were again generated by reference to the IMSL (1980) function GGUBFS which returns a random number in the range 0.0 to 1.0 drawn from a rectangular
distribution. The values returned by the function were then adjusted to provide values for the parameters I and G which were selected randomly from the following distributions:

I : Rectangular Distribution;
    integer in range 1 - 1200 secs.

G : Rectangular Distribution;
    integer in range 0 - 60 secs.

The second stage of the simulation of a time sampling study involved the BEHGEN program simulating time sampling under each level of D (time sampling procedure) using the previously generated values of I and G to simulate a time sampling study based on the first "true" behaviour pattern. The simulation program produced as output a value corresponding to the number of time intervals in which the behaviour was observed for each time sampling procedure (i.e. each value of D). This process continued for all subsequent true behaviour patterns with a new set of values for I and G generated for each true behaviour pattern.

Thus, after running the BEHGEN program an output file was created containing 1000 records of data for each of the three levels of D (making a total of 3000 records). Each record contained the data relating to one
simulated time sampling study, including the values of the parameters of the true behaviour pattern (i.e. the values of $X_b$, $s_b$, $X_g$, $s_g$, and $T$), the values of the design parameters (i.e. the values of $D$, $I$, and $G$), and the number of time intervals in which a behaviour occurred.

The number of time intervals in which the behaviour occurred is the basic outcome variable for time sampling studies, although usually this number is reported as a proportion of total time intervals. The number of time intervals scored as intervals in which the behaviour was observed is the last variable required to fully describe a time sampling study, and will be symbolized as follows:

$$O = \text{Number of intervals in which the behaviour was observed.}$$

### 3.3 The Dependent Variable

Since the data relating to true behaviour were computer-generated in this study, the number of behaviours which actually occurred (i.e. the value of $T$) was known. After the computer-generated behaviour pattern had been sampled by computer program simulating each of the time sampling procedures discussed above, there was available an estimate under each time sampling
procedure of the number of behaviours which occurred. The estimate of the number of behaviours which occurred is the number of intervals which would have been checked by a perfectly accurate observer (i.e. the value of the variable O).

The degree to which the estimate of the observed number of behaviours (i.e. the number of intervals in which the behaviour was recorded as occurring, symbolized by O) deviates from the true number of behaviours (i.e. the value of T) provided the basis for developing a variable which reflects the degree of accuracy in estimating true number of occurrences (T) of behaviour. However the absolute deviation of observed number of behaviours (O) from the true number of behaviours (T) reflects differing degrees of accuracy depending on the value of T. For example, if O differed from T by a value of 2 and there were 100 true occurrences, this would indicate a much more accurate estimate of T than if O differed from T by a value of 2 and there were 4 true occurrences. The basic measure of accuracy therefore is the difference between O and T, divided by T. However since
\[
\frac{O - T}{T} = \frac{O}{T} - \frac{T}{T}
\]

\[
= \frac{O}{T} - 1
\]

and since \((O/T) - 1\) is a linear transformation of \(O/T\), this study used the simpler form, i.e. \(O/T\), as the dependent variable.

Thus, the dependent variable in this study was the ratio of number of observed occurrences to the number of true occurrences. The rationale behind the choice of dependent variable is that it expresses the accuracy of estimating number of true occurrences (\(T\)) in a form which is comparable across time sampling studies. If there is no inaccuracy in the observed data then the ratio of number of observed occurrences of behaviour to true occurrences of behaviour will equal 1.0. To the degree that the sampling underestimates the true behaviour the ratio of number of observed occurrences to number of true occurrences will tend to zero, and to the degree that the sampling overestimates the true behaviour the ratio of number of observed occurrences to number of true occurrences will tend to infinity. A ratio of 2.0, for example, would mean that twice as many behaviours were observed as really happened. Thus the value of the
dependent variable provides an indication of the degree, and direction, of inaccuracy (i.e. whether more behaviours were observed than actually occurred or vice versa) in the data. For simplicity the dependent variable (i.e. O/T) will be referred to as the accuracy of the data, although strictly speaking this variable should be referred to as the degree of underestimation or the degree of overestimation of number of observed occurrences in relation to number of true occurrences of behaviour.

3.4 Method of Analysis

3.4.1 The Model

As noted above, the purpose of this study was to build a model which describes the degree of accuracy of time sampling procedures in terms of the parameters of the true behaviour pattern and the design decisions made by the researcher. The model was evaluated by use of multiple regression techniques. Thus the general form of the model is:

\[ Y = a + b_1X_1 + b_2X_2 + \ldots + b_8X_8 \]

where \( Y = O/T \)

and \( X_1 \) to \( X_8 \) = independent variables described above.
A model of this general form was built for each of the three time sampling procedures discussed above. It would be possible to build one model which incorporated time sampling procedure as a three-level categorical variable. However, the utility and simplicity of interpretation is enhanced by treating the three time sampling procedures in three separate models. The use of three separate models also makes comparisons with findings of other studies easier to interpret.

3.4.2 The Unit of Analysis

The final question to be addressed in this section is that of the unit of analysis. The most common use of multiple regression is in the situation where the researcher has data on a number of variables for a sample of subjects. In this instance, the subjects form the unit of analysis, and the data are the values on each of the variables in the model for each subject. In this study, the unit of analysis was the observational study. The data were the values of the parameters of true behaviour, the parameters of the time sampling design, and the observed number of occurrences of behaviour.
3.5 Summary

As noted earlier, this study used a Monte Carlo approach to generate the data. Data generation was achieved by selecting randomly the values for the five parameters of true behaviour pattern. Having selected these values, the true behaviour pattern was generated by a computer program. This true behaviour pattern was then sampled under each of three time sampling procedures using a randomly selected time interval and time gap to provide a simulation of a time sampling study. The simulation of a time sampling study was also carried out by a computer program, and yielded a value for the number of observed occurrences under each of three time sampling procedures. Thus, for each pass through the program there was generated one true behaviour pattern and one set of time sampling parameters, with three values of the number of observed occurrences, one relating to each time sampling procedure. The data generation and time sampling simulation was repeated 1000 times, simulating 1000 time sampling studies carried out under random combinations of true behaviour pattern and design parameters.

The output of this program was three sets of 1000 records of data, where each record related to the
simulated results of one time sampling study, and each of the three sets related to one of the three time sampling procedures described above. Each record contained the values of the five parameters of the true behaviour pattern (i.e. $X_b$, $s_b$, $X_g$, $s_g$ and $T$), the values of two of the time sampling design parameters (i.e. $I$ and $G$), and a coded value to represent the time sampling procedure used (i.e. the value of $D$).

The regression analysis was carried out separately on each set of data, where the unit of analysis was the data contained on one record (i.e. the summary results of one simulated time sampling study). Thus, as noted earlier, the unit of analysis was the observational study, with the analysis providing a regression model which attempted to account for the maximum standard deviation in the dependent variable (i.e. $O/T$) in terms of the independent variables across 1000 simulated observation studies.
CHAPTER IV
RESULTS: TIME SAMPLING PARAMETERS

4.1 Overview

As noted in Chapter 3 the following symbols were used to refer to the variables used in this study:

\[ \begin{align*}
\bar{x}_b &= \text{Mean Duration of Behaviour} \\
\bar{x}_g &= \text{Mean Duration of Non-behaviour} \\
s_b &= \text{Standard Dev. of Duration of Behaviour} \\
s_g &= \text{Standard Dev. of Duration of Non-behaviour} \\
T &= \text{True Number of Occurrences} \\
I &= \text{Length of Time Interval Observed (in seconds)} \\
G &= \text{Length of Gap between I's (in seconds)} \\
o &= \text{Number of intervals in which the behaviour was observed.}
\end{align*} \]

The validity of the findings of this study rests initially on the validity of the generated behaviour pattern data as a representation of real-life behaviour patterns. The generated behaviour patterns and time sampling designs were examined to ensure that (a) the distributions of values of the five parameters of true behaviour pattern (viz. \( x_b, s_b, x_g, s_g \) and \( T \)) and two design parameters (viz. \( I \) and \( G \)) were representative of
real-life situations, and (b) the values of the five parameters of true behaviour pattern (viz. \( X_b, s_b, X_g, s_g \) and \( T \)) and two design parameters (viz. \( I \) and \( G \)) across the 1000 simulated time sampling studies covered all possible combinations of these parameters so that the 1000 simulated studies represent all probable combinations of real-life behaviour pattern and time sampling design parameters.

The analyses reported in the first part of this chapter examine the nominal values of the five parameters of true behaviour pattern (viz. \( X_b, s_b, X_g, s_g \) and \( T \)) and two design parameters (viz. \( I \) and \( G \)) to ensure that (a) the values generated for each parameter were drawn from rectangular distributions, and (b) the values of each parameter were uncorrelated with the values of every other parameter. The analyses reported in the second part of this chapter examine the actual values of the five parameters of true behaviour pattern (viz. \( X_b, s_b, X_g, s_g \) and \( T \)) and two design parameters (viz. \( I \) and \( G \)) for (a) representativeness of the distributions of values of each parameter, and (b) representative coverage of all combinations of values of the parameters.
4.2 Nominal Values of Parameters

As noted in Chapter III the nominal values for the five parameters of true behaviour pattern (viz. $X_b$, $s_b$, $X_g$, $s_g$ and $T$) and two design parameters (viz. $I$ and $G$) were randomly selected from rectangular distributions for each of the 1000 simulated time sampling studies. Each value was selected independently of every other value. The analyses reported in this section examine the values generated to test that each set of values was drawn from a rectangular distribution, and that each variable is independent of every other variable.

4.2.1 Distributions of Nominal Values of Parameters

Frequency distributions of the values of the five parameters of true behaviour pattern (viz. $X_b$, $s_b$, $X_g$, $s_g$ and $T$) and two design parameters (viz. $I$ and $G$) were produced to examine the shapes of the distributions of the values generated. Tables 2 to 8 present frequency distributions for the nominal values of the five parameters of true behaviour pattern and two design parameters.
### TABLE 2
Frequency Distribution of Values of Nominal Mean Duration of Behaviour in Seconds

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 99</td>
<td>96</td>
<td>9.6</td>
</tr>
<tr>
<td>100 - 199</td>
<td>89</td>
<td>8.9</td>
</tr>
<tr>
<td>200 - 299</td>
<td>93</td>
<td>9.3</td>
</tr>
<tr>
<td>300 - 399</td>
<td>96</td>
<td>9.6</td>
</tr>
<tr>
<td>400 - 499</td>
<td>83</td>
<td>8.3</td>
</tr>
<tr>
<td>500 - 599</td>
<td>72</td>
<td>7.2</td>
</tr>
<tr>
<td>600 - 699</td>
<td>91</td>
<td>9.1</td>
</tr>
<tr>
<td>700 - 799</td>
<td>81</td>
<td>8.1</td>
</tr>
<tr>
<td>800 - 899</td>
<td>81</td>
<td>8.1</td>
</tr>
<tr>
<td>900 - 999</td>
<td>75</td>
<td>7.5</td>
</tr>
<tr>
<td>1000 - 1099</td>
<td>74</td>
<td>7.4</td>
</tr>
<tr>
<td>1100 - 1199</td>
<td>69</td>
<td>6.9</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1000</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>

### TABLE 3
Frequency Distribution of Values of Nominal Standard Deviation of Duration of Behaviour in Seconds

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 99</td>
<td>79</td>
<td>7.9</td>
</tr>
<tr>
<td>100 - 199</td>
<td>87</td>
<td>8.7</td>
</tr>
<tr>
<td>200 - 299</td>
<td>95</td>
<td>9.5</td>
</tr>
<tr>
<td>300 - 399</td>
<td>80</td>
<td>8.0</td>
</tr>
<tr>
<td>400 - 499</td>
<td>93</td>
<td>9.3</td>
</tr>
<tr>
<td>500 - 599</td>
<td>90</td>
<td>9.0</td>
</tr>
<tr>
<td>600 - 699</td>
<td>86</td>
<td>8.6</td>
</tr>
<tr>
<td>700 - 799</td>
<td>79</td>
<td>7.9</td>
</tr>
<tr>
<td>800 - 899</td>
<td>91</td>
<td>9.1</td>
</tr>
<tr>
<td>900 - 999</td>
<td>84</td>
<td>8.4</td>
</tr>
<tr>
<td>1000 - 1099</td>
<td>64</td>
<td>6.4</td>
</tr>
<tr>
<td>1100 - 1199</td>
<td>72</td>
<td>7.2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1000</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>
TABLE 4
Frequency Distribution of Values of Nominal Mean Duration of Non-Behaviour in Seconds

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 99</td>
<td>80</td>
<td>8.0</td>
</tr>
<tr>
<td>100 - 199</td>
<td>92</td>
<td>9.2</td>
</tr>
<tr>
<td>200 - 299</td>
<td>83</td>
<td>8.3</td>
</tr>
<tr>
<td>300 - 399</td>
<td>93</td>
<td>9.3</td>
</tr>
<tr>
<td>400 - 499</td>
<td>81</td>
<td>8.1</td>
</tr>
<tr>
<td>500 - 599</td>
<td>83</td>
<td>8.3</td>
</tr>
<tr>
<td>600 - 699</td>
<td>89</td>
<td>8.9</td>
</tr>
<tr>
<td>700 - 799</td>
<td>88</td>
<td>8.8</td>
</tr>
<tr>
<td>800 - 899</td>
<td>73</td>
<td>7.3</td>
</tr>
<tr>
<td>900 - 999</td>
<td>80</td>
<td>8.0</td>
</tr>
<tr>
<td>1000 - 1099</td>
<td>77</td>
<td>7.7</td>
</tr>
<tr>
<td>1100 - 1199</td>
<td>81</td>
<td>8.1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1000</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>

TABLE 5
Frequency Distribution of Values of Nominal Standard Deviation of Duration of Non-Behaviour in Seconds

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 99</td>
<td>78</td>
<td>7.8</td>
</tr>
<tr>
<td>100 - 199</td>
<td>93</td>
<td>9.3</td>
</tr>
<tr>
<td>200 - 299</td>
<td>88</td>
<td>8.8</td>
</tr>
<tr>
<td>300 - 399</td>
<td>88</td>
<td>8.8</td>
</tr>
<tr>
<td>400 - 499</td>
<td>93</td>
<td>9.3</td>
</tr>
<tr>
<td>500 - 599</td>
<td>85</td>
<td>8.5</td>
</tr>
<tr>
<td>600 - 699</td>
<td>65</td>
<td>6.5</td>
</tr>
<tr>
<td>700 - 799</td>
<td>77</td>
<td>7.7</td>
</tr>
<tr>
<td>800 - 899</td>
<td>87</td>
<td>8.7</td>
</tr>
<tr>
<td>900 - 999</td>
<td>90</td>
<td>9.0</td>
</tr>
<tr>
<td>1000 - 1099</td>
<td>83</td>
<td>8.3</td>
</tr>
<tr>
<td>1100 - 1199</td>
<td>73</td>
<td>7.3</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1000</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>
TABLE 6
Frequency Distribution of Values of Time Intervals in Seconds

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 99</td>
<td>63</td>
<td>6.3</td>
</tr>
<tr>
<td>100 - 199</td>
<td>90</td>
<td>9.0</td>
</tr>
<tr>
<td>200 - 299</td>
<td>102</td>
<td>10.2</td>
</tr>
<tr>
<td>300 - 399</td>
<td>75</td>
<td>7.5</td>
</tr>
<tr>
<td>400 - 499</td>
<td>97</td>
<td>9.7</td>
</tr>
<tr>
<td>500 - 599</td>
<td>80</td>
<td>8.0</td>
</tr>
<tr>
<td>600 - 699</td>
<td>77</td>
<td>7.7</td>
</tr>
<tr>
<td>700 - 799</td>
<td>91</td>
<td>9.1</td>
</tr>
<tr>
<td>800 - 899</td>
<td>83</td>
<td>8.3</td>
</tr>
<tr>
<td>900 - 999</td>
<td>78</td>
<td>7.8</td>
</tr>
<tr>
<td>1000 - 1099</td>
<td>83</td>
<td>8.3</td>
</tr>
<tr>
<td>1100 - 1199</td>
<td>81</td>
<td>8.1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1000</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>

TABLE 7
Frequency Distribution of Values of Gap between Time Intervals in Seconds

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 4</td>
<td>70</td>
<td>7.0</td>
</tr>
<tr>
<td>5 - 9</td>
<td>82</td>
<td>8.2</td>
</tr>
<tr>
<td>10 - 14</td>
<td>85</td>
<td>8.5</td>
</tr>
<tr>
<td>15 - 19</td>
<td>89</td>
<td>8.9</td>
</tr>
<tr>
<td>20 - 24</td>
<td>76</td>
<td>7.6</td>
</tr>
<tr>
<td>25 - 29</td>
<td>75</td>
<td>7.5</td>
</tr>
<tr>
<td>30 - 34</td>
<td>78</td>
<td>7.8</td>
</tr>
<tr>
<td>35 - 39</td>
<td>92</td>
<td>9.2</td>
</tr>
<tr>
<td>40 - 44</td>
<td>110</td>
<td>11.0</td>
</tr>
<tr>
<td>45 - 49</td>
<td>79</td>
<td>7.9</td>
</tr>
<tr>
<td>50 - 54</td>
<td>74</td>
<td>7.4</td>
</tr>
<tr>
<td>55 - 59</td>
<td>90</td>
<td>9.0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1000</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>
TABLE 8
Frequency Distribution of
Number of True Occurrences of Behaviour

<table>
<thead>
<tr>
<th>Number</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>2.5</td>
</tr>
<tr>
<td>2</td>
<td>29</td>
<td>2.9</td>
</tr>
<tr>
<td>3</td>
<td>31</td>
<td>3.1</td>
</tr>
<tr>
<td>4</td>
<td>34</td>
<td>3.4</td>
</tr>
<tr>
<td>5</td>
<td>37</td>
<td>3.7</td>
</tr>
<tr>
<td>6</td>
<td>21</td>
<td>2.1</td>
</tr>
<tr>
<td>7</td>
<td>39</td>
<td>3.9</td>
</tr>
<tr>
<td>8</td>
<td>42</td>
<td>4.2</td>
</tr>
<tr>
<td>9</td>
<td>28</td>
<td>2.8</td>
</tr>
<tr>
<td>10</td>
<td>37</td>
<td>3.7</td>
</tr>
<tr>
<td>11</td>
<td>33</td>
<td>3.3</td>
</tr>
<tr>
<td>12</td>
<td>26</td>
<td>2.6</td>
</tr>
<tr>
<td>13</td>
<td>35</td>
<td>3.5</td>
</tr>
<tr>
<td>14</td>
<td>43</td>
<td>4.3</td>
</tr>
<tr>
<td>15</td>
<td>31</td>
<td>3.1</td>
</tr>
<tr>
<td>16</td>
<td>38</td>
<td>3.8</td>
</tr>
<tr>
<td>17</td>
<td>37</td>
<td>3.7</td>
</tr>
<tr>
<td>18</td>
<td>39</td>
<td>3.9</td>
</tr>
<tr>
<td>19</td>
<td>23</td>
<td>2.3</td>
</tr>
<tr>
<td>20</td>
<td>41</td>
<td>4.1</td>
</tr>
<tr>
<td>21</td>
<td>33</td>
<td>3.3</td>
</tr>
<tr>
<td>22</td>
<td>28</td>
<td>2.8</td>
</tr>
<tr>
<td>23</td>
<td>37</td>
<td>3.7</td>
</tr>
<tr>
<td>24</td>
<td>39</td>
<td>3.9</td>
</tr>
<tr>
<td>25</td>
<td>36</td>
<td>3.6</td>
</tr>
<tr>
<td>26</td>
<td>35</td>
<td>3.5</td>
</tr>
<tr>
<td>27</td>
<td>40</td>
<td>4.0</td>
</tr>
<tr>
<td>28</td>
<td>40</td>
<td>4.0</td>
</tr>
<tr>
<td>29</td>
<td>43</td>
<td>4.3</td>
</tr>
<tr>
<td>Total</td>
<td>1000</td>
<td>100.0</td>
</tr>
</tbody>
</table>

Since the values of the five parameters of true behaviour pattern (viz. $X_b$, $s_b$, $X_g$, $s_g$ and $T$) and two design parameters (viz. $I$ and $G$) were randomly drawn from
rectangular distributions, the frequency distributions of these values would be expected to be approximately rectangular. The frequency distributions of the nominal values of the five parameters of true behaviour and two design parameters reported in Tables 2 to 8 do appear to be approximately rectangular.

Chi-square goodness-of-fit tests were carried out on the nominal values of the five parameters of true behaviour pattern (viz. $X_b$, $s_b$, $X_g$, $s_g$ and $T$) and two design parameters (viz. $I$ and $G$) to test that these values were drawn populations with rectangular distributions. Table 9 presents a summary of the results of the chi-square goodness-of-fit tests.

TABLE 9
Chi-square Goodness-of-fit Tests on Distribution of Values of Nominal Parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Chi-Square</th>
<th>D.F.</th>
<th>Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Duration of Behaviour</td>
<td>12.08</td>
<td>11</td>
<td>0.358</td>
</tr>
<tr>
<td>Std. Dev. of Duration of Behaviour</td>
<td>10.86</td>
<td>11</td>
<td>0.455</td>
</tr>
<tr>
<td>Mean Duration of Non-Behaviour</td>
<td>4.83</td>
<td>11</td>
<td>0.939</td>
</tr>
<tr>
<td>Std. Dev. of Duration of Non-Behaviour</td>
<td>9.63</td>
<td>11</td>
<td>0.564</td>
</tr>
<tr>
<td>Length of Time Interval</td>
<td>14.48</td>
<td>11</td>
<td>0.208</td>
</tr>
<tr>
<td>Length of Gap between Time Intervals</td>
<td>15.63</td>
<td>11</td>
<td>0.155</td>
</tr>
<tr>
<td>Number of True Occurrences of Behaviour</td>
<td>29.73</td>
<td>28</td>
<td>0.376</td>
</tr>
</tbody>
</table>
It can be seen from the results reported in Table 9 that none of the chi-square tests were significant at the 0.05 level. Thus there is no evidence to reject the assumption that the nominal values of the five parameters of true behaviour pattern (viz. $X_B$, $s_B$, $X_g$, $s_g$ and $T$) and two design parameters (viz. $I$ and $G$) were drawn from populations with rectangular distributions.

4.2.2 Intercorrelations of Nominal Values of Parameters

Since the nominal values of the five parameters of true behaviour pattern (viz. $X_b$, $s_b$, $X_g$, $s_g$ and $T$) and two design parameters (viz. $I$ and $G$) were generated independently of each other, the bivariate correlations between these variables have expected values of zero. A matrix of bivariate correlations was calculated to examine the independence of the parameters. Spearman rank-order correlation coefficients were calculated instead of the more usual Pearson product moment coefficients since the distributions of the variables were known to be non-normal.
### TABLE 10
Inter-correlations between Values of Nominal Parameters: Spearman Correlation Coefficients and Significance Level

<table>
<thead>
<tr>
<th></th>
<th>$\bar{X}_b$</th>
<th>$s_b$</th>
<th>$\bar{X}_g$</th>
<th>$s_g$</th>
<th>I</th>
<th>G</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\bar{X}_b$</td>
<td>$R_S$ -0.059</td>
<td>-0.059</td>
<td>0.016</td>
<td>0.013</td>
<td>0.024</td>
<td>0.082</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>$Sig$ -0.061</td>
<td>0.061</td>
<td>0.621</td>
<td>0.677</td>
<td>0.448</td>
<td>0.010</td>
<td>0.234</td>
</tr>
<tr>
<td>$s_b$</td>
<td>$R_S$ -0.059</td>
<td>-0.059</td>
<td>0.053</td>
<td>0.058</td>
<td>0.055</td>
<td>0.013</td>
<td>-0.013</td>
</tr>
<tr>
<td></td>
<td>$Sig$ 0.061</td>
<td>-0.059</td>
<td>0.093</td>
<td>0.067</td>
<td>0.083</td>
<td>0.677</td>
<td>0.677</td>
</tr>
<tr>
<td>$\bar{X}_g$</td>
<td>$R_S$ 0.016</td>
<td>0.016</td>
<td>0.053</td>
<td>-0.050</td>
<td>0.010</td>
<td>0.059</td>
<td>-0.022</td>
</tr>
<tr>
<td></td>
<td>$Sig$ 0.621</td>
<td>0.621</td>
<td>0.093</td>
<td>-0.115</td>
<td>0.749</td>
<td>0.062</td>
<td>0.492</td>
</tr>
<tr>
<td>$s_g$</td>
<td>$R_S$ 0.013</td>
<td>0.013</td>
<td>0.058</td>
<td>-0.050</td>
<td>-0.025</td>
<td>-0.054</td>
<td>-0.033</td>
</tr>
<tr>
<td></td>
<td>$Sig$ 0.677</td>
<td>0.677</td>
<td>0.067</td>
<td>0.115</td>
<td>-0.428</td>
<td>0.090</td>
<td>0.293</td>
</tr>
<tr>
<td>I</td>
<td>$R_S$ 0.024</td>
<td>0.024</td>
<td>0.055</td>
<td>0.010</td>
<td>0.025</td>
<td>-0.050</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td>$Sig$ 0.448</td>
<td>0.448</td>
<td>0.083</td>
<td>0.749</td>
<td>0.428</td>
<td>-0.114</td>
<td>0.911</td>
</tr>
<tr>
<td>G</td>
<td>$R_S$ 0.082</td>
<td>0.082</td>
<td>0.013</td>
<td>0.059</td>
<td>-0.054</td>
<td>-0.050</td>
<td>-0.057</td>
</tr>
<tr>
<td></td>
<td>$Sig$ 0.010</td>
<td>0.010</td>
<td>0.677</td>
<td>0.062</td>
<td>0.090</td>
<td>0.114</td>
<td>-0.070</td>
</tr>
<tr>
<td>T</td>
<td>$R_S$ 0.038</td>
<td>-0.038</td>
<td>-0.013</td>
<td>-0.022</td>
<td>-0.033</td>
<td>0.004</td>
<td>-0.057</td>
</tr>
<tr>
<td></td>
<td>$Sig$ 0.234</td>
<td>0.234</td>
<td>0.677</td>
<td>0.492</td>
<td>0.293</td>
<td>0.911</td>
<td>0.070</td>
</tr>
</tbody>
</table>

In making a number of comparisons on a set of data, such as was done with the intercorrelations of the true behaviour and design parameters, a problem arises with the Type I error rate. Keppel (1982) provides a succinct summary of the problem:

The major problem resulting from the performance of a series of analytical comparisons on a set of data is the unpleasant fact that the more comparisons we conduct, the more type I errors we will make when
the null hypothesis is true. In talking about this relationship, the distinction is often made between the type 1 error per comparison (PC) and the error rate familywise (FW). The PC error rate, which we will continue to call alpha, uses the comparison as the conceptual unit for the error rate. If we evaluated several comparisons in an experiment, each at \( \alpha = 0.05 \), we would be using a PC error rate; our probability of making a type I error would be 0.05 for each of the separate comparisons. In contrast, the type I FW error rate, \( \alpha_{FW} \), considers the probability of making one or more type I errors in the set of comparisons under scrutiny. (p. 145)

The relationship between PC and FW error rates is:

\[
\alpha_{FW} = 1 - (1 - \alpha)^c
\]

where \( \alpha_{FW} \) is the familywise error rate

\[ \alpha \] is the per comparison error rate

\[ c \] is the number of orthogonal comparisons conducted

To test the significance of a set of comparisons at a given FW error rate the above formula is re-expressed to yield a solution for the appropriate PC error rate.
\[
\begin{align*}
\alpha_{FW} &= 1 - (1 - \alpha)^c \\
(1 - \alpha)^c &= 1 - \alpha_{FW} \\
1 - \alpha &= (1 - \alpha_{FW})^{1/c} \\
\alpha &= 1 - (1 - \alpha_{FW})^{1/c}
\end{align*}
\]

In the present situation there were 21 comparisons made \((c = 21)\) at a familywise error rate of 0.05 \((\alpha_{FW} = 0.05)\). The appropriate per comparison error rate \((\alpha)\) is given by:

\[
\begin{align*}
\alpha &= 1 - (1 - \alpha_{FW})^{1/c} \\
&= 1 - (1 - 0.05)^{1/21} \\
&= 1 - (0.95)^{1/21} \\
&= 1 - 0.998 \\
&= .002
\end{align*}
\]

It should be noted that the above discussion relates to orthogonal comparisons. In the present situation the comparisons are not orthogonal, so that using the per comparison value of \(\alpha\) as calculated will still provide an inflated familywise error rate. In other words the probability of a type I error familywise will still be greater than the nominal value of the familywise type I error rate. The effect of using a familywise error rate for orthogonal comparisons on non-orthogonal
comparisons is to provide a less stringent test of the null hypothesis that all variables are uncorrelated, since the per comparison alpha value for non-orthogonal comparisons would be lower than the per comparison alpha value for orthogonal comparisons to test the null hypothesis at a given familywise error rate. In other words this approach is more likely to reject the null hypothesis that all variables are uncorrelated. Rejecting the null hypothesis that all variables are uncorrelated would mean accepting the alternative hypothesis that the variables are not all uncorrelated, and accepting this alternative hypothesis would raise serious doubts about the results of the data generation process. Since the test employed here allows rejection of the null hypothesis at somewhat less significant level than the nominal level of $\alpha_{FW}=0.05$ this amounts to a conservative test of the data generation process.

Inspection of the significance levels of the correlations reported in Table 10 reveals that none of the coefficients was significant at the 0.05 familywise level of significance. Thus there is no evidence to reject the premise that the values of the variables were generated independently of each other.
4.3 Actual Values of Parameters

4.3.1 Distribution of Actual Values of Parameters

As noted in Chapter III the actual values for mean and standard deviation of occurrence ($X_B$ and $s_B$) and mean and standard deviation of non-occurrence of behaviour ($X_g$ and $s_g$) were calculated from the randomly generated values for duration of occurrence and non-occurrence produced for each simulated behaviour pattern. Consequently the actual values of these four parameters were not expected to be the same as the nominal values of these four parameters. Because the data generation program prevented the selection of negative values for duration of occurrence and non-occurrence of behaviour, it was expected that the actual mean duration of occurrence and non-occurrence would be higher than the nominal values of mean duration of occurrence and non-occurrence. It was also expected that the actual standard deviation of duration of occurrence and non-occurrence would be lower than the nominal values of standard deviation of duration of occurrence and non-occurrence. The effect of preventing the selection of negative values for duration of occurrence and non-occurrence of behaviour was, therefore, to change the distributions of actual mean duration of occurrence and non-occurrence and actual standard deviation of duration of occurrence and
non-occurrence, relative to the rectangular distributions of the nominal values of these parameters. Tables 11 to 14 present frequency distributions of the actual values of mean duration of occurrence and non-occurrence and standard deviation of duration of occurrence and non-occurrence of behaviour.

It should be noted that no restrictions were placed on the actual values of interval (I), gap between intervals (G) and number of true occurrences of behaviour (T). Consequently the actual values of these three parameters are identical to the nominal values reported in Tables 6 to 8.
<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 99</td>
<td>6</td>
<td>0.6</td>
</tr>
<tr>
<td>100 - 199</td>
<td>25</td>
<td>2.5</td>
</tr>
<tr>
<td>200 - 299</td>
<td>47</td>
<td>4.7</td>
</tr>
<tr>
<td>300 - 399</td>
<td>55</td>
<td>5.5</td>
</tr>
<tr>
<td>400 - 499</td>
<td>76</td>
<td>7.6</td>
</tr>
<tr>
<td>500 - 599</td>
<td>76</td>
<td>7.6</td>
</tr>
<tr>
<td>600 - 699</td>
<td>96</td>
<td>9.6</td>
</tr>
<tr>
<td>700 - 799</td>
<td>109</td>
<td>10.9</td>
</tr>
<tr>
<td>800 - 899</td>
<td>106</td>
<td>10.6</td>
</tr>
<tr>
<td>900 - 999</td>
<td>114</td>
<td>11.4</td>
</tr>
<tr>
<td>1000 - 1099</td>
<td>120</td>
<td>12.0</td>
</tr>
<tr>
<td>1100 - 1199</td>
<td>78</td>
<td>7.8</td>
</tr>
<tr>
<td>1200 - 1299</td>
<td>50</td>
<td>5.0</td>
</tr>
<tr>
<td>1300 - 1399</td>
<td>28</td>
<td>2.8</td>
</tr>
<tr>
<td>1400 - 1499</td>
<td>14</td>
<td>1.4</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1000</strong></td>
<td><strong>100.0</strong></td>
</tr>
</tbody>
</table>
## TABLE 12
Frequency Distribution of Values of Actual Standard Deviation of Duration of Behaviour in Seconds

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 99</td>
<td>125</td>
<td>12.5</td>
</tr>
<tr>
<td>100 - 199</td>
<td>134</td>
<td>13.4</td>
</tr>
<tr>
<td>200 - 299</td>
<td>137</td>
<td>13.7</td>
</tr>
<tr>
<td>300 - 399</td>
<td>125</td>
<td>12.5</td>
</tr>
<tr>
<td>400 - 499</td>
<td>146</td>
<td>14.6</td>
</tr>
<tr>
<td>500 - 599</td>
<td>130</td>
<td>13.0</td>
</tr>
<tr>
<td>600 - 699</td>
<td>101</td>
<td>10.1</td>
</tr>
<tr>
<td>700 - 799</td>
<td>62</td>
<td>6.2</td>
</tr>
<tr>
<td>800 - 899</td>
<td>25</td>
<td>2.5</td>
</tr>
<tr>
<td>900 - 999</td>
<td>11</td>
<td>1.1</td>
</tr>
<tr>
<td>1000 - 1099</td>
<td>3</td>
<td>0.3</td>
</tr>
<tr>
<td>1100 - 1199</td>
<td>1</td>
<td>0.1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>1000</strong></td>
<td><strong>100.0</strong></td>
</tr>
<tr>
<td>Class Interval</td>
<td>Frequency</td>
<td>Percentage</td>
</tr>
<tr>
<td>----------------</td>
<td>-----------</td>
<td>------------</td>
</tr>
<tr>
<td>0 - 99</td>
<td>5</td>
<td>0.5</td>
</tr>
<tr>
<td>100 - 199</td>
<td>27</td>
<td>2.7</td>
</tr>
<tr>
<td>200 - 299</td>
<td>44</td>
<td>4.4</td>
</tr>
<tr>
<td>300 - 399</td>
<td>57</td>
<td>5.7</td>
</tr>
<tr>
<td>400 - 499</td>
<td>64</td>
<td>6.4</td>
</tr>
<tr>
<td>500 - 599</td>
<td>61</td>
<td>6.1</td>
</tr>
<tr>
<td>600 - 699</td>
<td>99</td>
<td>9.9</td>
</tr>
<tr>
<td>700 - 799</td>
<td>90</td>
<td>9.0</td>
</tr>
<tr>
<td>800 - 899</td>
<td>121</td>
<td>12.1</td>
</tr>
<tr>
<td>900 - 999</td>
<td>120</td>
<td>12.1</td>
</tr>
<tr>
<td>1000 - 1099</td>
<td>119</td>
<td>11.9</td>
</tr>
<tr>
<td>1100 - 1199</td>
<td>85</td>
<td>8.5</td>
</tr>
<tr>
<td>1200 - 1299</td>
<td>43</td>
<td>4.3</td>
</tr>
<tr>
<td>1300 - 1399</td>
<td>34</td>
<td>3.4</td>
</tr>
<tr>
<td>1400 - 1499</td>
<td>31</td>
<td>3.1</td>
</tr>
<tr>
<td>Total</td>
<td>1000</td>
<td>100.0</td>
</tr>
</tbody>
</table>
Data reported in Tables 11 and 13 suggest that the distribution of actual values of mean duration of occurrence and mean duration of non-occurrence were approximately normal. The change from the rectangular distributions of the nominal values of mean duration of occurrence and non-occurrence (Tables 2 and 4) was a result of the restriction that duration of occurrence and non-occurrence must be greater than zero.

The restriction that duration of occurrence and non-occurrence must be greater than zero had different
effects on the actual value of mean duration depending on the nominal value selected for mean duration. If the nominal value selected for mean duration was close to zero then many of the values of duration randomly selected from a normal distribution fell below zero and were excluded, thus producing a value for actual mean duration which was higher than the value for the nominal mean duration. Conversely if the nominal value selected for mean duration was high then few of the values of duration randomly selected from a normal distribution fell below zero and were excluded, thus the value for the actual mean duration was similar to that of the nominal mean duration.

The truncation of the normal distribution of values for duration was also affected by the nominal value selected for standard deviation of duration. If the nominal value selected for standard deviation of duration was small then few of the values for duration randomly selected from a normal distribution fell below zero, thus the value for actual mean duration was similar to that for nominal mean duration. If the nominal value selected for standard deviation of duration was large then more of the values for duration randomly selected from a normal distribution fell below zero, thus producing a value for
actual mean duration which was higher than the value for nominal mean duration.

The joint effect of nominal value of mean duration and nominal value of standard deviation of duration depended on the combinations of these values. If the nominal value selected for mean duration was low and the nominal value selected for standard deviation of duration was high then many values for duration selected from a normal distribution fell below zero, thus the actual value of mean duration was considerably higher than the nominal value of mean duration. If the nominal value selected for mean duration was low and the nominal value selected for standard deviation of duration was low then fewer values for duration selected from a normal distribution fell below zero, thus the actual value of mean duration was closer to the nominal value of mean duration.

Since the nominal values for mean duration and standard deviation of duration were selected independently, the number of combinations of small mean and small standard deviation was expected to be similar to the number of combinations of small mean and large standard deviation. Because of the interactive effect of
nominal mean duration and nominal standard deviation of duration on actual mean duration, half of the instances of low nominal mean (the half which were combined with large nominal standard deviation of duration) resulted in much higher actual mean duration of behaviour. The other half of the instances of low nominal mean duration (the half which were combined with small nominal standard deviation of duration) resulted in a value of actual mean duration which was similar to nominal mean duration. The net result of this interactive effect of nominal value of mean duration and standard deviation of duration was to produce an approximately normal distribution of values for actual mean duration of occurrence and non-occurrence as a result of having fewer instances of low values of actual mean duration than would occur in a rectangular distribution and more instances of medium values of actual mean duration than would occur in a rectangular distribution.

The results reported in Tables 12 and 14 show frequency distributions of the actual values of standard deviation of duration of occurrence and non-occurrence. The data presented in these tables show that the distribution of actual values for duration of occurrence and non-occurrence are approximately rectangular up to a
value of approximately 600 and the frequency of values above this level declines sharply.

The shape of the distributions of actual values of standard deviation of duration of occurrence and non-occurrence was again a result of the restriction that no value of duration of occurrence or non-occurrence was less than zero. In general the effect of this restriction was to reduce the actual standard deviation of duration since values below zero selected from a normal distribution were discarded, thus reducing the range of values selected.

The effect of the reduction of the range of values selected was not constant across all behaviour patterns. If the nominal value selected for mean duration was close to zero, then many values selected from a normal distribution fell below zero and were rejected, thus reducing the actual standard deviation considerably relative to the nominal standard deviation. If the nominal value selected for mean duration was high, then few values of duration selected from a normal distribution fell below zero and were rejected, thus the actual standard deviation was reduced marginally relative to the nominal standard deviation. The net result of
this reduction of actual standard deviation relative to nominal standard deviation was to produce a larger number of instances of low values of actual standard deviation relative to instances of high values of actual standard deviation, resulting in a "tapering off" of the frequency distribution of actual standard deviation at higher values of actual standard deviation.

4.3.2 Intercorrelations of Actual Values of Parameters

As noted in the introductory section of this chapter, the generalizability of the findings of this study rests initially on the degree to which the data generated represents real-life situations. In order to have all potential situations represented, it was necessary to generate data which covered all combinations of the five parameters of true behaviour pattern (viz. \( \bar{X} \), \( b \), \( s_b \), \( \bar{X}_g \), \( s_g \) and \( T \)) and two design parameters (viz. \( I \) and \( G \)). A measure of the degree to which the data generated represents all combinations of parameters is the bivariate correlations between all pairs of parameters. If the data generated do in fact represent all combinations equally then the bivariate correlations have expected values of zero. Table 15 presents a matrix of bivariate correlation coefficients between actual values of all pairs of the five parameters of true
behaviour pattern (viz. $X_b$, $s_b$, $X_g$, $s_g$ and $T$) and two
design parameters (viz. $I$ and $G$).

**TABLE 15**
Inter-correlations between Values of Actual Parameters:
Pearson Correlation Coefficients and Significance Level

<table>
<thead>
<tr>
<th></th>
<th>$X_b$</th>
<th>$s_b$</th>
<th>$X_g$</th>
<th>$s_g$</th>
<th>$I$</th>
<th>$G$</th>
<th>$T$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_b$</td>
<td>$R_s$</td>
<td>-0.596</td>
<td>0.040</td>
<td>0.031</td>
<td>0.044</td>
<td>0.098</td>
<td>0.037</td>
</tr>
<tr>
<td></td>
<td>$S_{ig}$</td>
<td>-0.000*</td>
<td>0.207</td>
<td>0.336</td>
<td>0.165</td>
<td>0.002*</td>
<td>0.246</td>
</tr>
<tr>
<td>$s_b$</td>
<td>$R_s$</td>
<td>0.596</td>
<td>-0.082</td>
<td>0.137</td>
<td>0.043</td>
<td>0.019</td>
<td>0.207*</td>
</tr>
<tr>
<td></td>
<td>$S_{ig}$</td>
<td>0.000*</td>
<td>-0.010</td>
<td>0.000*</td>
<td>0.176</td>
<td>0.540</td>
<td>0.000*</td>
</tr>
<tr>
<td>$X_g$</td>
<td>$R_s$</td>
<td>0.040</td>
<td>0.082</td>
<td>-0.601</td>
<td>-0.006</td>
<td>0.011</td>
<td>-0.018</td>
</tr>
<tr>
<td></td>
<td>$S_{ig}$</td>
<td>0.207</td>
<td>0.010</td>
<td>-0.000*</td>
<td>0.862</td>
<td>0.732</td>
<td>0.563</td>
</tr>
<tr>
<td>$s_g$</td>
<td>$R_s$</td>
<td>0.031</td>
<td>0.137</td>
<td>0.601</td>
<td>-0.009</td>
<td>-0.050</td>
<td>0.140*</td>
</tr>
<tr>
<td></td>
<td>$S_{ig}$</td>
<td>0.336</td>
<td>0.000*</td>
<td>0.000*</td>
<td>-0.767</td>
<td>0.114</td>
<td>0.000*</td>
</tr>
<tr>
<td>$I$</td>
<td>$R_s$</td>
<td>0.044</td>
<td>0.043</td>
<td>-0.006</td>
<td>0.009</td>
<td>-0.050</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td>$S_{ig}$</td>
<td>0.165</td>
<td>0.176</td>
<td>0.862</td>
<td>0.767</td>
<td>-0.115</td>
<td>0.877</td>
</tr>
<tr>
<td>$G$</td>
<td>$R_s$</td>
<td>0.098*</td>
<td>0.019</td>
<td>0.011</td>
<td>-0.050</td>
<td>-0.050</td>
<td>-0.058</td>
</tr>
<tr>
<td></td>
<td>$S_{ig}$</td>
<td>0.002*</td>
<td>0.540</td>
<td>0.732</td>
<td>0.114</td>
<td>0.115</td>
<td>-0.067</td>
</tr>
<tr>
<td>$T$</td>
<td>$R_s$</td>
<td>0.037</td>
<td>0.207</td>
<td>-0.018</td>
<td>0.140</td>
<td>0.005</td>
<td>-0.058</td>
</tr>
<tr>
<td></td>
<td>$S_{ig}$</td>
<td>0.246</td>
<td>0.000*</td>
<td>0.563</td>
<td>0.000*</td>
<td>0.977</td>
<td>0.067</td>
</tr>
</tbody>
</table>

* = $\text{Alpha}_{FW} < 0.05$

As noted earlier in this chapter, when making a
number of comparisons on the one set of data the
probability of a type I error is increased as the number
of comparisons is increased. The solution to this
problem is to reduce the per comparison (PC) significance
level such that the type I familywise (FW) error rate is
equal to the significance level chosen. For these data a PC significance level of alpha=0.002 results in a FW probability of type I error of alpha_FW=0.05. In other words, testing at the PC alpha level of 0.002 results in a 5% probability of making one or more type I errors in the set of comparisons under consideration.

Inspection of the significance levels reported in Table 15 reveals that five correlation coefficients were significant at the alpha_FW=0.05 level. Two of these correlations were between the mean and standard deviation of duration. Another two were between the standard deviation of duration and number of true occurrences. The final correlation significant at the alpha_FW=0.05 level was between standard deviation of duration of occurrence and standard deviation of duration of non-occurrence of behaviour.

The significant correlations between mean duration and standard deviation of duration result from the restriction that all values of duration must be greater than zero. If the nominal value selected for mean duration was low and the nominal value selected for standard deviation of duration was high many values of duration selected from a normal distribution fell below
zero and were discarded, thus reducing the actual value for standard deviation of duration. The result of this truncation of the normal distribution when nominal value of mean duration was low was to prevent many low mean/high standard deviation combinations from being generated and to force the generation of a large number of low mean/low standard deviation combinations. The truncation resulted in a relationship between low mean and low standard deviation and consequently a significant correlation between these two variables.

The correlation between mean duration and standard deviation of duration demonstrated by the data of this study reflects a statistical fact which would be expected in real-life behaviour patterns. If mean duration of behaviour is low this will tend to be associated with a lower standard deviation of duration of behaviour, since no instance of behaviour can occur for less than zero time, and thus the range of duration is truncated. Thus the correlations reported here suggest that the data generated reflect the expected relationships within real-life behaviour patterns.

The significant correlations between standard deviation of duration and number of occurrences results
from the use of a normal distribution to select values for the lengths of instances of occurrence. In selecting a single value from a normal distribution the probability is greater that the value will fall close to the mean than is the probability that the value will fall at a larger distance from the mean. In selecting two values from a normal distribution independently, the probability is greater that both values will fall close together near the mean than it is that the two values will fall farther apart. As the number of values selected is increased, the probability of extreme values is increased, and thus the probability is higher that the standard deviation of the set of values selected will be large.

In this instance the number of true occurrences \((T)\) is the number of values for duration selected from a normal distribution. As the number of true occurrences increases the probability of a larger standard deviation in the set of numbers increases, producing a correlation between number of true occurrences \((T)\) and standard deviation of duration of occurrence \((s_B)\) or non-occurrence \((s_g)\).

The correlation between number of true occurrences and standard deviation of duration demonstrated by the
data of this study reflects a statistical fact which would be expected in real-life behaviour patterns. Thus the data generated in this study reflect the expected relationships within real-life behaviour patterns.

The significant correlation between standard deviation of duration of behaviour \( (s_b) \) and standard deviation of duration of non-behaviour \( (s_g) \) does not have an immediately apparent explanation. Part of the explanation for this significant correlation may be due to the fact that standard deviation of behaviour \( (s_b) \) and standard deviation of non-behaviour \( (s_g) \) both correlate with true number of occurrences of behaviour \( (T) \). Part of the explanation for this significant correlation may reside in the fact that the nominal values of \( s_b \) and \( s_g \) correlated 0.058 (see Table 10). Part of the explanation for this significant correlation may be fortuitous random fluctuation or the effect of one or two outlying cases producing a spuriously significant correlation. Conversely the correlation between \( s_b \) and \( s_g \) may reflect an effect of the restriction that all values of duration of occurrence and non-occurrence must exceed zero.

Regardless of the reason for the significant correlation observed between \( s_b \) and \( s_g \) it should be noted
that the value of the correlation coefficient ($R_s=0.147$) was not large. The Pearson product moment correlation between $s_b$ and $s_g$ ($r=0.137$) indicates that less than 2% of the standard deviation of one variable is common with the standard deviation of the other variable. Because the correlation between $s_b$ and $s_g$ is low, these variables will be considered to be independent for this study.

The significant correlation between mean duration of behaviour ($X_b$) and length of the gap between time intervals ($G$) similarly does not have an immediately apparent explanation. However, as with the significant correlation between standard deviation of behaviour and standard deviation of non-behaviour, it should be noted that the magnitude of the correlation was small. The Pearson product-moment correlation coefficient ($r=0.098$) indicates that less than 1% of the variance of one variable is common with the variance of the other variable. Because the correlation between $X_b$ and $G$ is low, these variables will be considered to be independent for this study.
4.4 Distribution of Proportion of Occurrence

As noted in Chapter III a number of researchers using time sampling procedures report the proportion of time for which behaviour occurs as one of the statistics descriptive of the behaviour pattern under study. To enable comparisons between the data generated in this study for true behaviour patterns and the data reported in other studies a frequency distribution of actual proportion of time for which behaviour occurred was produced.

<table>
<thead>
<tr>
<th>Class Interval</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00 - 0.09</td>
<td>7</td>
<td>0.7</td>
</tr>
<tr>
<td>0.10 - 0.19</td>
<td>25</td>
<td>2.5</td>
</tr>
<tr>
<td>0.20 - 0.29</td>
<td>74</td>
<td>7.4</td>
</tr>
<tr>
<td>0.30 - 0.39</td>
<td>159</td>
<td>15.9</td>
</tr>
<tr>
<td>0.40 - 0.49</td>
<td>242</td>
<td>24.2</td>
</tr>
<tr>
<td>0.50 - 0.59</td>
<td>234</td>
<td>23.4</td>
</tr>
<tr>
<td>0.60 - 0.69</td>
<td>152</td>
<td>15.2</td>
</tr>
<tr>
<td>0.70 - 0.79</td>
<td>70</td>
<td>7.0</td>
</tr>
<tr>
<td>0.80 - 0.89</td>
<td>35</td>
<td>3.5</td>
</tr>
<tr>
<td>0.90 - 0.99</td>
<td>2</td>
<td>0.2</td>
</tr>
</tbody>
</table>

| Total          | 1000      | 100.0      |

The data reported in Table 16 suggest that across the 1000 simulated behaviour patterns generated for the
present study the distribution of the values for proportion of time behaviour occurs was approximately normal providing evidence that the data generated to represent true behaviour pattern in the present study does represent the range of potential proportion of occurrence of real-life behaviour patterns.

4.5 Summary of Characteristics of Generated Data

The actual values selected for mean duration of occurrence and non-occurrence of behaviour and standard deviation of duration of occurrence and non-occurrence of behaviour cover a range of situations from extremely short duration, regular, frequent behaviour to extremely long duration, irregular, infrequent behaviour. The range of situations covered by the values selected for mean duration of occurrence and non-occurrence of behaviour and standard deviation of duration of occurrence and non-occurrence of behaviour were examined by time sampling procedures using a range of values from extremely short observation interval and gap to extremely long observation interval and gap. The proportion of time for which behaviour occurred ranged from extremely low (less than 10% of the time) to extremely high (greater than 90% of the time). It seems reasonable to conclude that the data generated to represent true
behaviour pattern and design of time sampling studies is an adequate representation of real-life situations.
CHAPTER V
RESULTS: REGRESSION ANALYSES

5.1 Model One

As noted in Chapter 3 the following symbols were used to refer to the variables used in this study:

\[ \begin{align*}
    \bar{X}_b &= \text{Mean Duration of Behaviour} \\
    \bar{X}_g &= \text{Mean Duration of Non-behaviour} \\
    s_b &= \text{Standard Dev. of Duration of Behaviour} \\
    s_g &= \text{Standard Dev. of Duration of Non-behaviour} \\
    T &= \text{True Number of Occurrences} \\
    I &= \text{Length of Time Interval Observed (in seconds)} \\
    G &= \text{Length of Gap between I's (in seconds)} \\
    O &= \text{Number of intervals in which the behaviour was observed.}
\end{align*} \]

As noted in Chapter 3 the data generated in this study to represent time sampling studies were analysed through linear regression techniques. A separate linear regression model was built for each of the three time sampling procedures examined. Each regression model was used to examine the degree to which accuracy of time sampling for estimating frequency of occurrence could be predicted from the parameters of true behaviour and the
design parameters of time sampling studies. Estimates of the parameters of the model were made through stepwise multiple linear regression analysis. All calculations were performed using procedure Regression in the Statistical Package for the Social Sciences Version X (SPSSX) computer program package described in SPSS Inc. (1983).

The general form of each model was:

\[ Y = a + b_1X_1 + b_2X_2 + \ldots + b_6X_6 \]

where \( Y = O/T \)

and \( X_1 \) to \( X_6 = \) independent variables listed above

The dependent variable in this study was the ratio of observed number of occurrences of behaviour to true number of occurrences of behaviour (i.e. \( O/T \)). As discussed in Chapter 3 this ratio should strictly be referred to as the degree of overestimation or the degree of underestimation of number of observed occurrences in relation to number of true occurrences of behaviour. For simplicity the dependent variable will be referred to simply as the accuracy of estimation. Tables 17 to 19 present summary results of the analyses for whole
interval, partial interval and momentary time sampling procedures respectively.

TABLE 17
Summary of Stepwise Regression Solution for Whole Interval Time Sampling Procedure: Model 1

<table>
<thead>
<tr>
<th>Step Number</th>
<th>Variable</th>
<th>Overall R²</th>
<th>Model F</th>
<th>Change this Step R²</th>
<th>Change this Step F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I</td>
<td>0.2813</td>
<td>390.54</td>
<td>0.2813</td>
<td>390.54</td>
</tr>
<tr>
<td>2</td>
<td>Xₐ</td>
<td>0.3235</td>
<td>238.41</td>
<td>0.0423</td>
<td>62.29</td>
</tr>
<tr>
<td>3</td>
<td>G</td>
<td>0.3272</td>
<td>161.45</td>
<td>0.0037</td>
<td>5.41</td>
</tr>
</tbody>
</table>

* All F ratios were significant at p < .01

TABLE 18
Summary of Stepwise Regression Solution for Partial Interval Time Sampling Procedure: Model 1

<table>
<thead>
<tr>
<th>Step Number</th>
<th>Variable</th>
<th>Overall R²</th>
<th>Model F</th>
<th>Change this Step R²</th>
<th>Change this Step F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I</td>
<td>0.3090</td>
<td>446.24</td>
<td>0.3090</td>
<td>446.24</td>
</tr>
<tr>
<td>2</td>
<td>Xₐ</td>
<td>0.3676</td>
<td>289.71</td>
<td>0.0586</td>
<td>92.34</td>
</tr>
<tr>
<td>3</td>
<td>G</td>
<td>0.3738</td>
<td>198.18</td>
<td>0.0062</td>
<td>9.93</td>
</tr>
</tbody>
</table>

* All F ratios were significant at p < .01
TABLE 19
Summary of Stepwise Regression Solution for Momentary Time Sampling Procedure: Model 1

<table>
<thead>
<tr>
<th>Step Number</th>
<th>Variable Entered</th>
<th>Overall Model $R^2$</th>
<th>Model $F^*$</th>
<th>Change this Step $R^2$</th>
<th>Step $F^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>I</td>
<td>0.3101</td>
<td>448.55</td>
<td>0.3101</td>
<td>448.55</td>
</tr>
<tr>
<td>2</td>
<td>$\bar{X}_b$</td>
<td>0.3637</td>
<td>284.99</td>
<td>0.0537</td>
<td>84.08</td>
</tr>
<tr>
<td>3</td>
<td>G</td>
<td>0.3682</td>
<td>193.49</td>
<td>0.0045</td>
<td>7.04</td>
</tr>
</tbody>
</table>

* All $F$ ratios were significant at $p < .01$

Several points need to be noted in relation to the results presented in Tables 17 to 19. The first point is that all three models produced a statistically significant ($p < 0.01$) prediction of accuracy of time sampling for estimating frequency of occurrence of behaviour. Variance accounted for in the dependent variable ranged from 32.72% to 37.88%. Thus the accuracy of time sampling procedures for estimating frequency of occurrence of behaviour is significantly related to the variables selected through stepwise multiple regression analysis for the models.
The second point to note in relation to the results presented in Tables 17 to 19 is the consistency of the models for the three different time sampling procedures. Stepwise multiple regression is a technique which enters variables into a regression model one at a time and keeps variables in a regression model as long as the variables make a statistically significant increase in the proportion of variance accounted for. For all three time sampling procedures the same variables entered the model in the same order, and the same variables did not enter the model. Specifically, it seems that prediction of the accuracy of time sampling for estimating frequency of occurrence of behaviour can be based on the length of the time interval (\(I\)), the length of the gap between intervals (\(G\)), and the mean duration of behaviour (\(X_b\)). The other parameters of true behaviour pattern do not add significantly to the predictive power of a linear model based on these three variables.

The third point to note in relation to the results presented in Tables 17 to 19 is that the models account for between 32% and 38% of the variation in accuracy of time sampling for predicting frequency of occurrence of behaviour. While the proportion of explained variation is statistically significantly different from zero, the
important issue is whether the proportion of variation explained is important in terms of time sampling studies. Models such as those reported here which account for only about one-third of the variation in the dependent variable do not provide enough predictive power to enable researchers to confidently select values for parameters for time sampling designs. About two-thirds of the variation in accuracy was not explained by the variables in the model.

5.2 Model Two

There are three possible explanations for the failure of Model One to account for a practically important (as distinct from statistically significant) proportion of variation in the dependent variable: (a) a large proportion of the variation in the dependent variable was due to random error, and therefore was not predictable; (b) there are variables contributing to the explanation of the variation in the dependent variable which were not included in the model; and (c) a non-linear or non-additive relationship exists between the dependent variable and the independent variables in the model.
The issue now was to decide whether, and if so how, to proceed with analyses of the data. The first possible explanation for the failure of Model One listed above, that a large proportion of the variation in the dependent variable was due to random error, did not provide any insight into further directions for analysis. In fact if this was the correct explanation for the failure of Model One then no further analyses would be warranted.

The second possible explanation for the failure of Model One listed above, that the variation in the dependent variable was related to variables which were not included in the model, was not congruent with the conceptual model of time sampling presented in Chapter 3. The conceptual analysis of true behaviour pattern and design variables presented in Chapter 3 identified the independent variables involved in true behaviour pattern and design aspects of a time sampling study. Consequently it did not seem plausible that the failure of Model One to explain a larger proportion of the variation in the dependent variable was due to the existence of variables not considered in Model One.

The third possible explanation for the failure of Model One to account for a larger proportion of the
variation in the dependent variable was that the relationship between the independent variables and the dependent variable was non-linear or non-additive or both. This explanation suggested that further analyses incorporating non-linear and non-additive relationships would be warranted.

Since the first possible explanation of the inadequacy of Model One did not suggest further direction for analyses, and the second possible explanation did not seem plausible, analysis proceeded in the direction suggested by the third possible explanation of the inadequacy of Model One. There was a virtually unlimited number of combinations of multiplicative and exponential terms which could be produced from the seven fundamental independent variables examined in Model One. Consequently initial consideration was limited to those variables which contributed significantly to the explanation of variation in the dependent variable of Model One (viz. $X_b$, $I$ and $G$).

In considering the type of functions to produce from the three variables under consideration (viz. $X_b$, $I$ and $G$) a simple hypothetical situation was developed such that the effects of all the other independent variables
listed in Chapter 3 were controlled. Specifically, consideration was given to the relationship of $X_b$, $I$ and $G$ when $s_b$ and $s_g$ were zero and $X_g$ was equal to $X_b$. The true behaviour pattern described by these values of the parameters of true behaviour consisted of occurrences of the behaviour with every occurrence having the same duration (i.e. $s_b=0$), and gaps between the occurrences of behaviour with every gap having the same duration (i.e. $s_g=0$). The duration of occurrences and non-occurrences was equal, resulting in a perfectly regular behaviour pattern. Under these conditions it is intuitively apparent that if the value of $I + G$ equals the value of $X_b$ (i.e. $X_b/(I+G) = 1.0$) and observation commences with the onset of a behaviour, then the observed number of behaviours ($O$) would equal the true number of behaviours ($T$), resulting in perfect accuracy (i.e. $O/T = 1.0$). Perfect accuracy would be achieved because the pattern of observation intervals would coincide exactly with the pattern of occurrences of behaviour. This would be true for all three time sampling procedures considered in the present study. The foregoing analysis suggested the use of the ratio of $X_b$ to the sum of $I + G$ as a predictor variable.
Consideration was next given to the effect of relaxing some of the constraints imposed in order to develop a perfectly regular behaviour pattern. The purpose of considering the effect of accuracy of relaxing some of the constraints imposed in the simple situation developed above was to determine which other variables could possibly influence accuracy, and therefore which other variables should be incorporated in Model Two.

If the duration of occurrences of behaviour varied (i.e. $s_b \neq 0$) it seemed possible that accuracy would be affected since the duration of $I + G$ would no longer equal the duration of every occurrence of behaviour. Allowing variance in duration of occurrence would result in the pattern of observation intervals no longer being synchronized with the pattern of occurrences of behaviour. If duration of the gaps between occurrences of behaviour varied (i.e. $s_g \neq 0$) it seemed similarly possible that accuracy would be affected since again the pattern of observation intervals would no longer be synchronized with the pattern of occurrences of behaviour. If $X_g$ was not equal to $X_b$ this again was expected to affect accuracy since the pattern of occurrences of behaviour would not be synchronized with the pattern of observation intervals. If the pattern of
occurrences of behaviour was not synchronized with the pattern of observation intervals it seemed possible that longer observation sessions (i.e. where T had a larger value) would produce more inaccuracy than shorter observation sessions.

On the basis of the preceding conceptual analysis of the relationship between pattern of true behaviour and design parameters, a second model (Model Two) was developed and examined through stepwise multiple linear regression. Model Two was:

\[ Y = a + b_1X_1 + b_2X_2 + \ldots + b_7X_7 \]

where \( Y = \frac{O}{T} \)

\( X_1 \) to \( X_6 \) = independent variables described in Chapter 3 and listed at the start of this chapter

and \( X_7 = \frac{X_P}{(I+G)} \)

Model Two was examined separately for the three time sampling procedures under consideration (viz. whole interval, partial interval and momentary time sampling). Tables 20 to 22 present a summary of the results of the stepwise multiple linear regression analysis of Model Two.
TABLE 20
Summary of Stepwise Regression Solution for Whole Interval Time Sampling Procedure: Model 2

<table>
<thead>
<tr>
<th>Step Number</th>
<th>Variable Entered</th>
<th>Overall Model</th>
<th>Change this Step</th>
<th>F</th>
<th>R²</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$X_b/(I+G)$</td>
<td>0.9961</td>
<td>0.9961</td>
<td>253643.75**</td>
<td>0.9961</td>
</tr>
<tr>
<td>2</td>
<td>$X_b$</td>
<td>0.9972</td>
<td>0.0012</td>
<td>180626.77**</td>
<td>0.0000</td>
</tr>
<tr>
<td>3</td>
<td>I</td>
<td>0.9977</td>
<td>0.0004</td>
<td>140981.37**</td>
<td>0.0000</td>
</tr>
<tr>
<td>4</td>
<td>G</td>
<td>0.9978</td>
<td>0.0002</td>
<td>114695.24**</td>
<td>0.0000</td>
</tr>
<tr>
<td>5</td>
<td>$s_b$</td>
<td>0.9979</td>
<td>0.0000</td>
<td>93823.30**</td>
<td>0.0000</td>
</tr>
<tr>
<td>6</td>
<td>$s_g$</td>
<td>0.9979</td>
<td>0.0000</td>
<td>78540.07**</td>
<td>0.0000</td>
</tr>
<tr>
<td>7</td>
<td>$X_g$</td>
<td>0.9979</td>
<td>0.0000</td>
<td>67576.59**</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

* $p < .05$
** $p < .01$
TABLE 21
Summary of Stepwise Regression Solution for Partial Interval Time Sampling Procedure: Model 2

<table>
<thead>
<tr>
<th>Step Number</th>
<th>Variable</th>
<th>Overall Model R²</th>
<th>Overall Model F</th>
<th>Change this Step R²</th>
<th>Change this Step F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Xₖ/(I+G)</td>
<td>0.9965</td>
<td>284742.32</td>
<td>0.9965</td>
<td>284742.32</td>
</tr>
<tr>
<td>2</td>
<td>X₉</td>
<td>0.9975</td>
<td>196289.82</td>
<td>0.0010</td>
<td>377.64</td>
</tr>
<tr>
<td>3</td>
<td>I</td>
<td>0.9978</td>
<td>148835.72</td>
<td>0.0003</td>
<td>137.61</td>
</tr>
<tr>
<td>4</td>
<td>G</td>
<td>0.9978</td>
<td>114188.31</td>
<td>0.0001</td>
<td>23.80</td>
</tr>
<tr>
<td>5</td>
<td>s₉</td>
<td>0.9979</td>
<td>93840.45</td>
<td>0.0001</td>
<td>28.06</td>
</tr>
<tr>
<td>6</td>
<td>sₙ</td>
<td>0.9979</td>
<td>79446.34</td>
<td>0.0000</td>
<td>16.80</td>
</tr>
</tbody>
</table>

* All F ratios were significant at p < .01
TABLE 22
Summary of Stepwise Regression Solution for Momentary Time Sampling Procedure: Model 2

<table>
<thead>
<tr>
<th>Step Number</th>
<th>Variable Entered</th>
<th>Overall Model $R^2$</th>
<th>Model $F^*$</th>
<th>Change this Step $R^2$</th>
<th>Change this Step $F^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$X_b/(I+G)$</td>
<td>0.9972</td>
<td>361065.85</td>
<td>0.9972</td>
<td>361065.85</td>
</tr>
<tr>
<td>2</td>
<td>$X_g$</td>
<td>0.9979</td>
<td>234740.56</td>
<td>0.0006</td>
<td>299.84</td>
</tr>
<tr>
<td>3</td>
<td>$I$</td>
<td>0.9982</td>
<td>186234.01</td>
<td>0.0003</td>
<td>190.07</td>
</tr>
<tr>
<td>4</td>
<td>$X_b$</td>
<td>0.9983</td>
<td>142661.49</td>
<td>0.0000</td>
<td>22.25</td>
</tr>
<tr>
<td>5</td>
<td>$s_b$</td>
<td>0.9983</td>
<td>115665.58</td>
<td>0.0000</td>
<td>14.37</td>
</tr>
<tr>
<td>6</td>
<td>$s_g$</td>
<td>0.9983</td>
<td>98020.84</td>
<td>0.0000</td>
<td>17.81</td>
</tr>
<tr>
<td>7</td>
<td>$G$</td>
<td>0.9983</td>
<td>85036.26</td>
<td>0.0000</td>
<td>13.01</td>
</tr>
</tbody>
</table>

* All $F$ ratios were significant at $p < .01$

The results of the stepwise multiple linear regression analyses reported in Tables 20 to 22 were extremely consistent across the three time sampling procedures. For each time sampling procedure the new prediction variable $X_b/(I+G)$ was entered into the model first and accounted for more than 99.5% of the variation in the dependent variable. While most other variables produced a statistically significant increase in the proportion of variation in the dependent variable.
accounted for, they did not improve the predictive power of the model in a practical sense. Clearly accuracy (i.e. O/T) was almost perfectly predicted by Xb/(I+G).

5.3 Model Three

On the basis of the results reported in Tables 20 to 22, Model Three was developed as a simplification of Model Two. Model Three was:

\[ Y = a + b_1 X_1 \]

where \( Y = O/T \)

and \( X_1 = X_b/(I+G) \)

Since the variable \( X_b/(I+G) \) was entered on the first step in the stepwise regression analysis of Model Two, the \( R^2 \) value was the same for \( X_b/(I+G) \) in Model Three as it was for this variable in the first step of Model Two. Model Three is in effect the model developed in the first step of the stepwise regression solution for Model Two. Table 23 presents a summary of the linear regression analysis of Model Three for the three time sampling procedures. Table 24 presents a summary of the parameter estimates for Model Three for the three time sampling procedures.
TABLE 23
Summary of Regression Solutions for Three Time Sampling Procedures: Model 3

<table>
<thead>
<tr>
<th>Time Sampling Procedure</th>
<th>Sum of Squares Regression</th>
<th>Sum of Squares Residual</th>
<th>$r^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole</td>
<td>11925.69</td>
<td>46.92</td>
<td>0.9961*</td>
</tr>
<tr>
<td>Partial</td>
<td>11763.75</td>
<td>41.23</td>
<td>0.9965*</td>
</tr>
<tr>
<td>Momentary</td>
<td>12271.94</td>
<td>33.92</td>
<td>0.9972*</td>
</tr>
</tbody>
</table>

* $p < .001$

TABLE 24
Parameter Estimates from Regression Solutions for Three Time Sampling Procedures: Model 3

<table>
<thead>
<tr>
<th>Time Sampling Procedure</th>
<th>Intercept</th>
<th>Raw Regression Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole</td>
<td>-0.6906</td>
<td>1.0051*</td>
</tr>
<tr>
<td>Partial</td>
<td>0.7032</td>
<td>0.9982*</td>
</tr>
<tr>
<td>Momentary</td>
<td>-0.2281</td>
<td>1.0196*</td>
</tr>
</tbody>
</table>

* $p < .001$

It can be seen from the results presented in Table 23 that more than 99.5% of the variation in accuracy (i.e. $O/T$) can be predicted by the variable $X_b/(I+G)$ for each time sampling procedure. The equations based on the
regression solution of Model Three for predicting accuracy from the raw values of $X_b$, $I$ and $G$ for each time sampling procedure were:

**Whole Interval Time Sampling:**

$$Y = 1.0051(X) - 0.6906$$

**Partial Interval Time Sampling:**

$$Y = 0.9982(X) + 0.7032$$

**Momentary Time Sampling:**

$$Y = 1.0196(X) - 0.2281$$

where $Y = O / T$

and $X = X_b/(I+G)$
CHAPTER VI
DISCUSSION AND IMPLICATIONS

6.1 Summary of Findings

As noted in Chapter 3 the following symbols were used to refer to the variables used in this study:

\[ \begin{align*}
\bar{X}_b &= \text{Mean Duration of Behaviour} \\
\bar{X}_g &= \text{Mean Duration of Non-behaviour} \\
S_b &= \text{Standard Dev. of Duration of Behaviour} \\
S_g &= \text{Standard Dev. of Duration of Non-behaviour} \\
T &= \text{True Number of Occurrences} \\
I &= \text{Length of Time Interval Observed (in seconds)} \\
G &= \text{Length of Gap between I's (in seconds)} \\
O &= \text{Number of intervals in which the behaviour was observed.}
\end{align*} \]

As noted in Chapter 2 the purpose of this study was to investigate the effect of design factors and true behaviour pattern on the accuracy of time sampling for estimating frequency of occurrence of behaviour. This purpose was achieved by developing a regression model which aimed to account for as much of the variation in accuracy as possible on the basis of the variation in design variables and true behaviour pattern. The effect
of design factors and true behaviour pattern was investigated separately for whole interval, partial interval, and momentary time sampling. The results reported in Chapter 5 clearly show that this aim was achieved. For all three time sampling procedures a regression model was developed which accounted for in excess of 99.5% of the variation in accuracy of time sampling. In practical terms this means that accuracy of time sampling for estimating frequency of occurrence can be perfectly predicted by the independent variable (viz. $X_b/(I+G)$) in the final regression model reported in Chapter 5.

6.2 Relation of Findings to Previous Research

Previous researchers examining the accuracy of time sampling for estimating frequency of occurrence have attempted to provide recommendations for the appropriate length of the observation interval. In 1939 Arrington (cited in Irwin and Bushnell, 1980) proposed as a rule of thumb based on her experience that the time interval should be approximately the length of a single instance of the behaviour. Given that duration of instances of behaviour would normally vary, this recommendation translates into the suggestion that the time interval should approximate the mean duration of behaviour. If
time intervals are interspersed with gaps for recording observations, then presumably Arrington's rule of thumb would translate into the recommendation that the sum of the length of observation interval and gap between intervals should equal the mean duration of behaviour.

Recently researchers have empirically examined the question of factors affecting the accuracy of time sampling for estimating frequency of occurrence of behaviour. However results of recent research have been inconclusive. Powell and Rockinson (1978) conducted a study in which they used a computer to simulate time sampling studies conducted with differing behaviour patterns and differing time intervals for recording behaviour. Powell and Rockinson concluded that there were many combinations of behavioural frequency and duration where time sampling did not produce acceptably accurate observational data.

Tyler (1979) conducted a similar study with computer generated data and concluded that the accuracy of time sampling for estimating frequency of occurrence was a function of the length of the time interval used and the true behaviour pattern. He suggested that accuracy was highest when the time interval was short.
In contrast with the findings reported by Powell and Rockinson (1978) and Tyler (1979), McDowell (1973) reported the results of a study he conducted from which he concluded that time sampling procedures produced practically identical data to continuous recording. Since continuous recording involves no sampling error, the implication of McDowell's conclusion is that time sampling is an accurate method for estimating frequency of occurrence of behaviour.

The results of the present study clearly indicate the relationship between accuracy of time sampling procedures for estimating frequency of occurrence of behaviour, and mean duration of behaviour, length of interval, and length of gap between intervals. Regression analyses reported in Chapter 5 show that, in practical terms, accuracy of time sampling is perfectly related to a composite of mean duration of behaviour, length of interval, and length of gap.

Consequently the results of the present study quite clearly contradict the conclusions drawn by Powell and Rockinson (1978) and Tyler (1979), and describe the circumstances under which McDowell's (1973) conclusions
are correct. Furthermore the present study extended the scope of research in the field by developing a model which described the degree and nature of the interrelationships between the accuracy of time sampling, and the design and true behaviour pattern factors. The findings of the present study provide empirical evidence for the rule of thumb proposed by Arrington (cited in Irwin and Bushnell, 1980).

6.3 Application of Findings to Time Sampling Studies

The findings of the present study have two direct applications to time sampling studies. The first application is of value to the researcher planning a time sampling study, since it enables the researcher to determine the optimum length of the time interval to produce maximum accuracy \( O/T \) of the data collected. The present study has ascertained the mathematical relationship between accuracy \( O/T \) and mean duration of behaviour \( X_b \), length of observational interval \( I \), and length of gap between intervals \( G \).

Specifically,

\[
\frac{O}{T} = a + b \frac{X_b}{I+G}
\]
Since perfect accuracy occurs when $O/T = 1.0$, then the requirement for perfect accuracy is that

$$a + b \frac{X_b}{(I+G)} = 1.0$$

By restating this relationship it is possible to provide a formula for ascertaining the optimum value of length of interval and length of gap between intervals as follows:

$$a + b \frac{X_b}{(I+G)} = 1.0$$

$$b \frac{X_b}{(I+G)} = 1.0 - a$$

$$\frac{X_b}{(I+G)} = \frac{1.0 - a}{b}$$

$$(I+G)(1.0 - a) = bX_b$$

$$I+G = \frac{bX_b}{1.0 - a}$$

The results reported in Chapter 5 provide empirically determined values for the constants $a$ and $b$
for each of the three time sampling procedures examined in the present study. To apply the above formula and ascertain the optimum value of $I + G$ requires only an estimate of the mean duration of behaviour ($X_b$). An estimate could be provided by previous research in the area of interest or by a short pilot study using continuous recording methods. The researcher planning a time sampling study can then substitute an estimate of $X_b$ in the above formula to arrive at a value for the sum of $I$ and $G$. He may then determine the value of $I$ and $G$ on pragmatic grounds such as the time required to record observations.

The second application of the findings of the present study is in relation to data previously collected by time sampling procedures. Using the mathematical relationship ascertained in this study between accuracy of time sampling procedures and mean duration of behaviour, length of time interval, and length of gap between intervals, it is now possible to assess the accuracy of data collected previously using time sampling. The accuracy of data previously collected is given by:
Furthermore by restating this relationship it is possible to calculate the true number of occurrences of behaviour based on the observed number of occurrences as follows:

\[
\frac{O}{T} = a + b \frac{X_B}{(I+G)}
\]

\[
\frac{O}{T} = a + b \bar{X}_B/(I+G)
\]

\[
O = T \left( a + b \bar{X}_B/(I+G) \right)
\]

\[
T = \frac{O}{a + b \bar{X}_B/(I+G)}
\]

The values of the constants a and b were empirically determined for each time sampling procedure in this study and are reported in Chapter 5. The values of length of interval (I) and length of gap between intervals (G) would presumably be available from the study under consideration. To apply the above formula requires only an estimate of mean duration of occurrence of behaviour. This estimate could be provided from previous research, or could be obtained from a short pilot study using
continuous recording methods. With this information the above formula can be used as a "correction factor" to determine the true number of occurrences of the behaviour based on the observed number reported in a previous study.

6.4 Implications of Findings of Present Study

The results of the present study have two major implications for time sampling studies. The first implication is that data collected in many previous studies using time sampling procedures is less than perfectly accurate. Powell and Rockinson (1978) concluded from their study based on computer simulated time sampling that there were many combinations of behaviour pattern and design factors under which time sampling did not produce acceptably accurate data. The findings of the present study lead to the inference that for a particular situation there is a specific value of I + G which will enable time sampling to produce accurate data. Any other values will produce data which is inaccurate to some degree. Since it is unlikely that researchers using time sampling would select the optimum value of I + G by chance or by using a rule of thumb, it seems reasonable to suggest that most researchers using time sampling have used a value other than the optimum
value for I + G, and have therefore conducted studies which have resulted in inaccurate observational data.

The second major implication of the results of the present study relates to the use of time sampling procedures to measure the degree of change in frequency of occurrence of behaviour over time. The use of time sampling to measure degree of change in behavioural frequency over time is applicable to any longitudinal study, and is particularly relevant to studies in which the experimenter introduces some treatment aimed at affecting the subject's behaviour. Time sampling procedures are often used to measure the change in behaviour, with the assumption that any observed change in behaviour is a reflection of the effect of the treatment. Thus the observational data collected by time sampling procedures is used to assess the efficacy of the treatment. In educational settings the treatment is often some form of behaviour modification program. In clinical settings the treatment may be any of a wide range of techniques which, if effective, would result in a change in behavioural pattern of the subject.

The implication of the present study is that, presuming the usual time sampling approach under which
the length of interval (I) and gap (G) are held constant for the duration of the study, the accuracy of the observational data will be a function of the change in behaviour. It follows that using observational data to assess the degree of change produces data of differing accuracy over the period of the study if in fact a change occurs. It is quite possible that the observational data may contain enough inaccuracy to either obscure a change in true behaviour pattern or to spuriously suggest a change in true behaviour pattern when in fact there was no change.

Consider the following hypothetical situation in which the therapist's object is to reduce the rate of occurrence of a behaviour. The therapist decides to use time sampling to collect data upon which to base an assessment of the efficacy of the therapy, and selects a time interval of five seconds (I = 5) and a gap of five seconds (G = 5) for collecting the data. To facilitate the arithmetic involved, the expression

$$\frac{0}{T} = a + b\frac{\bar{x}_b}{(I+G)}$$

will be simplified to read
\[
\frac{O}{T} = \frac{X_b}{(I+G)}
\]

This simplification does not affect the nature of the relationship reported in Chapter 5, but does affect the estimate of accuracy obtained. For illustrative purposes the interest is in the trends rather than the actual value of \(O/T\), so the calculated value of \(O/T\) need not be exact.

Assume that before treatment the mean duration of behaviour is five seconds. The predicted accuracy of time sampling under these conditions is 0.5. In other words the estimate of frequency of behaviour obtained before treatment under these conditions will underestimate the true frequency by 50% (i.e. \(0 = T/2\)), and therefore the rate of occurrence will be underestimated by 50%. If the true rate of occurrence of behaviour before treatment was 20 occurrences per hour, the therapist in this hypothetical example would obtain an observed rate of 10 occurrences per hour. Assume further that the treatment had two effects on the subject's true behaviour pattern: (a) the rate of occurrence fell from 20 per hour to 5 per hour, (b) the mean duration of occurrence rose from 5 seconds to 20
seconds. The first effect of the treatment, the reduction in the rate of occurrence, suggests that the treatment was successful. However, because of the change in mean duration of occurrence the therapist in this example would obtain an observed rate after treatment of 10 occurrences per hour, which is identical with his observed rate before treatment. The therapist would presumably conclude that the treatment had no effect, when in fact it reduced the true rate of occurrence from 20 per hour to 5 per hour.

It is not difficult to construct similar hypothetical situations in which the treatment did not have the desired effect, but because of changes in mean duration of behaviour, data obtained using time sampling procedures would indicate that the treatment did have the desired effect.

The second implication to be drawn from the results of this study, therefore, is that if time sampling procedures are to be used to assess change of behaviour, the length of the observation interval (I) and the gap between intervals (G) should be adjusted over the period of the study in line with the changes in mean duration of behaviour.
APPENDIX A

Listing of Fortran Program BEHGEN
PROGRAM NAME  BEHGEN

AUTHOR  TIM MATTHEWS

LANGUAGE  FORTRAN 77 (USING ONLY FORTRAN 66 SUBSET)

HARDWARE  IBM 3081

FILES REQUIRED

FILE 6 (FT06F001) = OUTPUT FILE FOR SIMULATED DATA
FILE 9 (FT09F001) = OUTPUT FILE FOR ACTUAL PARAMETERS

SUBROUTINES REQUIRED

IMSL SUBROUTINE GGUBFS
IMSL SUBROUTINE GGQNF

PURPOSE OF PROGRAM  THIS PROGRAM GENERATES 100 SIMULATED BEHAVIOURS UNDER RANDOMLY SELECTED PARAMETERS. IT THEN SIMULATES THREE TYPES OF TIME SAMPLING FOR EACH OF THE SIMULATED BEHAVIOUR PATTERNS UNDER RANDOMLY SELECTED VALUES OF TIME INTERVAL AND GAP BETWEEN TIME INTERVALS. THE SUMMARY RESULTS OF ACTUAL PARAMETERS USED AND NUMBER OF OBSERVED OCCURRENCES FOR EACH TYPE OF TIME SAMPLING IS OUTPUT TO FILE 6. A SUMMARY OF THE NOMINAL AND ACTUAL PARAMETERS CAN OPTIONALLY BE OUTPUT TO FILE 9 (SET NOMOUT = 1).
<table>
<thead>
<tr>
<th>Line</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>00360</td>
<td>C</td>
</tr>
<tr>
<td>00370</td>
<td>C</td>
</tr>
<tr>
<td>00380</td>
<td>C</td>
</tr>
<tr>
<td>00390</td>
<td>C</td>
</tr>
<tr>
<td>00400</td>
<td>C</td>
</tr>
<tr>
<td>00410</td>
<td>C Main Variables Used</td>
</tr>
<tr>
<td>00420</td>
<td>C</td>
</tr>
<tr>
<td>00430</td>
<td>C</td>
</tr>
<tr>
<td>00440</td>
<td>C</td>
</tr>
<tr>
<td>00450</td>
<td>C</td>
</tr>
<tr>
<td>00460</td>
<td>C</td>
</tr>
<tr>
<td>00470</td>
<td>C</td>
</tr>
<tr>
<td>00480</td>
<td>C</td>
</tr>
<tr>
<td>00490</td>
<td>C</td>
</tr>
<tr>
<td>00500</td>
<td>C</td>
</tr>
<tr>
<td>00510</td>
<td>C</td>
</tr>
<tr>
<td>00520</td>
<td>C</td>
</tr>
<tr>
<td>00530</td>
<td>C</td>
</tr>
<tr>
<td>00540</td>
<td>C</td>
</tr>
<tr>
<td>00550</td>
<td>C</td>
</tr>
<tr>
<td>00560</td>
<td>C</td>
</tr>
<tr>
<td>00570</td>
<td>C</td>
</tr>
<tr>
<td>00580</td>
<td>C</td>
</tr>
<tr>
<td>00590</td>
<td>C</td>
</tr>
<tr>
<td>00600</td>
<td>C</td>
</tr>
<tr>
<td>00610</td>
<td>C</td>
</tr>
<tr>
<td>00620</td>
<td>C</td>
</tr>
<tr>
<td>00630</td>
<td>C</td>
</tr>
<tr>
<td>00640</td>
<td>C</td>
</tr>
<tr>
<td>00650</td>
<td>C</td>
</tr>
<tr>
<td>00660</td>
<td>C</td>
</tr>
</tbody>
</table>

**Main Variables Used**

- `BEH` Array of durations of occurrences and non-occurrences
- `AVBN` Nominal mean duration of behaviour
- `AVGN` Mean duration of behaviour
- `SDBN` SD of duration of behaviour
- `SDGN` SD of gap
- `TRUE` Actual number of occurrences of behaviour
- `INTV` Integer form
- `AVBA` Mean duration of behaviour
- `AVGA` Mean duration of gap
- `SDBA` SD of duration of behaviour
- `SDGA` SD of gap
- `GAP` Length of gap for time sampling
- `NOMENT` Momentary time sampling
- `TOTAL` Total duration of behaviour in minutes
- `PROP` Proportion of time for which behaviour occurs
- `SEQN` Sequence number of behaviour pattern
- `INDEX` Index to behaviour array
- `HINTS` Number of elements in behaviour array
- `RSEED` Double precision seed for random number generator
- `NOJOUT` Flag to optionally write parameters used to file 9
VARIABLES SETTING MIN AND MAX VALUES OF MAIN VARIABLES

AVBNHI MAX VALUE OF AVBN
AVBNLO MIN " " "
AVGNHI MAX " " AVGN
AVGNLO MIN " " "
TRUENI MAX " " TRUEN
TRUENL MIN " " "
SDBNHI MAX " " SDBN
SDGNHI MAX " " SDGN
AVBAHI MAX " " AVBA
AVGAHI MAX " " AVGA
SDBahi MAX " " SDBA
SDGAHI MAX " " SDGA
GAPHI MAX " " GAP
INTVHI MAX " " INTV
MAXE MAX " " MAXE
MAXP MAX " " MAXP
MAXN MAX " " MAXN
00880 C
START OF MAIN PROGRAM

DIMENSION DEH(5000)
DOUBLE PRECISION DSEED, AVBA, AVGA, SDBA, SDGA

FORMAT(1X, I4, F6.3, F5.2, 4F8.2, I2, I5, F4.0, P5.0, I6)
FORMAT(8F8.2, I5, F4.0, P5.0)

INITIALIZE VARIABLES

DSEED = 123457.0
AVBA = 0
AVGA = 0
SDBA = 0
SDGA = 0

GAPHI = 60.0
ANTVHI = 1200
MAXW = 300
MAXP = 300
MAXH = 300
START MAIN LOOP - 1 ITERATION PER BEHAVIOUR PATTERN

NBEH=0
NBEH=NBEH+1
IF(NBEH.GT.1000) GO TO 999

GENERATE NOMINAL VALUES FOR THIS BEHAVIOUR PATTERN

AVBN=AVBNHI*GGUBFS(DSEED)
IF(AVBN.LT.AVBNLO) GO TO 10
AVGN=AVGNHI*GGUBFS(DSEED)
IF(AVGN.LT.AVGNLO) GO TO 15
TRUEN=(TRUENH+1.0)*GGUBFS(DSEED)
IF((TRUEN.LT.TRUENL).OR.(TRUEN.GT.TRUENH)) GO TO 20
TRUEN=AINT(TRUEN)
SDBN=SDBNHII*GGUBFS(DSEED))
SDGN=SDGNHI*(GGUBFS(DSEED))
NTRUE=IFIX(TRUEN)

INITIALIZE VARS FOR TOTAL LENGTH OF OCCS AND NON-OCCS.

TOTB=0.0
TOTG=0.0
**GENERATE LENGTH OF INSTANCES OF OCCURRENCE AND NON-OCC.**

FOR THIS BEHAVIOUR PATTERN FROM NORMAL DISTRIBUTION

**DO 130 I=1,NTRUE**

**X=GGNQF(DSEED)**

**X=X*SDGN+AVGN**

**IF(X.LE.0.0)GO TO 110**

**NTERV=(2*I)-1**

**BEH(NTERV)=X**

**TOTB=TOTB+X**

**DO 120 I=1,NTRUE**

**X=GGNQF(DSEED)**

**X=X*SDGN+AVGN**

**IF(X.LE.0.0)GO TO 120**

**NTERV=2*I**

**BEH(NTERV)=X**

**TOTG=TO TG+X**

**CONTINUE**

**TOTBEH=(TOTB+TOTG)/3600.0**

**PROP=TOTB/(TOTB+TOTG)**

**NINTS=2*NTRUE**

**CALL AVER(BEH,1,NINTS,TRUE,AVBA,SDNA)**
CALCULATE ACTUAL PARAMETERS - MEAN AND SD OF NON-OCC OF BEH

CALL AVER(BEH,2,HINTS,TRUEN,AVGA,SDGA)

TEST THAT ACTUAL MEAN AND SD ARE IN DESIGNATED RANGE

IF(AVBA.GT.AVDAHI)GO TO 10
IF(AVGA.GT.AVGAHI)GO TO 10
IF(SDBA.GT.SDBAHI)GO TO 10
IF(SDGA.GT.SDGAHI) GO TO 10

GENERATE INTERVAL AND GAP FOR SIMULATING TIME SAMPLING STUDY

GAP=AINT(GAPHI * GGUBFS(DSEED))
INTV=FLOAT(INTVHI) *(GGUBFS(DSEED))
IF(INTV.LT.1)GO TO 200

INITIALIZE VARIABLES FOR OUTPUT OF SIMULATION OF TIME SAMPLING

MWHOLE=0
MPART=0
MORIANT=0
CALL SUBROUTINE WHICH SIMULATES TIME SAMPLING STUDY

CALL SAMPLE(BEH, INTV, GAP, NINTS, MWHOLE, MPART, MOMENT)

TEST THAT NO. OF OBSERVED OCCURRENCES IS IN DESIGNATED RANGE

IF(MWHOLE.GT.MAXU)GO TO 10
IF(MPART.GT.MAXP)GO TO 10
IF(MOMET.GT.MAXM)GO TO 10

WRITE ACTUAL PARAMETERS AND OBSERVED NO. OF OCCCS. TO FILE 6

HDEF=1
WRITE (6,6002) NBEH, TOTBEH, PROP, AVBA, AVGA, SDAB, SDGA, HDEF, INTV, * GAP, TRUEN, MWHOLE

HDEF=2
WRITE (6,6002) NBEH, TOTBEH, PROP, AVBA, AVGA, SDAB, SDGA, HDEF, INTV, * GAP, TRUEN, MPART

HDEF=3
WRITE (6,6002) NBEH, TOTBEH, PROP, AVBA, AVGA, SDAB, SDGA, HDEF, INTV, * GAP, TRUEN, MOMENT

OPTION TO WRITE ACTUAL AND NOMINAL PARAMETERS TO FILE 9

IF(NOMOUT.EQ.0)GO TO 900
WRITE (9,9001) AVBN, AVBN, SDEN, SDAB, AV3N, AV3N, SDGN, SDGA, INTV, * GAP, TRUEN

GO TO 5
STOP
END
SUBROUTINE TO CALCULATE MEAN AND SD OF LENGTH OF BEHAVIOUR OR GAP

SUBROUTINE AVEB(BEH,N,NINTS,TRUEN,AVE,SD)

DIMENSION BEH(5000)
DOUBLE PRECISION SUMX,SUMX2,AVE,SD
SUMX=0.0
SUMX2=0.0
DO 10 J=N,NINTS,2
  SUMX=SUMX+DBLE(BEH(J))
  SUMX2=SUMX2+DBLE(BEH(J))**2
AVE=SUMX/DBLE(TRUEN)
SD=SUMX2/DBLE(TRUEN) - AVE**2
IF(SD.LT.0.0) SD=0.0
SD=DOSQRT(SD)
RETURN
END
SUBROUTINE TO SIMULATE TIME SAMPLING STUDY

SUBROUTINE SAMPLE(BEH,INTV,GAP,MINTS,HWHOLE,MPART,MOMENT)

DIMENSION BEH(5000)

INT=1
WHOLE=0
MPART=0
MOMENT=0
SINT=FLOAT(INTV)
STARTI=0.0
STOPI=SINT
STARTB=0.0
STOPB=BEH(INT)
03060  10  MP=0
03070  IF((STARTI.GE.STARTB).AND.(STARTI.LE.STOPB)).OR.
03080        *( (STOPI.GE.STARTB).AND.(STOPI.LE.STOPB)).OR.
03090        *( (STARTI.LE.STARTB).AND.(STOPI.GP.STOPB)))  MP=1
03100  IF(MP.EQ.0) GO TO 20
03110  MPART=MPART+1
03120  IF((STARTI.GE.STARTB).AND.(STOPI.LE.STOPB)) MWHOLE=MWHOLE + 1
03130  IF(STOPI.LE.STOPB) MOMENT=MOMENT + 1
03140  STARTI=STOPI + GAP
03150  STOPI =STARTI + SINT
03160  GO TO 10
03170 20  IF(STARTI.GT.STOPB) GO TO 30
03180  STARTI=STOPI + GAP
03190  STOPI =STARTI + SINT
03200  GO TO 10
03210 30  INT=INT + 1
03220  IF(INT.GE.WINTS) RETURN
03230  STARTB=STOPB + BEH(INT)
03240  INT=INT + 1
03250  STOPB=STARTB + BEH(INT)
03260  GO TO 10
03270  END
BIBLIOGRAPHY


