The Role of the Sampling Distribution in Developing Understanding of Statistical Inference

Kay Lipson
Swinburne University of Technology

2000

Abstract

There has been widespread concern expressed by members of the statistics education community in the past few years about the lack of any real understanding demonstrated by many students completing courses in introductory statistics. This deficiency in understanding has been particularly noted in the area of inferential statistics, where students, particularly those studying statistics as a service course, have been inclined to view statistical inference as a set of unrelated recipes. As such, these students have developed skills that have little practical application and are easily forgotten.

This thesis is concerned with the development of understanding in statistical inference for beginning students of statistics at the post-secondary level. This involves consideration of the nature of understanding in introductory statistical inference, and how understanding can be measured in the context of statistical inference. In particular, the study has examined the role of the sampling distribution in the students' schemas for statistical inference, and its relationship to both conceptual and procedural understanding. The results of the study have shown that, as anticipated, students will construct highly individual schemas for statistical inference but that the degree of integration of the concept of sampling distribution within this schema is indicative of the level of development of conceptual understanding in that student. The results of the study have practical implications for the teaching of courses in introductory statistics, in terms of content, delivery and assessment.

Acknowledgements

This thesis would not have been possible without the invaluable support of my colleagues, friends and family. I would like to thank in particular Professor Peter Jones, whose interest and enthusiasm for the task from the beginning were inspirational, and whose insightful analysis ensured that the research was systematic and logical. My thanks go also to Mr Brian Phillips, who showed considerable faith both in the value of the research, and in my ability to pursue it to its conclusion. I would also like to thank Prof Barbara van Ernst and Mr Steve Weal for their encouragement and support. Finally, I would like to thank my students for their patience, tolerance and willingness to participate in this study.

Declaration

I, Kay Louise Lipson, declare that this thesis contains no material which has been accepted for the award to the candidate of any other degree or diploma, except where due reference is made in the text of the thesis. To the best of my knowledge, this thesis contains no material previously published or written by another person except where due reference is made in the text of the thesis. Where the work is based on joint research or publications, the relative contributions of the respective authors has been disclosed.

Table of Contents

Table of Figures

List of Tables

Chapter 1
Thesis Overview

1.1 Introduction

"Statistics is choosing how best, how most efficiently, to investigate and communicate the probable truth" (Hawkins, 1990 p.25). It is a discipline which provides insight into the solution of real world problems based in the most part on the use of numerical data. And, in this information age, there is no shortage of data. However, there is a real shortage of those in the workforce who are able to evaluate the quality of that data, and then use it appropriately to inform decision making. The need for a quantitatively literate society has been widely recognised for at least the last twenty years, with the resultant proliferation of statistics subjects offered over a wide range of courses. These days, whether one is a student at post-secondary level of engineering or literature, psychology or marketing, most will be required to undertake at least one course in statistics, and many will study several courses. The underlying objective of such courses is to instill in the students a knowledge basis which will inform them as citizens and members of the workforce, real-world applicable knowledge grounded in their everyday experiences.

Such an objective, however, is often quite unrealistic. Many students are not confident in quantitative disciplines, having had previous unhappy experiences with the subject with which they usually associate statistics, mathematics (for example Allan & Lord, 1991). Add to this the practical constraints of small numbers of contact hours, large group sizes and the pressures of balancing paid work and study, and the realisation of meaningful learning in introductory statistics courses seems a remote possibility. Yet, engendering statistical understanding in such students is of paramount importance if we are to achieve informed decision making across in the professions. It is thus the belief of the author that research which enlightens the statistics education profession be carried out in order that we may foster the growth of knowledge and understanding in this discipline.

1.2 Background and statement of the problem

Whilst introductory statistics courses differ in many ways, including name, who they are for, who teaches them, the way they are taught and methods of assessment, the core content of such courses is remarkably similar. Generally, a traditional post-secondary course in statistics consist of three strands: descriptive statistics (or data analysis), probability theory and statistical inference. According to Garfield and Ahlgren (1988) the content of a typical introductory statistics course can be summarised as follows:

> *Descriptive statistics*: Measures of central tendency (mean, median, mode), measures of variability (range, variance, standard deviation), measures of position (percentiles, *z* scores), frequency distributions and graphs
> *Probability theory*: Rules (addition, multiplication), independent and mutually exclusive event, random variables, probability distribution, binomial distribution, normal distribution, sampling, central limit theorem
> *Inferential statistics*: Estimating parameters (mean, variance, proportion, correlation coefficient), testing hypotheses.
>
> (Garfield & Ahlgren, 1988 p.46)

How successful are introductory statistics courses in general? Personal experience, and a survey of the literature, suggests that success, or lack of success, may be described in terms of both the cognitive and affective domains. Certainly, in the past, the introductory statistics course has not been extremely popular. As stated by Gordon and Gordon (1992a p. vii):

> ..far too many students who take a statistics course describe it as the worst or most boring class they have experienced in college.

This feeling is reinforced by Hogg (1992 p.3), who states:

> Over the years, there has been considerable dissatisfaction expressed about the first course in statistics. Many instructors have been less than happy about the success of such courses. Students have usually been far more vehement about their dislike for the subject. Thus, far too often, introductory statistics has earned the reputation for being "the worst course that I took in college."

Additionally, there is a consensus amongst those involved in statistical education that statistics is not an area which has been in general particularly well taught. From the cognitive point of view, there is also a growing body of opinion to suggest that many courses in statistics have not been successful in promoting student understanding (Konald, 1991; Well, Pollatsek, & Boyce, 1990).

Over the years two main reasons have been advanced to explain the problems experienced in teaching and learning in an introductory statistics course. These explanations are concerned with

- the overly mathematical nature of such courses and
- the emphasis in such courses on calculation over concept.

Let us firstly consider the mathematics versus statistics dilemma. Many statistics educators feel that in the past statistics courses have been too mathematical. There has been an overemphasis on the teaching of probability, which, because of its symbolic nature, is the area most mathematicians feel more comfortable teaching (Burrill et al., 1992; Moore, 1992b). Whilst probability theory in itself is an important branch of mathematics, very little knowledge of formal probability is actually required to grasp the basic concepts in statistics (Moore, 1992a). Much of the probability currently taught in many introductory statistics courses is not only irrelevant to the statistics actually taught but, because it is intrinsically difficult for many non-traditional mathematics students, its teaching may impede rather than enhance their understanding of statistics. Garfield and Ahlgren (1988) suggest that:

> The intrusion of technical probability issues that are not likely to be understood will stall the learning process. (p. 57)

And, according to David Moore (1992a):

> …a basic statistics course should cover no more probability than is actually needed to grasp the statistical concepts that will be presented. (p.23)

While some statistics educators are of the opinion that studies in probability are crucial to statistical understanding (for example, Lindley, 1990), these educators are distinctly in the minority. Whatever the reality, there is an ongoing debate concerning the basic probability requirements of an introductory course in statistics. This issue will be discussed in more detail later in this thesis.

The second common problem with introductory statistics courses concerns the concentration on the computational aspects of the subject rather than strategies for data collection, data analysis, pattern finding and interpretation. The evaluation of formulae for arbitrary lists of numbers is not statistics. As stated succinctly by David Moore (1992a):

> Calculating the mean of five numbers is an exercise in arithmetic. Calculating the mean alcohol content of five vats of wine as they complete fermentation is statistics, particularly because we look also at the variation among the vats and compare their mean with the mean alcohol content in other years and at other vintners. (p.15)

The recognition of data as numbers in context and calculation as merely a preamble to interpretation, is a recent but fundamental change in the attitude of statistics educators.

This preoccupation with calculation in introductory statistics courses may have arisen because of concern for the limited quantitative background of many of the students taking such courses. Whilst statistics was once a course taken only by the more mathematically able students, most students currently taking an introductory statistics course have little mathematical aptitude or expertise (Tanis, 1992). As a result, most introductory statistics courses have moved away from including much of the statistical theory, which underpins inference, to become basically technique orientated "recipe book" courses. This trend can be easily verified from a review of statistics texts of the eighties and nineties (see for example, Goldman & Weinberg 1985; Devore & Peck 1986; Kirk 1990; Johnson & Bhattacharyya 1992; Bluman 1997). The inherent danger of producing students capable of performing various complex statistical tests without really knowing what they are doing is of great concern. The problem is compounded if the course includes the use of a sophisticated statistical computer package which enables the student to access an enormous range of potentially inappropriate techniques.

Statistics courses today span the whole spectrum, from those which deal solely with theories which are never applied to practical problems, to those which teach recipes for statistical tests and make no attempt to justify these tests theoretically. And both types of courses, those emphasising the mathematics behind the statistics, and those emphasising the calculations which go into the statistics, can fail to develop understanding of the concepts of statistics within the students. According to Shaughnessy (1992), current courses in statistics offered at most universities:

> ...continue to be either rule-bound, recipe type courses for calculating statistics, or overly mathematical introductions to statistical probability that were the norm a decade ago. (p.466)

Thus, the challenge for statistics educators has become how to structure and teach a course in introductory statistics that promotes an understanding of statistical concepts. A recent move to develop understanding in non-mathematically trained students of introductory inference has stemmed in part from the developments in technology.

There exists now a large volume of statistical software which has been designed to facilitate understanding in introductory inference (see for example Biehler, 1993; Cohen et al, 1994; Shaughnessy, 1992). Teaching for understanding is the current objective (Moore, 1992a, 1992b).

In summary, while has become a widely taught subject, it enjoys limited success with students both in terms of their attitude to, and understanding of, the subject. Whilst professional journals are full of suggestions as to how to improve the current situation, actual research in the area of statistics education is relatively new (Shaughnessy, 1992). Further research is needed concerning the source and nature of the problems which the students experience in the development of understanding of introductory statistics, in order to offer suggestions to the profession for the solution of these problems.

1.3 The focus of this study

The current study has focussed on the transition from data analysis to statistical inference, a known area of difficulty for students (see, for example Garfield & Ahlgren, 1988, Gordon & Gordon, 1992a).

Personal experience suggests that, whilst many students show reasonable levels of understanding of data analysis, the concepts that underlie introductory statistical inference are a mystery to most, and remain so after they have completed most courses in statistics. On numerous occasions I have seen students able to proceed step-by-step through a complex inference problem only, at the very last stage, to show their lack of understanding of what is involved by drawing an incorrect conclusion based on the P-value. I suspect the words "I can never remember whether I reject or accept the null hypothesis if the P-value is less than 0.05" have been heard in statistics classrooms hundreds of thousands of times!

What are the key concepts which underpin understanding in statistical inference? Traditionally, the concept of sampling distribution has been seen as fundamental to an understanding of introductory statistical inference. Several researchers have diagnosed student misconceptions concerned with the sampling process itself, on which the theory of statistical inference is based (Green, 1982; Rubin, Bruce, & Tenney, 1990; Tversky & Kahneman, 1971; Tversky & Kahneman, 1982), and this is confirmed by this researcher's own experience. Is sampling distribution a key concept, underpinning an understanding of introductory statistical inference? Is it possible to develop a teaching strategy which addresses the key concepts, and hence supports the development of understanding in statistical inference for a non-mathematically trained group of students, as well as the necessary technical competence which they need? And, further to this, what is the role for technology within this strategy? These are the central issues which direct this study. The emphasis here is on the development of a teaching strategy which is classroom focussed, and has practical application in a post-secondary environment.

1.4 Theoretical framework

This thesis is concerned with the development of understanding in students of introductory statistical inference. In order to carry out a study concerning the development of understanding, it is necessary to define understanding in the context of the study.

The theoretical framework for this research views the learning process as an interaction between the student learning experiences and their existing mental structures, mediated by the social and cultural context within which the experiences take place. Knowledge is stored in long term memory in mental structures which are constructed and organised by the individual, with learning seen as modification of these mental structures (Ausubel, Novak, & Hanesian, 1978; Piaget, 1970; Skemp, 1971). According to this model, understanding may be interpreted as the degree to which this knowledge is accessible to the individual in a variety of contexts, indicating the level of connectedness of the mental structures constructed by the learner (Hiebert & Carpenter, 1992; Marshall, 1995). Thus, those students who have not developed an adequate understanding of statistical inference may be seen to have constructed mental structures of statistical inference which are incomplete or incorrect. That is, they either exclude concepts which are fundamental to flexible application of the knowledge, or the links between concepts are incorrect, inappropriate or missing.

Applying this model of understanding to an analysis of the content domain of statistical inference leads to a theory concerning the form of the mental structure that is likely to support understanding in statistical inference. In particular, the study is concerned with the role of the sampling distribution in the mental structure of the students, and the relationship between this mental structure and the level of understanding of statistical inference which the student demonstrates. Consideration of these issues forms the basis of the research questions which are addressed through this research.

1.5 Research Hypothesis

This study is concerned with the content and form of the mental structures that are likely to be associated with the development of conceptual and procedural understanding in students of introductory statistical inference.

The specific hypothesis addressed by the research can be stated as follows:

> Students whose schema for sampling distribution demonstrates links to the sampling process and whose schema for statistical inference includes links to the sampling distribution, will show evidence of both conceptual and procedural understanding of statistical inference, whilst those who have not integrated these concepts into their cognitive structure may still demonstrate high levels of procedural understanding, but not of conceptual understanding.

1.6 Design and Methodology

In order to address the questions outlined in the previous section a teaching experiment methodology (Schoenfeld, 1994; Steffe, 1991) applied to the post-secondary teaching environment was adopted. The methodology of the teaching experiment was appropriate for this study, as in order to address the research questions it was necessary to document the nature of the students' mental structures during and after an appropriate course of instruction. These data would enable the researcher to investigate the nature of any qualitative differences observed between students' mental structures and also to address the predictions made concerning the level of understanding developed by students and their mental structures associated with statistical inference.

The course chosen for the study was a post-graduate introductory statistics course, undertaken by students from a broad range of essentially non-mathematical backgrounds. The subjects of the study were 23 mature-age students, all studying part-time. The teaching/learning strategy employed was based on a theoretical analysis of the knowledge domain of statistical inference, and the constructivist view of how people learn. In particular, the teaching/learning strategy recognised the multiplicity of mathematical and symbolic representations available for the concepts associated with statistical inference, aiming to make explicit, using a variety of media, the links between these representations.

A preliminary study undertaken with a group of students enrolled in the same course in the previous year (N=35) in order to trial and evaluate tools for externalising the mental structures of students. This would enable the researcher to investigate changes in mental structures of the students' occurring as students progressed through the course. The preliminary study was also used to trial and evaluate instruments used to measure students' resultant level of understanding in statistical inference.

The main study included both qualitative and quantitative components. Data were collected on the background training and knowledge of the students, along with some demographic information. During the conduct of the teaching experiment, the researcher documented the ways in which students constructed, organised and reconstructed their knowledge, and the ways in which new knowledge was integrated into their existing mental structures. A variety of tasks were used to gain some insight into the students' levels of understanding in inferential statistics at the end of the period of instruction. Factor analysis was used to develop constructs from these tasks that measured different aspects of student understanding, procedural understanding and conceptual understanding. The relationship between the form of the student's schemas related to statistical inference and the levels of procedural and conceptual understanding in statistical inference demonstrated by the student could thus be addressed.

1.7 General Conclusions

The findings of this research can be summarised as follows. Firstly, it was shown by investigation of the mental structures constructed by students that these structures were qualitatively different, confirming the individuality of the student's experience and their resultant mental structures. The mental structures were also observed to be dynamic, evidencing both the growth of understanding and the development of misconceptions over time. On the basis of the observed mental structures, students were divided into three groups:

Group 1	Students whose mental structures showed evidence of the development of the concept of sampling distribution and subsequently integrated sampling distribution appropriately with their mental structure for statistical inference.
Group 2	Students whose mental structures showed evidence of the development of the concept of sampling distribution but did not relate this to their mental structure for statistical inference.
Group 3	Students who did not show evidence of the development the concept of sampling distribution.

Each student was also allocated separate scores in procedural understanding and conceptual understanding. As hypothesised, the role of the sampling distribution was found to be fundamental in developing understanding in inferential statistics. Those students allocated to Group 1 scored higher on the measure of conceptual understanding than those in Group 2, who in turn scored higher than those in Group 3. As also hypothesised, no significant difference was found between the three groups on the procedural measure.

The empirical analysis identified differences in the level of conceptual understanding demonstrated by the students. In order to investigate the reasons for the differing levels of performance, case study analyses were carried out using two students from each of the three groups. These analyses investigated the relationship between the form of the students' mental structures and the subsequent performance measures, and confirmed the important role of the sampling distribution within the mental structure for statistical inference.

1.8 Thesis Organisation

This thesis consists of eight chapters, references and appendices.

Chapter 1 gives an overview of the thesis. It gives some background to the problem, and describes the need for such research. It also includes a brief description of the theoretical framework, the research questions, an overview of the methodology and a summary of the results of the study.

Chapter 2 is a review of the literature as it pertains to the development of understanding. It contains sections concerned with the nature and process of knowledge construction, the meaning of understanding, issues of representation, implications of this theory for teaching and learning, and the potential role of technology within this theoretical framework.

Chapter 3 is a theoretical analysis of the knowledge domain that is encompassed by statistical inference. It also catalogues the mathematical and symbolic representations that are related to this knowledge domain. The desirable features of the conceptual structures which are constructed by experts within this knowledge domain are established, and the potential for computer intensive methods to support student learning is assessed. Together, these analyses lead to the specific hypothesis of the study. A rationale for the use of concept maps as external representations of mental structures in this study is also presented.

Chapter 4 describes the teaching experiment methodology used in the study. The instruments used to evaluate student knowledge structures and determine subsequent success in statistical inference are presented and argued. An evaluation of the data collection instruments developed, based on the results of the preliminary study, is included.

Chapter 5 gives the results of the study at a group level. This includes an evaluation of students' conceptual structures based on a comparison of their concept maps with those prepared by experts in the area, and the relationship between these mental structures and subsequent performance on a range of statistical tasks designed to measure conceptual and procedural understanding.

Chapter 6 documents the case studies analysis, providing a detailed analysis for six students, in order to investigate more fully the nature of the relationship between the form of the student's schema and performance of the outcome measures which was established through the analysis in Chapter 5.

Chapter 7 looks at the implications of the results of the study for teaching and learning statistical inference. In particular, the findings are related to the structure and content of traditional courses, the potential for the use of technology in teaching statistical inference, and the role of assessment in facilitating and evaluating meaningful learning.

Chapter 8 summarises the findings and describes the researcher's reflections on the study as well as recommendations for future research in the area.

References and appendices follow Chapter 8.

Chapter 2
The Theoretical Framework

2.1 Introduction

Statistical inference is recognised by many statistics educators as proving difficult for large numbers of students (for example Well, Pollatsek & Boyce, 1990; Konald, 1991). However, suggestions from the literature to date have generally focussed on recommending alternative teaching strategies, often incorporating the use of computer technology (for example Adams & Stephens 1991; Gordon & Gordon 1992b; Pukkila & Putanen 1986) in order to help students overcome difficulties with the concepts of statistical inference. Some of these suggestions will be discussed later in this chapter.

The aim of this study is to investigate the cognitive development of students beginning a study of statistical inference, in order to examine

- how students' mental structures change over the period of instruction,
- whether or not there are any qualitative differences in the mental structures of students, and
- if a student's subsequent level of understanding in inferential statistics can be related to the form of their mental structure. That is, are there concepts which when assimilated into the students' cognitive structure indicate which students will be more successful in demonstrating an understanding of elementary statistical inference?

In order to proceed with the study, the nature of the students' mental structures needs to be established, as well as a framework for defining and evaluating student understanding in statistical inference. These issues are the main themes of this chapter, and each will be investigated by analysis of the literature.

Much of the literature that has been synthesised for this chapter has arisen from the study of the development of understanding in mathematical concepts. Although it may be argued that there are fundamental differences between mathematics and statistics (Moore, 1992b), the themes to be developed here are considered to be fundamental to the development of both mathematics and statistics concepts.

In Section 2.2 the constructivist view of learning is introduced, whilst in Section 2.3 models for the development and storage of knowledge are discussed. In Section 2.4 literature pertaining to understanding is reviewed, and a framework presented for discerning different aspects of understanding. In Section 2.5 various meanings of the terms representation are identified, and the relationship between them discussed. Arguments are also presented which identify and validate the concept map as a legitimate tool for externalising a student's mental structure. In Section 2.6 the principles of the cognitive view of learning and understanding are applied to the development of an instructional strategy, with reference to the potential role of current technology in such a strategy.

2.2 The nature of knowledge construction

Cognitive development refers to the intellectual development of an individual. The cognitive approach to understanding intellectual development embodies three important characteristics (Howard, 1983):

1. The emphasis is on what an individual knows, as a mental process, rather than on how the individual behaves. The goal is an understanding of the mind and how it works, which is necessarily not directly observable.
2. An individual's knowledge is organised into a (highly personal) mental structure, which is then called into play when that individual responds to external stimuli. This notion of the organisation of a mental structure was first proposed by Piaget, who named this mental structure a *schema* (Piaget, 1970).
3. The individual is an active participant in the construction of the knowledge, rather than a passive recipient of it. According to Howard (1983):

> The individual is thought of as actively constructing a view of reality, selectively choosing some aspects of experience for further attention, and attempting to commit some information to memory. (p. 6)

Using these defining tenets as a starting point leads one to a contructivist view of learning. That is, students are not regarded as passive receivers of information, but as active constructors of their own knowledge that is built up as they participate in experiences which are socially and culturally situated. Thus, what one knows can be considered to be a product of an individual's perception of an external experience, and how that experience has influenced or has been influenced by their existing cognitive structure. An important feature of contructivism is the emphasis on understanding. The goal of the learning process is not only that the student is able to carry out a task, but that he or she is also undergoing appropriate cognitive development.

2.3 Cognitive structures

Knowledge stored in long term memory may be considered to be organised into mental structures (Sweller, 1993). What is the nature of these structures, and how do they change as part of the learning process? These questions are important for the present research, as the answers to these questions will lead to a theoretical framework within which the data collected in this study may be interpreted.

A *concept* is the abstraction of the important features of an experience which enables a new experience to be recognised as belonging to that class of experiences (Skemp, 1987). In terms of developing advanced mathematical concepts, Tall and Vinner (1981, Vinner, 1983) have found it useful to differentiate between the *concept definition* (the taught concept) and the *concept image*, which describes the student's personal concept. The concept image for a concept is the set of all pictures that have ever been associated with the concept in a particular person's mind together with the set of properties associated with a concept by an individual. The need for the notion of concept image grew from observations by Tall and Vinner that students could have complex cognitive structures associated with a concept which enabled the image of the concept and the definition of the concept to co-exist comfortably for many students, even when contradictory.

As they learn, individuals aggregate their knowledge into more refined cognitive structures, which are termed by Piaget as schemas (see Section 2.2). A schema may be considered to consist of a number of inter-related concepts (Skemp, 1971; Skemp, 1987) which have been accumulated into a more refined and complex structure.

According to Gerhard Steiner (1994):

> ..a schema is always a representational, permanently modifiable unit, a meaning structure of a particular (although restricted) scope that represents actions, operations (these latter ones as systems of internalised actions in Piaget's sense) or concepts. (p. 250)

The multitude of schema available to an individual makes up that individual's total cognitive structure, also called a *semantic network*, with the schema seen as the "activated part of the semantic network" (Steiner, 1994 p.250).

Necessarily, knowledge construction begins with some experience in the outside world. Whilst each experience may be unique, after a number of experiences the similarities may be recognised by an individual and integrated into an existing schema. According to Marshall (1995) a schema can be though of as follows:

> It is a mechanism in human memory that allows for the storage, synthesis, generalisation, and retrieval of similar experiences. (p. vii)

A cognitive structure is necessarily dynamic, since all experiences will influence the existing schemas in some way. New information is either *assimilated* into existing schema, or causes the appropriate schema to be modified in some way. This process of modification is sometimes called *accommodation* (Piaget, 1970; Skemp, 1987). Accommodation can be considered as the process of refining the schema into one that is more appropriate and hence more useful. Piaget used the term *equilibration* to indicate the satisfactory match between the new situation and new schema, from the individual perspective. According to McInerney and McInerney (1994):

> For Piaget then, cognitive development involves an interaction between assimilating old facts to new knowledge and accommodating old knowledge to new facts and the maintenance of structural equilibration. (p. 79)

In order to think about student knowledge it helps to adopt a representation for the structure of a schema. Several metaphors have been proposed for the nature of this structure. Since an essential property of a schema is that, when one piece of information concerning a situation is accessed from the schema, all of the other concepts associated with that schema are available, it makes sense to think of the schema as a connected network of concepts. The notion of a schema as a connected network of concepts fits well with the role that the schemas play in the construction of knowledge. Marshall (1995)summarises this succinctly when she says:

> A distinctive feature of a schema is that when one piece of information associated with it is retrieved from memory, other pieces of information connected to the same schema are also activated and available for mental processing. (p. vii)

Hiebert and Carpenter suggest two metaphors that may be useful when thinking about schemas (Hiebert & Carpenter, 1992). Their first metaphor is a network structured hierarchically, with successively higher stages in the hierarchy representing more generalised concepts. Their second metaphor is that of a spider's web, where each cross-link is characterised by a concept, and the threads represent relationship between concepts. Some relationships are direct, and some are indirect through other concepts, making all nodes in the web connected. This web-like metaphor fits well with the notion of semantic networks, and their external representation through concept mapping which will be discussed in more detail later in this chapter.

2.4 What does in mean to understand?

This thesis is concerned with the development of understanding in introductory statistical inference. In order to determine the level of understanding developed by students, it is necessary firstly to consider what is meant by understanding and on the basis of this to develop appropriate instruments with which to assess understanding in statistical inference.

Consider firstly the meaning of understanding in terms of the conceptual structures described in the previous section. In the words of Skemp (1987):

> To understand something means to assimilate it into the appropriate schema. (p. 29)

Hiebert and Carpenter (1992), further develop this view of understanding in relation to schema formation:

> A mathematical idea or procedure or fact is understood if it is part of an internal network. More specifically, the mathematics is understood if the mental representation is part of a network of representations. The degree of understanding is determined by the number and strength of connections. A mathematical idea, procedure, or fact is understood thoroughly if it is linked to existing networks with stronger or more numerous connections. (p. 67)

Marshall (1995) also suggests understanding is related to connections between concepts:

> The degree of connectivity among the schema's components determines its strength and accessibility. (p. 46)

Others who consider connections between representations as indicative of understanding include Hiebert & Lefevre (1986), Kaput (1992) and Davis (1986).

This metaphor for understanding is consistent with the constructivist position that learning is necessarily a highly individual experience, largely dependent on what the learner already knows:

> The processes of reorganising networks and adjoining new representations to existing networks both depend, to some degree, on the networks that have already been created. (Hiebert & Carpenter, 1992, p. 70)

Putnam, Lampert and Peterson (1990) consider that this basic idea of understanding, described purely in terms of the cognitive structure, can be expanded and operationalised in relation to understanding in mathematics. To this end they described five key themes around which understanding in mathematics can be explicated. These are:

1 *Understanding as representation*. Mathematics as a discipline has developed a powerful and efficient symbol system and notations for representing mathematical ideas. These may be referred to as external representations, as they exist outside the learner. The cognitive structures, which have been discussed earlier in this chapter, are internal representations, specific to the individual. These internal representations cannot be observed, and so any match between the external representation and the internal representation is purely speculative. However, in order to be able to make progress in this area, it is useful to consider certain external constructs as evidence of particular internal mental representations. Other researchers, such as Janvier (1987), view understanding in terms of the ability of the student to move between the various external representations, thus evidencing the internal connectness of the appropriate schemas.

2 *Understanding as knowledge structures*. By detailed study of the concepts involved in some aspect of mathematical learning, together with the way in which learners go about applying such mathematical knowledge, researchers are constructing models of the knowledge structures which would reflect student understanding. In this way, the extent of a learner's understanding can be evaluated, as the degree to which the student knowledge structure agrees with the theoretical analysis of the knowledge content can be seen of indicative of student understanding.

3 *Understanding as connections amongst types of knowledge*. As previously mentioned, much of what takes place in a mathematics classroom is couched in a language and notation system which is not seen or used in any other situation. And yet, many of the concepts which are studied in mathematics have application outside the classroom and mathematical problem solving skills are already used by many students in the real world. These two types of knowledge, relating to mathematics inside and outside the classroom, have been termed *formal* and *informal* knowledge. The ability to relate formal and informal knowledge, to interpret the real world problem in terms of the mathematical theories and vice-versa, can also be taken as evidence of understanding of the mathematical knowledge in question.

4 *Understanding as the active construction of knowledge*: As already stated, the constructivist view of learning recognises that the learner constructs his or her knowledge on the basis of what he or she already knows. Thus, a teacher cannot assume that the curriculum as presented equates to what the student will learn. Acknowledging then the individuality of the student experience, consideration of understanding may need to focus more on how the student's conceptual structures have changed, rather than what the student can be determined to know of the taught curriculum.

5 *Understanding as situated cognition*: The cognitive model for understanding described thus far proposes that understanding can be viewed in terms of the content and degree of connectivity of a student's cognitive structures. However, many educators have recognised that knowing and understanding in mathematics is also an interaction between those structures and the social and physical situation in which the knowledge is to be applied. If the schema relevant to a particular aspect of mathematics knowledge is appropriately activated in a real-world situation, then this could be considered as evidence of understanding of the concept which is represented by that schema.

The notion of understanding in mathematics (and statistics) is abstract and complex. The five themes previously described present multiple views of understanding, which together provide a framework for consideration of what it means to understand introductory statistical inference. These themes provide a basis for assessment of understanding in statistical inference, which is the topic of a later section in this thesis.

2.4.1 Procedural and Conceptual Understanding

It is well recognised by many educators and researchers that students are able to successfully complete certain mathematical tasks but at the same time demonstrate no real understanding of the concepts and rules inherent in these tasks. Skemp (1978) coined the terms *instrumental* understanding and *relational* understanding to describe these different facets of understanding, which were familiar to many mathematics educators.

Cognitive psychologists (for example Hiebert & Carpenter, 1992) have coined the terms *procedural* and *conceptual* understanding to describe this phenomenon. Procedural or instrumental understanding describes the student's ability to carry out routine tasks successfully, whereas conceptual or relational understanding implies knowledge on the student's behalf of what they are doing and why they are doing it. Nowhere is this dichotomy between procedural and conceptual understanding more evident than in statistical inference. Whilst many students are able to solve routine hypothesis testing problems, or calculate confidence intervals from standard formulae, statistical educators have no doubt that many students have little understanding of the concepts involved (Garfield & Ahlgren, 1988).

Precise definitions of procedural and conceptual knowledge, which relate these types of knowledge to the properties of the students' schemas, have been advanced by Hiebert and Carpenter (1992), who state:

> Conceptual knowledge is equated with connected networks.......A unit of conceptual knowledge is not stored as an isolated piece of information; it is conceptual knowledge only if it is part of a network. On the other hand, we define procedural knowledge as a sequence of actions. The minimal connections needed to create internal representations of a procedure are connections between succeeding actions in the procedure. (p. 78)

Conceptual knowledge is thus seen as knowledge which has been assimilated with existing knowledge, appropriately connected with existing schema. However, clearly there may be varying degrees of connection, which would imply varying levels of understanding rather than discrete categories. Hiebert and Lefevre (1986) distinguish between the *primary* level of connectedness, where the information is constructed at the same level of abstraction as the information itself, and the *reflective* level, where relations are connected at a more abstract level, and are able to be generalised to a variety of contexts.

Hiebert and Lefevre (1986) also distinguish between two types of procedural understanding. In the first type, they include the understanding of the notation system and language of mathematics. In the second, they include understanding of the rules and methods which students are trained to apply, often in a recipe like fashion. They state:

> An important feature of the procedural system is that it is structured. Procedures are hierarchically arranged so that some procedures are embedded in others as subprocedures. An entire sequence of step-by-step prescriptions or subprocedures can be characterised as a super procedure. The advantage of creating superprocedures is that all subprocedures in a sequence can be accessed by retrieving a single superprocedure. (p. 7)

The definition of procedural understanding which is adopted in this research encompasses both a knowledge of the language and notation of statistics, as well as an ability to apply the super procedures described here.

2.5 The notion of representation

Understanding in mathematics education is often discussed in terms of *representation*, and in particular multiplicity of representations. The term representation is much used in discussions of mathematics education, and clearly has different meanings when cited in different contexts. To help clarify the term, four different types of representations have been proposed by Kaput (1987). These are

(1) mental representation, the representation in the mind of an individual;

(2) explanatory representation involving the models proposed by psychologists and educators to describe the mental representation;

(3) representation within mathematics, the representation of one mathematical construct by another;

(4) external symbolic representation, the material forms used to express abstract mathematical ideas.

The *mental representation* referred to in (1) above is sometimes called an *internal* representation since it is located inside the student's head, in contrast to (4), which is designated as an external representation because it is located in the environment (von Glaserfeld, 1987). Putnam, Lampert and Peterson suggest that:

> cognitive psychology is about hypothesising the sorts of mental representations that individuals have and use. (1990 p. 68)

Whilst the mental representation held by the student may be considered the essence of a student's understanding, others are unable to know that mental representation first hand. Psychologists therefore build models or *explanatory representations* (2), to help us to know this mental representation. For example, an explanatory representation used in advanced mathematics is the concept image described earlier (Tall & Vinner, 1981; Vinner, 1983).

Issues involving the *mathematical representation* (3) of mathematics are of critical interest in statistics learning. For example, mathematical representation is a fundamental concern when introducing the concept of probability. As stated by Shaughnessy (1994), in his review:

> Perhaps the predominant theme is that the probability concept is not a single clear cut mental construct. Historically, theoretically, didactically, and psychologically there has been a tension between probability based on Laplacian equiprobable events and probability considered as the limit of relative frequencies. (p. 70)

This tension exists because probability cannot be seen only as one or other of these mathematical representations, because it is in fact both. Much discussion has taken place amongst researchers as to the relative merits of teaching probability from either the axiomatic or the relative frequency approach (for example Steinbring, 1992). Either way, it is true that educators would ultimately like both mathematical representations of the concept to be present in the students' schema for probability, and for students to be able to refer to whichever mathematical representation of probability is more helpful in a given situation.

The final type of representation proposed by Kaput (4) is that of *external symbolic representation*. In order to both think about and communicate our (internal) mental representations it is necessary to express these as (external) symbolic representations. There are many symbolic representation systems available for most mathematical concepts. One list of representations suggested by Lesh, Post and Behr (1987) is:

(a) experience based scripts (which consist of almost recipe-like procedures which student learn to follow through constant repetition);
(b) manipulable models;
(c) pictures or diagrams;
(d) spoken languages;
(e) written symbols.

An example of the variety of symbolic representations available for the different mathematical representations of a concept in statistics can be illustrated by looking at the various representations used when discussing the sample mean. The concept of the mean may be represented in several ways, for example (following the previous categorisation of symbolic representations):

(a) The mean of a set of numbers is obtained by summing the numbers and then dividing by the number of values. (*experience based scripts*)

(b) The mean of a set of numbers may be found physically using a three dimensional model, by finding the balance point when blocks are placed at the data values on a balance beam. (*manipulable models*)

(c) The mean of a set of numbers may be found by finding the balance point of a density curve (Moore & McCabe, 1993, p.62). (*pictures or diagrams*)

(d) The sample mean gives a typical value for that variable.(*spoken languages*)

(e) For the data set $\{x_1, x_2, \ldots\ldots, x_n\}$ the mean $\bar{x}$ is:

$$\bar{x} = \frac{x_1 + x_2 + \ldots\ldots\ldots.x_n}{n}$$ (*written symbols*)

These diverse representations of the mean can be considered to each embody different facets of the concept of mean, all of which are important in building a schema for the mean which is rich in relationships. The connection between these representations, and the ability to see the equivalence of them, may be considered to indicate that the student has an understanding of the mean which is both useful and correct.

As well as those representations suggested by Lesh et al, Kaput (1992) has suggested that the advent of modern computer technology has meant that new notation systems (ways of recording and /or displaying information) may be created. These notation systems allow for mathematical representations which can provide multiple, dynamic, linked images of the concept or system of concepts. These ideas are discussed in more detail later in section 2.5.1, as they relate to the learning of statistical inference.

In summary, consideration of knowledge in terms of representation means one can consider a person's knowledge of a concept as a mental representation (schema) that is described by external representations of this concept. Within a person's schema for a concept one or more mathematical representations of the concept may be present and accessible (in that they can be evoked under particular circumstances), and, within each mathematical representation, there may be several symbolic representations available to the learner. Links may or may not have been formed between the various mathematical representations and/or the symbolic representations. Increasing the number of representations present in the schema, and developing the links between them, may be considered as increasing the level of understanding in that person of the concept in question.

2.5.1 Concept maps: External Representations of Conceptual Structures

The research documented in this thesis was concerned with the nature and form of the schemas constructed by students concerning statistical inference, and how this relates to the level of understanding demonstrated by students. These schemas are not directly observable, and as a consequence, their nature cannot be explicitly known. Thus, external representations of that mental structure need to be developed. That is, ways must be found to make that internal conceptual structure explicit, or at least explicit enough to be useful to researchers and educators.

In the past Piaget used the clinical method as his technique for tapping into the student's schemas. This was an open-ended interview session, with subsequent questions determined by previous answers. Marshall (1995) also suggests researchers use structured interviews with subjects in order to gain access to their schema. However, such approaches are extremely time consuming and impractical in the everyday classroom. Fortunately techniques for providing external representations of the students' schema have been recently developed that can be employed in a classroom situation with a larger group of students.

Jonasson, Beissner and Yacci (1993) discuss several methods which may be used for obtaining external representations of an individual's schema, one of which, the concept map, is suitable for everyday application. The *concept map* arose from the work of Ausubel (Ausubel, Novak, & Hanesian, 1978) and has been developed both as a research instrument and learning tool by Novak and Gowin (1984) who describe it thus:

> A concept map is a schematic device for representing a set of concept meanings embedded in a framework of propositions. (p. 15)

An example of a student concept map is shown in Figure 2.1.

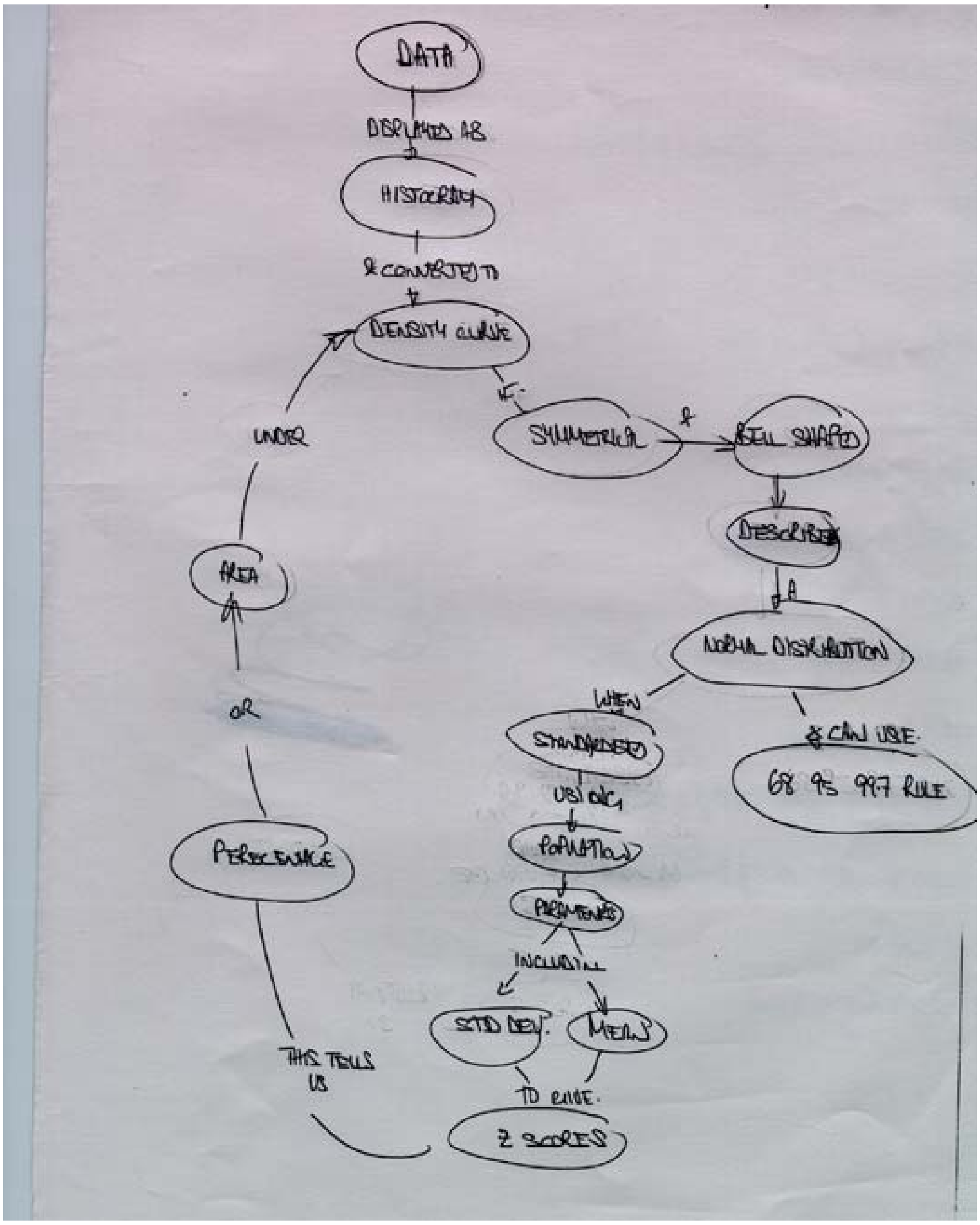


Figure 2.1 A concept map constructed by a student for the Normal Distribution

Concept maps are an explicit method for externalising the knowledge structure of a learner in a particular content domain at a given point in time. They are two-dimensional diagrams in which relationships between concepts are made explicit. By linking terms associated with a concept, propositions can be formed which indicate understanding of some aspect of the concept. According to Jonassen, Beissner and Yacci (1993):

> Ausubel believes that knowledge structures are organised hierarchically, with more inclusive concepts subsuming more detailed concepts. These concepts are defined and described through propositions, which identify relationships between concepts. Thus interrelated networks of concepts and propositions are an essential element of

> human learning. Concept maps are an explicit representation of these integrated networks.
> (p. 155)

Constructing a concept map requires the student to identify important concepts concerned with the topic, rank these hierarchically, order them logically, and recognise relationships where they occur.

Concept maps were originally used for research purposes to facilitate the study of the students' schemas before and after instruction (Novak, 1990b). Since the concept map may be considered as a representation of an existing schema, then when learning occurs and new concepts are integrated with existing knowledge, the resultant change in the student's schema will be evidenced by changes in the concept map. In a longitudinal study, researchers at Cornell University used concept maps to chart changes in cognition in a group of children over a 12 year period of schooling (Novak, 1990a).

There are numerous classroom applications of concept mapping to be found in the literature. Though largely in science education, many other disciplines have now recognised the potential of concept mapping activities. Examples include concept maps being used by teachers in curriculum planning to determine the content and structure of courses (Peterson & Treagust, 1989; Starr & Krajcik, 1990), as an assessment tool to give insight into students' understanding (Schau, 1997; Shavelson, 1993), and as a teaching strategy to facilitate the development of student understanding (in much the same way as the preparation of a summary may help) and promote discussion (Roth & Roychoudhury, 1993).

Concepts maps have continued to prove useful as a research tool, where they have been used to explore understanding and help diagnose misconceptions and document changes in understanding over a period of instruction (Wallace & Mintzes, 1990; Lipson, 1996), and compare groups of students with different backgrounds or undergoing different educational experiences (Kinnear, Glesson, & Commerford, 1985; Markham, 1994; Williams, 1995).

Using concept maps one can gain some measure of the value of particular educational experiences by looking at how students' concept maps alter as the result of these experiences. Wallace and Mintzes (1990) used concept maps as well as traditional psychometric testing to study the changes in conceptual development which took place as a result of a Computer Assisted Instruction (CAI) session with a group of elementary education majors. The students were randomly allocated into two groups, with one group of 42 students (the experimental group) participating in a CAI activity relevant to the study, whilst the other group of 49 students (the control group) undertook a similar activity on an unrelated topic. Their results showed a small difference between the groups based on the traditional testing methods. In order to compare the concept maps for the two groups, experts were also asked to prepare maps to establish a list of critical concepts. The maps were then scored using two different methods:

- a quantitative method where the number of valid propositions (1 point), hierarchies (five points), branches (1 point), cross-links (10 points) and general-to-specific examples (1 point) contribute to a numerical score for each map, and
- the number of critical concepts and propositions.

Subsequent analyses showed significant increases in the experimental group over the control group using both scoring systems. Using the number of critical concepts and propositions as a measure indicator revealed no change for the control group, whilst in the experimental group 'the number of critical concepts increased by over two-thirds and the number of propositions jumped by nearly 150%' (Wallace & Mintzes, 1990 p. 1040).

Williams (1995) used concept maps for the concept of function to compare the conceptual structure of 14 students from a remedial calculus class and 14 students from a traditional calculus class, for the concept of function. She also compared the maps prepared by both groups with those prepared by experts. Quantitative analysis of the maps indicated that the student maps from both groups showed poor evidence of the core concepts when compared to the expert maps. Qualitative analysis of the maps indicated differences between the two student groups, in terms of the sophistication of the propositions used, although both groups exhibited less well-structured maps than the experts.

Edwards and Fraser (1983) suggested that the concept map might be a more practical method of eliciting the student's cognitive structure than the interview, which has practical limitation for the classroom teacher:

> Teachers roll their eyes in disbelief at the suggestion of a regular regime of individual interviews. The restructuring required to enable such a system to operate is massive and unlikely to be widely implemented in classrooms. Concept mapping, on the other hand, offers a technique for revealing cognitive structure which appears manageable within classroom constraints. (p. 19)

They carried out a research project using twenty-four Grade 9 Science students, in order to determine how the conceptual structure as determined by the concept map compared to that resulting from an interview. They compared the propositions evidenced by the map with the knowledge subsequently revealed by the student at interview, and found the results were almost identical.

> The degree of agreement between the propositions revealed by the students' maps and the subsequent interviews is impressive…These results indicated that in this study concept maps were as accurate as interviews for revealing student comprehension of concepts. (Edwards & Fraser, 1983 p. 24)

Laturno (1994) carried out a comparable study in mathematics, working with 118 remedial mathematics students at a community college. Laturno did not analyse the student maps using propositions, but used a scoring system similar to the one advanced by Novak and Gowin (1984). However, similar results were observed:

> The results of this study would indicate that student generated concept maps show indications of validity as a research tool, giving comparable results to both interviews concerning student knowledge of relationships between concepts and academic progression of students through as individually paced course.....The connection between concept maps and clinical interviews appears quite promising as a research tool. (Laturno, 1994, p. 65)

In summary then, previous research has established that the concept map can be used to provide an external representation of the students' schema, and over the course of instruction to document any changes that may occur in that schema.

2.6 Implications of the cognitive view of learning for instructional strategies

An essential element of the cognitive view of learning is that individuals construct their own meaning from experiences based on their own particular cognitive structure, which has in turn been constructed by that individual (for example Putnam, Lampert, & Peterson, 1990). These cognitive structures, also known as schemas, provide the facts, processes and strategies that an individual accesses when attempting to interpret the external environment. The extent of the connectedness of the schemas can be taken as a measure of the individual's understanding.

In theory, learning involves either the assimilation of knowledge into an existing schema, or adaptation of the schema to accommodate this knowledge, but in practice generally both of these mental actions occur. What is not in doubt is that, in an instructional setting, the learner will construct knowledge, establishing a schema no matter what his or her goal or the form of the teaching and learning strategy implemented. According to Cobb, Yackel & Wood (1992):

> The central issue is not whether students are constructing, but the nature or quality of those constructions. (p. 28)

Can the learning environment, and the activities that take place within it, influence the nature of the learner's schema, or does the constructivist view imply that meaningful learning takes place only when students discover for themselves the characteristics of relationships between mathematical ideas? Not according to Cobb, Yackel & Wood (1992) who say:

> It should be clear from the arguments we have made that we do not believe that mathematical learning can ever be natural, if by natural we mean the unconstrained organic growth of mathematical knowledge independent of social and cultural circumstance. Further, the conclusion that teachers should not attempt to influence students' constructive efforts seems indefensible. (p. 27)

Romberg (1992) also believes that formal education experiences can encourage an individual to modify a person's existing schema. In fact, Romberg suggests that, although schema are continually changing in the light of one's experiences, both formal and informal, individuals who possess incomplete or ill–formed schemas will "actively search for experiences which provide them with structure" (1992 p.62).

Thus, it is proposed that students can be encouraged to modify, restructure and connect schemas by planned instructional interventions. What is the role of the teacher then? Marshall (1995) suggests:

> Under schema theory, the student is given the dynamic role of active learner rather than the passive recipient. It is up to him or her to take incoming new information and attach it meaningfully to previously stored knowledge. The teacher in this scenario becomes the facilitator, making sure the new information is pertinent, pointing out explicitly its links to other, known information, and providing understandable examples to help the student make the appropriate connections. (p.111)

Before developing a schema based instructional strategy, it is necessary for the instructor to undertake an analysis of the knowledge domain, and recognise explicitly links between concepts within the particular knowledge domain and to other knowledge domains. As stated by Marshall (1995):

> The goal of schema-based instruction is the creation and expansion of students' schema for the domain in which instruction occurs. To implement this approach, one must first identify the major ideas of a field and the circumstances in which these ideas are manifested, and then one can construct a curriculum that concentrates in learning to recognise these ideas (situations), on developing sound mental models about how they function, on formulating ways to use the ideas creatively, and on developing the skills and procedures that are requisite for the field. (p. 120).

What sort of instructional strategies encourage the formation/modification of appropriate schema? Students need a variety of learning experiences that will encourage them to assimilate new knowledge in appropriate ways into their existing schemas, or reorganise these structures in useful ways. According to Marshall (1995), the differences between schema-based instruction and other approaches are that schema-based instruction:

- de-emphasises facts which are not central to the instruction;
- has as its goal both conceptual and procedural understanding;
- is designed so that the students are given the total picture right from the beginning, rather the a series of independent chunks which are integrated at the end, 'top-town rather than bottom-up' (1995 p.120); and
- stresses links between concepts that are important to the establishment of the schema.

Designing an instructional strategy based on the cognitive model discussed in this chapter thus requires each of these points to be addressed, and forms part of the discussion in Chapter 3.

2.6.1 The place for technology in schema based instruction

As mentioned earlier, many statistics educators have recommended the use of technology in teaching strategies designed to facilitate the development of understanding in statistics (for example Moore, 1992a; Biehler, 1993). Is this position supported by the theory of what it means to understand which has been presented in this thesis?

Pea (1987) and others have argued that intelligence does not reside totally in the mind of the individual. They recognise the impact on cognition that has historically been achieved by such advances as the development of written language:

> I take as axiomatic that intelligence is not a quality of the mind alone, but a product of the relation between mental structures and the tools of the intellect provided by the culture. (p.91)

This argument develops a Vygotskian position. Vygostsky (1978) saw cognitive development as an interplay between biological development and the social and cultural environment in which an individual was placed. As a consequence of this, educational experiences are seen to be mediated by the context of the student, taking into account the language and technology with which the student is familiar.

Anything that supports the development of learning has been termed by Pea (1987) as a cognitive technology.

> A cognitive technology is any medium that helps transcend the limitations of the mind (eg. attention to goals, short term memory span) in thinking, learning and problem solving activities. (p.91)

Examples of cognitive technologies include the pen and pencil, the slide rule, the electronic calculator, and television. The form of technology which is now popular is the computer. Like all other cognitive technologies, computer technology offers the potential to facilitate restructuring of a cognitive model. How can this be achieved? Pea (1985) recognised two roles for technology, which he refers to in a later publication (Pea, 1993, p.58) as

> …uses of the computer that focus on achieving the cognitive self-sufficiency of their users…pragmatic systems which allow for precocious intellectual performances which may be incapable without the system's support.

Salomon, Perkins and Globerson (1991) referred to these different cognitive effects on intelligence as effects *with* technology and effects *of* technology. Salomon et al. describe the effects *with* technology as the capability of the combined system of technology and user, and effects *of* technology as the effect of the technology on the cognitive ability, and hence cognitive structure, of the individual.

In what ways can computer technology support cognitive development of the individual? At its simplest level, the technology may be seen to reduce the cognitive load on the mind, allowing the working memory to pay attention to properties and relationships that may otherwise not be apparent (Sweller, 1993). However, any change in schema occurring here would be unpredictable, unless the students' attention is overtly drawn to important relationships between concepts.

Can computer technology facilitate the formation of appropriate linkages in an individual's schema relating to a specific content domain? According to Pea (1987) the answer to this is an unqualified yes. He states:

> The dynamic and interactive media provided by computer software make gaining an intuitive understanding (traditionally the province of the professional mathematician) of the interrelationships among graphic, equational and pictorial representations more accessible to the software user. Doors to mathematical thinking are opened and more people may wander in. (p. 96)

A similar view is expressed by Kaput (1992):

> Technologies based on dynamic interactive electronic media embody fundamental attributes that distinguish them from traditional static media in ways likely to have tremendous long term impact on mathematics education. (p. 525)

The theoretical arguments for the potential of the technology to develop understanding are well documented by Kaput (1992). Firstly, with computers, we are now in the position to create new notations (ways of recording and /or displaying information), that are more capable of conveying a complex idea than the traditional paper based notations. Notations that are dynamic rather than static, interactive rather than inert, offer potential representations which were once not possible. Dynamic representations are those that change as a function of time. Interactive notations are ones with which the learner may undertake a form of dialogue. According to Kaput (1992):

> .. the key difference with notations instantiated in interactive media is the addition of something new to the result of a user's actions, something that the user must then respond to. (p.526)

Due to the interactive and dynamic nature of the media, the learner is able to experience interactively links between representations that are dynamically connected by the technology. As a consequence, the possibility exists of making links overt visually which were difficult to demonstrate in inert media and, as a consequence, the learner is helped to establish desirable and appropriate links within and between schema.

The facility and availability of the technological resources now available for application in the learning environment have both expanded enormously in recent years. This advancement has caused educators to re-think some of their current beliefs about the nature of knowledge and the construction of knowledge for the individual. Papert and Turkle (1992) have argued that recent technological developments have allowed learners to successfully solve problems without developing the hierarchical cognitive structures that would be required if the technology was not available, such as solving simultaneous equations graphically by reading the co-ordinates of the point of intersection directly from the graph. They have termed this multiplicity of methodologies *epistemological pluralism*, and argue that these conceptually more simple approaches are as valid and valuable as those once required. The implication of epistemological pluralism for schema-based teaching is that it is no longer obvious which are the necessary concepts for the content domain, as the learner invents new strategies of problem solving. This is an important notion, as statistics software packages enable the student to carry out complex and sophisticated analyses without requiring knowledge of advanced statistics, and has far-reaching implications for the teaching of the subject.

2.6.2 The current role of technology in the introductory statistics course

The recent developments in technology discussed in the previous section have a huge potential to change both what is taught in statistics and the way that it is taught. As a result, it should also impact on the educator's view of what comprises an essential body of student knowledge in a discipline. Traditional courses in statistics, and in fact traditional statistical practice, have evolved using mathematical models to describe data sets. These models are then used to underpin the subsequent development of the principles of statistical inference which, in turn, rely on students having developed a reasonably good understanding of probability theory. And herein lies the stumbling block for many students.

While statistics was once a course chosen only by the more mathematically able student, today the majority of students of statistics have little mathematical aptitude or expertise (Tanis, 1992, for example). As a result, many introductory statistics courses have moved away from including much of the probability theory which underpins statistical inference. Such a course of study is then in danger of becoming basically a technique orientated "recipe book" course, relying solely on students' ability to memorise unrelated tasks.

The general feeling amongst many statistics educators (see, for example, Biehler, 1991; Brown & Schrage, 1989; Burrill, et al., 1992; Gordon & Gordon, 1992b; Hogg, 1992; Kader, 1990; Moore, 1992b) is that the probability content of the statistics course should be minimal. As stated by Chervany et al (1977), introductory statistics courses have been criticised for:

> ...placing far too much emphasis on mathematics and probability, for offering too little insight into applications of statistics...(p.18)

Many statistics educators recommend that, to overcome the need for formal probability theory, where possible computer simulation be used to illustrate important theoretical probability concepts:

> Simulation can help convey both the hard idea that random variation has a pattern in the long run and specific facts such as the central limit theorem. (Moore, 1992a, p. 23)

and

> Teachers can use simulation to illustrate ideas that are not otherwise accessible to beginning students. For example, there are many proofs of the central limit theorem, and many of them are short. But none are particularly intuitive or accessible to students who know nothing of moment generating functions. On the other hand, it is easy to demonstrate the central limit theorem, making the underlying definitions clearer and illustrating methods of simulation in the process. (Thisted & Velleman, 1992, p. 49)

There is, however, a lack of substantive research into the effectiveness of simulation exercises and opinion is divided amongst statistics educators as to the advantages of such activities. As noted by Hawkins:

> ICOTS 2 delegates were treated to "101 ways of prettying up the Central Limit Theorem on screen", but if the students are not helped to see the purpose of the CLT, and if the software does not take them beyond what is still, for them, an abstract representation, then the software fails. (1990, p. 28)

One area in which some theoretical work has been done is in determining the features of the software which may in theory prove to be more effective, such as the use of multiple representations, and dynamic interactive displays (Hawkins, 1990; Kaput, 1992; Rubin, Bruce, Roseberry, & DuMouchel, 1988) as described in the last section. This has led educators to develop specialised software, such as *ELASTIC*, the forerunner of a new generation of educational statistics package developed by Rubin et al (1988). Written for the Macintosh and using interactive graphics to teach fundamental statistical concepts, *ELASTIC* was designed according to three educational principles:

- that software should provide students with dynamic visual representations of basic concepts;
- that software should provide students with linked representation;
- that software should be interactive.

Many other software packages have since been developed which adhere to some or all of these principles, such as *JMP* (SAS, 1990) and *Models 'n' Data* (Stirling, 1991b). It would seem, then, theoretically plausible to facilitate the development of statistical understanding in non-mathematical students by including appropriate computer based experiences, based on commercially available software, in the teaching/learning strategy to be implemented.

2.7 Summary

This chapter has surveyed the literature on the nature of knowledge construction, cognitive structures and what it might mean to understand in terms of the metaphor used for cognitive structure. Important ideas such as schema and conceptual and procedural understanding were introduced and discussed. The value of concept maps as an external representation of cognitive structures was also argued.

Instructional strategies that facilitate the formation and/or modification of a theoretically desirable schema within a contructivist framework were examined, together with the potential role of computer technology to enhance these instructional strategies. Various theories concerning the use of technology, such as epistemological pluralism, and distributed cognition were discussed. The chapter concluded by reviewing the role seen by statistics educators for recently developed computer technology.

The theoretical framework for this study is based in the ideas surveyed in this chapter. In the next chapter, this framework is applied specifically to the teaching and learning of statistical inference.

Chapter 3
Teaching and Learning Statistical Inference

3.1 Introduction

My theoretical framework results from the synthesis of the work of many researchers and educators whose writings have influenced my own thinking and teaching, in particular the works of Piaget, Ausubel, Skemp, Hiebert, Carpenter and Lefevre. It leans heavily on the view of knowledge organised in semantic networks, these in turn consisting of various interrelated schemas consisting of both information and actions able to be activated by various external stimuli. The degree of activation of the semantic network depends not only on the nature of the stimulus, but also on environmental situation in which the individual is placed at the time.

A student's cognitive structure is a unique creation, dependent on that student's own mental architecture, as well as the body of social, cultural and education experiences to which he or she has been exposed throughout their lifetime. The constructivist position recognises the importance of student participation and reflection in any modification of the cognitive structure that could lead to learning. At the same time, the constructivist position acknowledges the role of formal educational experiences in facilitating the learning process, with learning itself seen to be both the acquisition of knowledge, and the process of that acquisition.

An important task for the educator is then to develop and implement a teaching strategy which recognises the importance of the underlying cognitive structure in the learning process and makes overt the key concepts and links between these concepts, which are important in the creation of the relevant schema. Underpinning such a teaching strategy is an analysis of the content of the relevant knowledge domain. This enables important concepts to be identified, and links between these concepts to be recognised.

In this chapter the key concepts involved in developing an understanding of introductory statistical inference are identified, with the aim of determining what it means to know statistics from the perspective of an expert. A result of this analysis will be the identification of the links that are fundamental to the development of both conceptual and procedural understanding. After these conceptual links have been identified, ways of facilitating the formation of these links are explored based of the theoretical framework established in Chapter 2

3.2 Concepts in statistical inference

Since it is impossible to directly observe an individual's schema, researchers must hypothesise the structure of schemas appropriate to particular mathematical tasks. To do this it is necessary to look carefully at the mathematical content of the task in order to ascertain the knowledge required in order to carry out that task successfully. By analysing the knowledge domain specific to an area of statistics, it should be possible to determine theoretically the desirable features of the relevant knowledge structure which are fundamental to an understanding of that content.

Two of the most important underlying abstractions in inferential statistics are the concepts of *population* and *sample*. The behaviour of a population is often described by a mathematical model, known as a *probability distribution*. This model is then used to make predictions concerning the nature of a sample to be selected. The properties of a sample are described by *statistics*. It is the information contained in these statistics, together with knowledge of the behavior of the probability distribution used to model the population that underlies the process of *statistical inference*. That is, that inferences are made about the properties of a population, which is usually unknown, based on the information contained in the sample, which is known.

Statistical inference requires students to recognise that the sample with which they are working is just one of a potentially infinite set of samples which may be drawn from that population. The student then needs to appreciate that, in order to make an inference, the distribution of all such samples must be known, or modelled. The distribution of a sample statistic is called a *sampling distribution*, and this is a key concept in the study of statistical inference. Many statistics educators (for example Tversky and Kahneman (1971), Rubin, Bruce, & Tenney (1990), and Shaughnessy (1992) have suggested that the sampling distribution is a core idea in the understanding of statistical inference, something that many teachers of the subject have intuitively recognised. One need only look at the proliferation of computer activities dedicated to the Central Limit Theorem to confirm this (for example Astruc et al., 1993; Betteley, 1990; Holmes, 1985; Hunt, 1986; Kader, 1990; Kreiger & Pinter-Lucke, 1992; Thomas, 1984). Yet, despite its critical role in understanding inference, experience and research have shown that the idea is generally poorly understood. For example, Cox and Mouw (1992) state:

> ...the process of sampling was poorly understood by most of the subjects...a sample was still seen by most subjects as a fixed and accurate representation of a population phenomenon. The role of sample size and the fallacy of parameters as likelihood indicators would be clearer if subjects could envisage any statistic as an estimator coming from a sampling distribution, rather than a fixed and true representation of a parameter. It may be that statistics instructors could devote a great deal of time and energy to the topic of sampling and its probabilistic nature. (p. 173)

In order to investigate the critical concepts in statistical inference, an analysis of the underlying knowledge, and the way in which each of the component ideas relates to each other, is crucial. This enables the desirable features of a schema which will support both procedural and conceptual understanding in statistical inference to be identified. It is also fundamental to the development of a teaching strategy that will facilitate the development of an organised and connected schema for statistical inference. A useful tool for doing this analysis is again the concept map, a technique developed by Novak and Gowin (1984), and used for the purpose of content analysis by some educators (see, for example, Starr & Krajcik, 1990; Jonassen, Beissner &Yacci, 1993).

3.2.1 Expert analysis of the content

The content development in introductory statistical inference can be viewed in four stages, as shown in Figure 3.1. This is also the content development that is followed by many statistics textbooks, although stages 2 and 3 are often reversed.

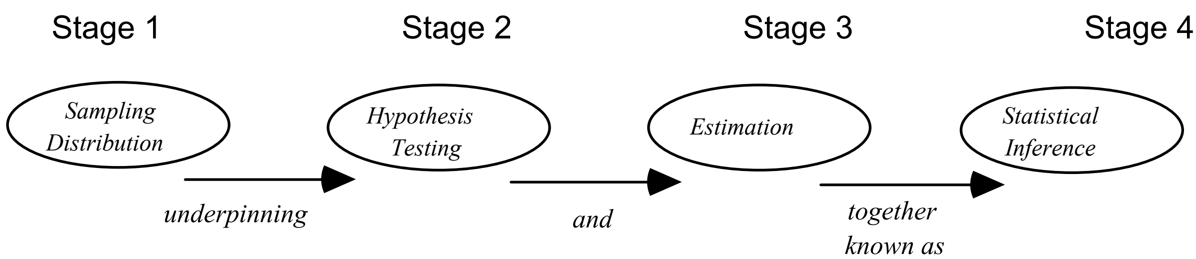


Figure 3.1 Four stages in the development of the idea of statistical inference

In order to carry out the analysis of the content domain a series of concept maps were constructed by the researcher and a colleague, both content experts in the area of introductory statistical inference. Concept mapping exercises were carried out in order to deconstruct the structural knowledge implicit in a study of statistical inference, and to identify important links between key concepts in the schema constructed by experts. Previous research has shown that as students learn the schema they create become closer in structure to those of their instructors, and thus that the students' knowledge structure can be evaluated by comparing the students' maps with the expert maps (Jonasson, Beissner & Yacci, 1993).

Several concept maps were constructed by both of the experts for each of the stages, and, by a process of negotiation, the maps arrived at met with unanimous approval. These concept maps were then termed the *expert* maps, in that they exhibited all the key features required at that particular stage in terms of concepts and connections between concepts included. From these expert maps, certain propositions could be identified, which summarised both the knowledge domain and the connections between aspects of knowledge, which identify a connected schema.

3.2.1.1 Propositions identified in the expert map for sampling distribution

Most introductory courses in statistics begin the study of statistical inference by introducing the sampling distribution, generally in a specific context, such as the sampling distribution of the mean. Once sampling distributions for the mean, the proportion, and the correlation coefficient for example have been discussed, to fully understand the concept it is necessary that students recognise the commonality of the features of sampling distributions for a variety of statistics. That is, to recognise each of these distinct measures as examples of the concept, rather than the concept itself. This mental process is termed *integrative reconciliation* (Novak & Gowin, 1984), and is indicative of the development of understanding at the reflective level (Hiebert & Wearne, 1986).

The expert concept map constructed for the generalised sampling distribution is given in Figure 3.2.

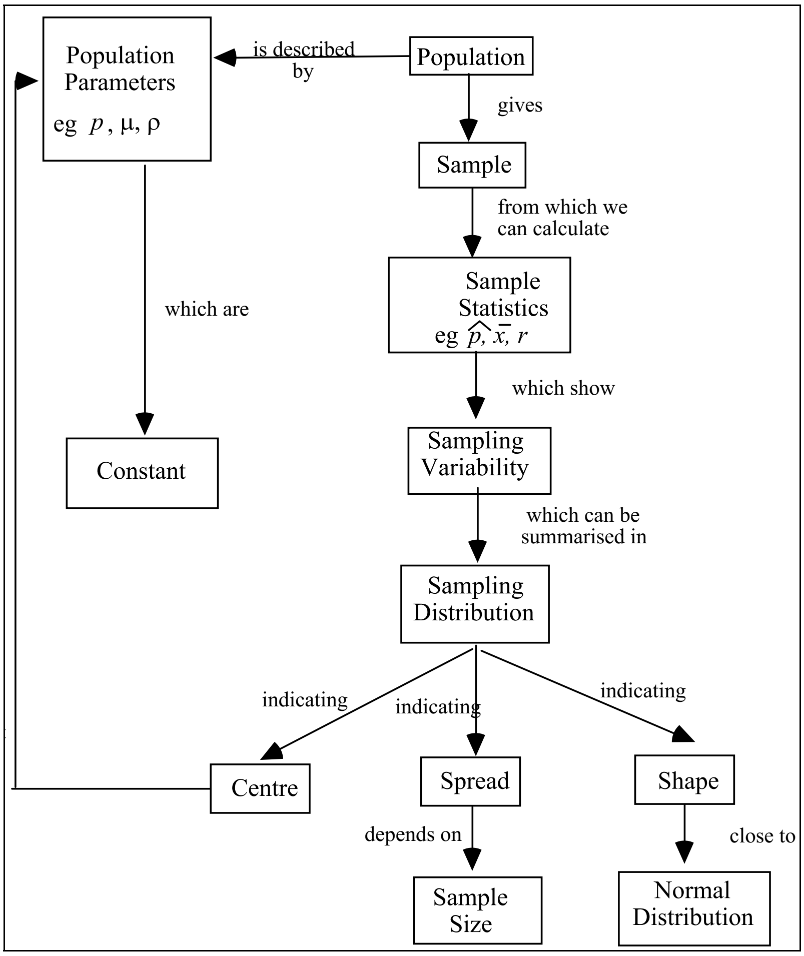


Figure 3.2 Expert concept map for the sampling distribution

The propositions given in Table 3.1 were derived from the links between concepts shown in the expert map in Figure 3.2.

Table 3.1 Propositions identified in the expert concept map for the idea of sampling distribution

Samples are selected from populations.
Populations (distributions) are described by parameters.
Parameters are constant in value.
Samples are described by statistics.
Statistics are variable quantities.
The distribution of a sample statistics is known as a sampling distribution.
The sampling distribution of the sample statistic is approximately normal*.
The sampling distribution of the sample statistic is characterised by shape, centre, spread.
The spread of the sampling distribution is related to the sample size.
The sampling distribution is centred at the population parameter.

3.2.1.2 Propositions identified in the expert map for hypothesis testing

Hypothesis testing is generally introduced in a particular context, usually the mean or the proportion for a single population. Most introductory statistics courses will quite quickly extend the students' exposure to hypothesis testing to five or six different scenarios, or tests. The ability to recognise each of these an application of the same logical process, and relate the process in each case to a sampling distribution, is fundamental to the mathematical principals that underpin the statistical argument. Hypothesis testing, in its simplest form, is concerned with the plausibility of concluding that a given sample could have been drawn from a hypothesised population. An appreciation of the variability of samples, and consequently the associated sample statistics is crucial to the logic of hypothesis testing. If the distribution of the sample statistic is known, or can be modelled, then the probability of the given sample arising from the hypothesised population can readily be calculated. This probability, or P-value, is used to aid decision-making. If the probability is small, we conclude that the link between the sample and the hypothesised population is tenuous and this is reflected in the final decision to reject the null hypothesis. The P-value is thus integral to an accept/reject procedure of decision making. As stated by Weldon (1986):

> Whenever we observe something that is unusual under ordinary circumstances, we usually suspect that the circumstances are not ordinary. (p. 325)

*Clearly this is not always true, but it is true for the sample statistics considered here

This important relationship between hypothesis testing and the sampling distribution is evident from the expert concept map, which is shown in Figure 3.3.

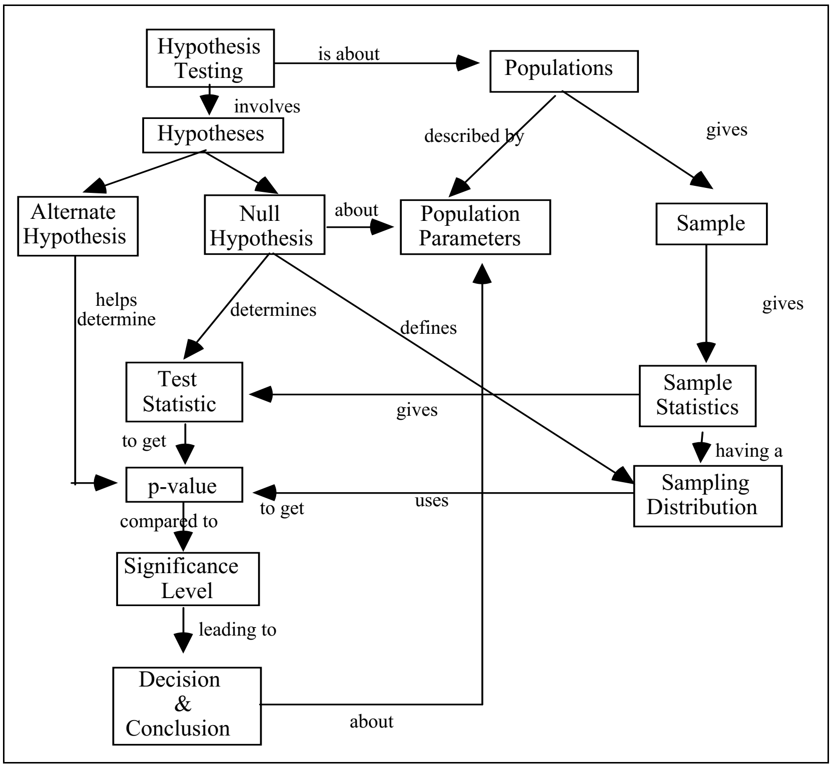


Figure 3.3 The expert concept map for hypothesis testing

From Figure 3.3 the propositions given in Table 3.2 were identified.

Table 3.2 Propositions identified in the expert map for hypothesis testing

Population (distributions) are described by parameters.
Parameters are constant values.
Hypothesis testing is about populations (parameters).
Hypotheses are about parameters.
Populations give rise to samples.
Sample distributions are described by statistics.
Statistics are variable quantities.
The test statistic is formed from the sample statistic.
Sample statistics form a distribution known as a sampling distribution.
The exact sampling distribution depends on the null hypothesis.
The value of the test statistic and its sampling distribution together determine the P-value.
The P-value also depends on the alternate hypothesis.
A decision is based on comparing the P-value to the significance level.
The decision is concerned with the null hypothesis.

3.2.1.3 Propositions identified in the expert map for estimation

The other facet of the development of the basic concepts of statistical inference involves interval estimation of the population parameter based on the value of the sample statistic. Once again the form of this interval estimate, the confidence interval, depends on the sampling distribution of the sample statistic. And once again, many formulae may be given to students to enable the calculation of confidence intervals for an unknown parameter in various situations. However, there is an underlying logic to the forming of a confidence interval not revealed by these formulae, an understanding of which is necessary if confidence intervals are to be interpreted correctly. To correctly interpret a confidence interval, students need to understand that the sample statistic is a variable, and thus that the confidence interval to which it leads is a variable interval. Furthermore, the value of this confidence interval is dependent both on the value of the sample statistic obtained from the particular sample selected, and on the properties of its sampling distribution.

The link between the process of estimation and the sampling distribution is clearly shown in the expert concept map, which is shown in Figure 3.4.

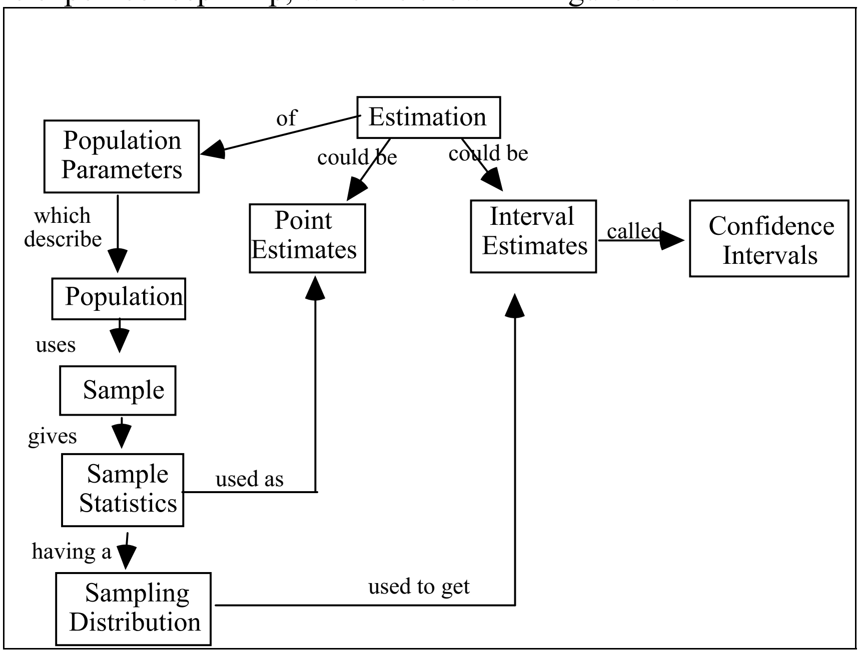


Figure 3.4 The expert concept map for estimation

The propositions identified from the expert concept map for estimation are given in Table 3.3.

Table 3.3 Propositions identified in the expert concept map for estimation

Populations are described by parameters.
Estimation is concerned with population parameters.
Samples can be drawn from a population.
Samples are described by sample statistics.
A point estimate for a population parameter is given by the sample statistic.
The values of sample statistics are variable, and their behaviour can be summarised by a sampling distribution.
Knowledge of the sampling distribution and the value of the sample statistics enables us to calculate an interval estimate.
This interval estimate is called a confidence interval.

3.2.1.4 Propositions identified in the expert map for statistical inference

To demonstrate a full understanding of the ideas of statistical inference, all of the relationships recognised in the propositions previously tabled should be accessible in the student schema, together with the added recognition of the equivalence of hypothesis testing and estimation in certain situations. The sampling distribution is present as the central statistical principle, underpinning both the hypothesis testing process and interval estimation.

The pivotal role of the sampling distribution in statistical inference is clear from the expert concept map, which is shown in Figure 3.5.

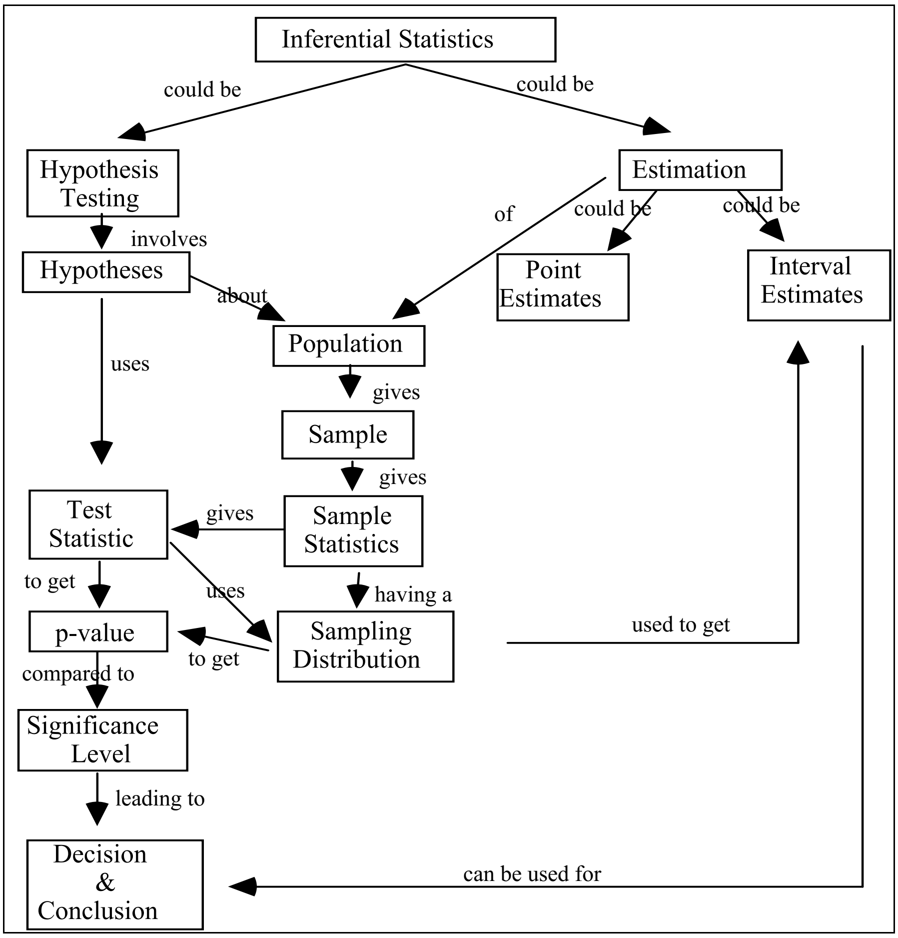


Figure 3.5 The expert concept map for statistical inference

A summary of the propositions distilled from the expert concept map for statistical inference is given in Table 3.4.

Table 3.4 Propositions identified in the expert concept map for statistical inference

Populations give rise to samples.
Population (distributions) are described by parameters.
Parameters are constant values.
Samples are described by statistics.
Statistics are variable quantities.
Sample statistics form a distribution known as a sampling distribution.
Hypothesis testing is about populations
Hypotheses are about parameters
The value of the test statistic and its sampling distribution together determine the P-value.
A decision is based on comparing the P-value to the significance level
Knowledge of the sampling distribution enables us to calculate an interval estimate.
Statistical inference is concerned with both hypothesis testing and estimation
Both aspect of inference are concerned with knowing more about a population parameters
Consideration of the confidence interval is an equivalent act, and should lead to the same conclusion.

3.2.1.5 Summary

The analysis undertaken in this section has enabled identification of the links between concepts in the expert schema for statistical inference, which can be assumed to be required for an understanding of statistical inference. Based on this analysis, a teaching strategy could be developed which acknowledged not only the relevant concepts, but also the important links that need to be established between concepts.

In particular, the analysis has confirmed the central role in statistical inference of the sampling distribution, as conjected by others (for example Cox & Mouw, 1992; Moore, 1992b; Rubin, Bruce, & Tenney, 1990; Shaughnessy, 1992). In summary, understanding of the sampling distribution, and maintenance of links between the sampling distribution and of both aspects of statistical inference, hypothesis testing and estimation, is necessary in building understanding of statistical inference.

Before discussing the nature of instructional experiences which could support the development of the appropriate schema for statistical inference in students, further analysis of the sampling distribution concept in terms of the mathematical and symbolic representations used or able to be used in classrooms must be undertaken.

3.3 Representations of the sampling distribution

The sampling distribution arises as a means of describing the results of the sampling process. Intrinsic to an understanding of the idea of a sampling distribution is the students' recognition of the important features of this sampling process, such as the notions of invariance of the elements in a population and the variability of the elements in a sample. These two key ideas concerning sampling are succinctly summarised by Rubin (1990):

> There are two almost opposing ideas: sampling representativeness and sample variability. *Sampling representativeness* is the idea that a sample taken from a population will often have characteristics similar to those of its parent population (eg the proportion of girls in a class is likely to be close to the proportion of girls in the whole school). *Sample variability* is the contrasting idea that samples from a single population are not all the same and thus do not match the population (eg some classes in the school are likely to have many more girls than boys, even if the school population is evenly divided). (p.314)

Because the concept of the sampling distribution is a multi-faceted and complex, being associated with both the selected sample and the dynamic process of sampling, there are many images that can be, and arguably should be, associated with a schema for sampling distribution. As discussed in Section 2.5, there are a range of mathematical, symbolic and computer generated representations that can be associated with a concept. Those of relevant to the concept of sampling distribution, used in both the teaching and application of this statistical construct, are discussed in this section.

3.3.1 Mathematical representations: Probability density versus frequency distributions

As identified earlier, a particularly important idea in the development of an understanding of statistical inference is the recognition that when samples are drawn from a population they will vary, and that this variation will conform to a predictable pattern. This idea has been traditionally introduced in statistics courses using a deductive approach based on probability theory (for example Johnson & Bhattacharyya, 1992; Mendenhall, Wackerly, & Scheaffer, 1990). Such explanations are usually expressed in a highly symbolic form which tends to make the argument inaccessible to all but the mathematically able, now a very small minority of the students taking introductory courses in inferential statistics. But perhaps more importantly, the sampling distribution described by a probability density curve is a theoretical development that is difficult to relate to the actual physical process of drawing a sample from a population.

As a result writers of statistics textbooks have come to recognise that there are deficiencies with a random variable based explanation and now often accompany or replace this with an empirical argument (for example Devore & Peck, 1986; Ott & Mendenhall, 1990). This approach uses the long run relative frequency argument, where the sampling distribution is viewed as the result of taking repeated samples of a fixed size from a population and calculating the value of the sample statistic for each and then forming a frequency histogram to display the distribution of the sample statistic values. In this approach, the probability density function forms the theoretical sampling distribution, and is seen as the limit of the empirical sampling distribution.

Thus, sampling distributions may be defined through either of two alternative statistical arguments, one from a theoretical, random variable perspective and the other from an empirical perspective. Whichever approach is used, and despite the closer relationship of the relative frequency approach to the sampling process, the end product of the discussion on the sampling distribution is generally seen as a probability density curve. This theoretical distribution can usually be described mathematically in terms of a well-known distribution, such as the normal distribution. Once the idea of a theoretical sampling distribution has been established, students are generally not again reminded of the link between the sample statistic and the empirical form of the sampling distribution which arose from the sampling process. Determination of the P-value becomes an abstract exercise in the calculation of the probability in the tail of a probability density curve. And any cognitive link which was established between the process of hypothesis testing and the sampling process is possible unlikely to remain over time, as it is no longer made explicit.

The empirical approach to the development of the sampling distribution sees the sampling distribution as arising from repetitive sampling, has the advantage of being more readily related to the actual physical process of sampling than the theoretical approach, which derives the sampling distribution as a probability density curve. Analysis of the expert schema (concept and links) associated with statistical inference show that it is this view of a sampling distribution, as a way of describing the sampling variability of a sample statistic, that is needed to correctly interpreting the results of the inference. From this perspective the observed value of a sample statistic is readily viewed as one of many possible values, some of which were more likely to be observed from in a sample drawn from a particular population than others. Similarly, a confidence interval is seen as a variable interval estimate of the population parameter, which is quite likely to contain that parameter.

Thus, the empirical view of sampling distribution is an essential component of a schema for sampling distribution which facilitates understanding of statistical inference. It is important then for both the empirical and theoretical mathematical representations of sampling distribution to be part of a student's schema. And, more than this, it is desirable that the schema associated with sampling distribution contain links between these two mathematical representations, so that whichever part of the schema is activated, the alternative mathematical representation is available.

From the point of view of understanding the process of statistical inference, it is useful for the sampling distribution to be seen by students as the distribution of a sample statistic, based on the observation of many, many samples, which can be modelled by a particular theoretical probability distribution under certain assumptions (rather than just as a probability density curve).

3.3.2 Symbolic representations

Symbolic representations are the external representations used to exemplify a concept (see Section 2.5). Many mathematical concepts are complex and multi-faceted, and several symbolic representations are often used to illustrate such concepts. Consider, for example a linear relationship such as

$$\boldsymbol{y} = 2\boldsymbol{x} + 3$$

This relationship can be expressed in symbolically many ways, such as an algebraic equation, as above, or graphically, or as a table of values and so on. Each representation illustrates a different aspect of the concept. The co-existence of many representations in a student's schema does not necessarily indicate a depth of understanding. Lesh, Post and Behr (1987) have distinguished between movement within a representation *(transformation)*, and movement between representations *(translation)*. The ability to recognise a concept within a variety of representation systems, to successfully manipulate the concept within a representational system and to translate the concept between symbolic representations are recognised an indicators of understanding of the concept (Lesh, Post & Behr, 1987). These abilities are synonymous with the metaphor of each schema as a complex but connected network, available from any point of entry into the structure through a complex system of links and nodes, and each concept within the schema similarly available.

Based on this model of understanding, to develop the student's understanding it is desirable to include in the student's experience a variety of symbolic representation for each of the mathematical representations of a concept, and to emphasise the similarities and differences between them. The constructivist position on learning would also support this contention (Cobb, Yackel & Wood, 1992) as different students will obviously derive different information and varying levels of benefit from the various representations offered.

As discussed in Section 2.5, possible types of symbolic representations suggested by Lesh, Post and Behr (1987) include experience based scripts, manipulable models, pictures or diagrams, spoken languages and written symbols. Can this list of possible representations be applied to sampling distribution in order to clarify further the knowledge domain and its representation in regard to sampling distribution? Are there other representations not listed which are applicable here? In order to answer these questions a detailed analysis of the symbolic representations possible must be undertaken.

3.3.2.1 Analysis of representations of the sampling distribution

In section 3.3.1 it was proposed that there are two distinct mathematical representations available for the sampling distribution, both of which seem to be important for the student learning introductory statistical inference. To make the analysis of symbolic representations available for sampling distribution easier to follow, it was carried out in a simple binomial context. Suppose a sample of size *n* is drawn from a population, with the population proportion denoted by *p*. The number of individuals in the sample with the desired attribute is *X*. This means that the sample statistic that would be used for further inference would be $\frac{X}{n}$, usually denoted as $\hat{p}$, and the sampling distribution of $\hat{p}$ would be the sampling distribution of interest.

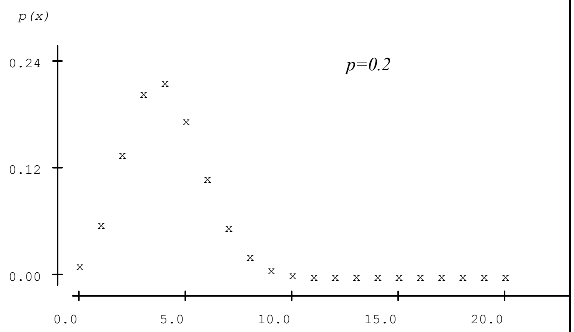
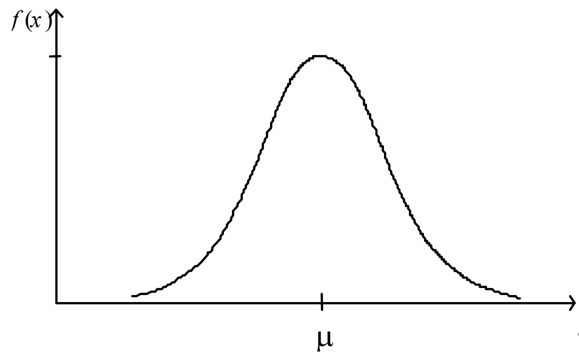
Consider firstly the symbolic representations appropriate for the *theoretical* representation of the sampling distribution of the sample proportion. The distribution of the random variable X, the distribution of the random variable, and the relationship between them together comprise the theoretical concept of sampling distribution. An analysis of the representations available for this mathematical representation showed that there are four which are commonly used:

- Experience based script
- Pictures or diagrams
- Spoken language

- Written symbols

An example of each of these, using particular values for *n* and *p*, is given in Table 3.5.

Table 3.5 Symbolic representations for the theoretical sampling distribution

Distribution of X	*Symbolic representation of the sampling distribution*
Experience based scripts	
The distribution of X is: $X \overset{d}{=} \text{Bin}(10, 0.6)$	The distribution of $\hat{p}$ is: $\hat{p} \overset{d}{=} \text{N}(6, 2.4)$
Pictures or diagrams	
The graph of the probability function can be represented as shown below, when $n = 20$ and $p = 0.2$. 	The sampling distribution for is modelled by a normal distribution, with the graph of the density function similar to the one shown below. 
Spoken language	
X is a binomial random variable, describing the distribution of the number of successes observed in a sample of size *n*, when the probability of observing a success on any trial is *p*.	$\hat{p}$ is the mean of *n* independent identically distributed random variables. Thus, by the central limit theorem, $\hat{p}$ is (asymptotically) normal.
Written symbols	
$\Pr(\boldsymbol{X} = \boldsymbol{x}) = \binom{\boldsymbol{n}}{\boldsymbol{x}} \boldsymbol{p}^{x}(1-\boldsymbol{p})^{n-x} \quad x = 0,1,\ldots,n$	As $n \to \infty$, $\hat{p} \to \text{N}(p, \frac{p(1\text{-}p)}{n})$

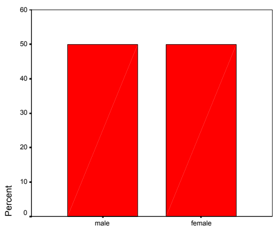
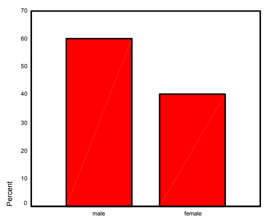
Using the proposed model of understanding, a full understanding of the theoretical mathematical representation of sampling distribution would imply that the student was able to move within and between these symbolic representations suggested in Table 3.5. However, it is the opinion of some researchers that many students are only able to work within the experience based script, able to apply routine mathematical algorithms to statistical inference problems (for example, Williams, 1993).

Consider now the symbolic representations available for the empirical mathematical representation of sampling distribution, which is formed by summarising the data resulting from repetitive sampling from a population. Here, the focus is on representation of the population from which the sample is drawn, and on the sample itself. The sample statistic is determined from the sample, and the sampling distribution formed by repetition of the process. Once again the binomial context is used to illustrate the symbolic representations suggested.

Table 3.6 Representations for the empirical sampling distribution

Population	*Sample*	*Value of the sample statistic*	*Symbolic representation of the sampling distribution*
Manipulable models			
Physical model, such as a box of beads, with difference colours representing success and failure.	A random selection procedure, such as a sampling shovel is used to select the sample.	Determination of the sample proportion by counting the number of coloured beads in the sample.	Sampling variability leading to the sampling distribution is observable from the variation in the sample statistic, summarised in a statistical graphical display such as a histogram.
Spoken language			

A population consists of 60% females and 40% males, for example.	A random sample is chosen from the population.	The percentage of females in the sample is calculated. This will not necessarily be 60%.	If another sample was chosen then the contents of that sample would most likely be different from the first sample, and hence the proportion of females in the sample different again.

Population	*Sample*	*Value of the sample statistic*	*Symbolic representation of the sampling distribution*
Pictures or diagrams			
A schematic diagram is used to represent the individuals in a population, such as the following population of 13 males and 11 females. 	A sample of size 5 is chosen at random and also represented schematically. 	The proportion of females in the sample is determined by counting. (The number of females in this sample is 2, so $\hat{p}$ = 0.4)	Repeated sampling illustrates sampling variability, leading to the sampling distribution which can be summarised in a statistical graphical display such as a histogram.
The proportion of males in the population is shown in a bar chart (p =0.5) 	The proportion of males in the sample is shown in a bar chart ($\hat{p}$ = 0.6). 	The value of $\hat{p}$ can be read from the bar chart.	Repeated sampling illustrates sampling variability, leading to the sampling distribution which can be summarised in a statistical graphical display such as a histogram

The symbolic representations for the empirical sampling distribution discussed in Table 3.6 tend to more directly reflect the physical process of sampling and, as such, more readily relate to the student experience. The variability of the sample and in turn the value of the sample statistic, are able to be demonstrated through physical examples. In this way the sampling distribution can be built up from the students' direct experiences. This is in contrast to the theoretical sampling distribution, which is based on relatively abstract mathematical constructs.

The list of symbolic representations presented here (for each mathematical representation) for the sampling distribution is not claimed to be exhaustive, but indicative of the range of representations which can be used at each stage of the derivation of sampling distribution. In the next section, the potential for creating external linkages between representations offered by the emerging technologies will be discussed.

3.3.3 Computer generated representations

As well as the symbolic representations considered in the previous tables, there are many computer-generated representations available for the sampling distribution, some of which will be discussed in this section. Their contribution to the images otherwise available to students may be considered through the following analysis.

3.3.3.1 Paper based computer generated images

A computer may be used to generate images (usually histograms) of empirical sampling distributions. These are generally prepared by an instructor and communicated to the student in the form of printed output, such as in printed notes or overhead images. Because the instructor has the flexibility to generate many such images without much effort, several can be used in order to illustrate the shape, centre and spread of the resultant distribution, and how each of these are affected by sample size. An example of a computer generated sampling distribution for the sample mean is shown in Figure 3.6.

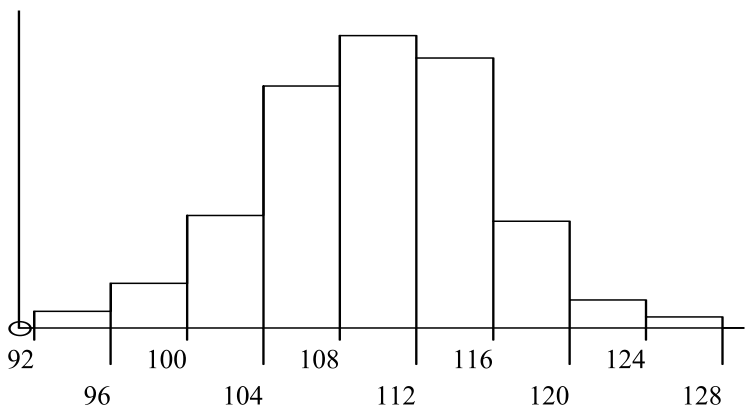


Figure 3.6 Histogram showing the distribution of the means of 200 samples of size 6 drawn from a population with a mean μ = 110 and standard deviation σ = 15.

3.3.3.2 Computer based images

The continued development of statistical software has resulted in a proliferation of computer based images, with many texts now including student activities concerned with the sampling distribution (for example Martin, Roberts, & Pierce, 1994; Astruc et al, 1993). For the purpose of this analysis the symbolic representations available while working with a computer will be considered within two general categories, static and dynamic.

Static computer-based images

The usual sort of image produced by computer software is a static display, generally a single representation, although more may be present. Students are given access to the computer and software that enable them to produce these for themselves. An example of the type of software commonly used for such activities is *Minitab*. The resulting images produced on the computer screen are essentially the same as those that could be produced on paper, such as that shown in Figure 3.6.

A clear advantage of these computer-based images over the paper-based images is that the students use the computer to construct these images of the sampling distribution for themselves. Thus, the student is able to manipulate the distribution of the population, as well as the sample size, in order to investigate the effects of these on the sampling distribution. This potential for investigation seems to offer the possibility for students to recognise the key feature of the sampling distribution, and construct links which reflect this growth in understanding in their schema for sampling distribution. However, the links between the changes that the student can make and the resulting sampling distribution are not overt in that they appear on different screens at different times. Considerable cognitive effort is required by the student on order for this learning experience to realise its potential.

Dynamic computer images

As well as the general categories of symbolic representation listed earlier (experience-based scripts, manipulable models, pictures or diagrams, spoken languages and written symbols), Kaput (1992) has suggested that the advent of modern computer technology has meant that we are now in the position to create new notation systems (ways of recording and /or displaying information). These new notation systems have a greater capacity for conveying a complex idea than the traditional paper based notations, by providing multiple, dynamic, linked representations of the concept (see for example *Sampling Laboratory* (Rubin, 1990) which is described in detail later in this section).

From consideration of multiple images, different insights into a concept are available. Dynamic linking of the images allows the student to observe first hand the effect on one image of representation in another representation. Kaput suggest that computer based linkages between representations (a linking of symbolic representations) should facilitate integration of cognitions (a linking within mental representations). He suggest that this has the potential to increases the students' level of understanding, as the mental links possibly remain when the physical (computer-based) link is removed. Others support this notion of a cognitive residue which remains after the experience, in particular Salomon, Perkins and Globerson (1991).

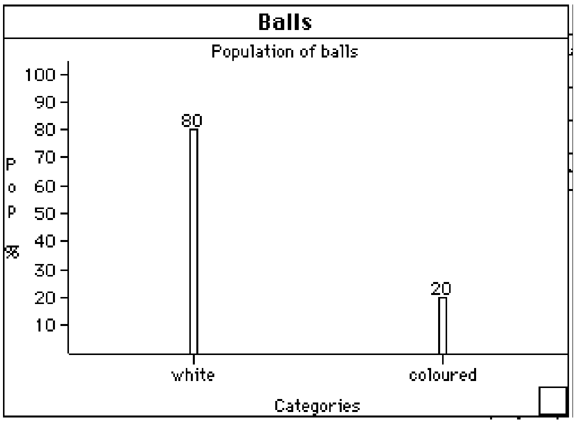
It is for this reason the dynamic computer-based images seem to offer the greater potential to contribute to the development of the concept of a sampling distribution in students. The sampling distribution is the product of a sampling process. Hence, understanding the sampling distribution cannot be dissociated from an understanding of the process by which it is formed (Jones & Lipson, 1993). The sampling process is dynamic and involves the linking of several elements: a parent population, the samples drawn from the population, the values of the test statistic extracted from each of the samples, and the sampling distribution which is generated. In a purely paper-based explanation, the written word is used to describe the process outlined above and graphics are used to illustrate the end product of the process, the empirical distribution of the sample statistic. The bringing together of the sampling process and the resulting sampling distribution requires a high degree of cognitive reconciliation, which seems to be difficult for many students.

An example of computer software which supports the teaching of sampling distribution through dynamic, linked images is *Sampling Laboratory* (Rubin, 1990). The strategy used in this computer package is to create a working model of the sampling process which has as its product a histogram displaying the resulting sampling distribution (similar to that produced by *Minitab*), but also displays the process by which the sampling distribution is obtained.

When *Sampling Laboratory* is used investigate the sampling process, the required number of samples is drawn sequentially, and the screen shows simultaneously the three windows illustrated and explained in Table 3.7.

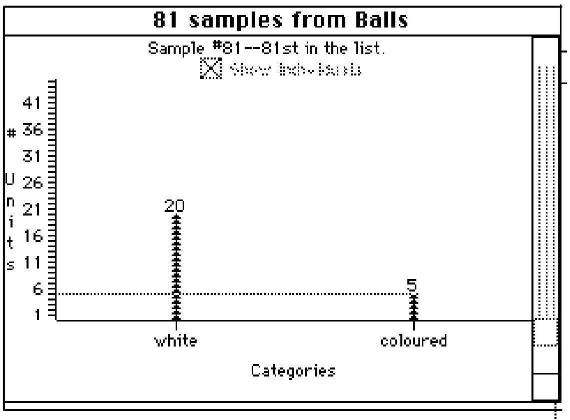
Table 3.7 Dynamically linked windows shown simultaneously by *Sampling Laboratory*

Window 1



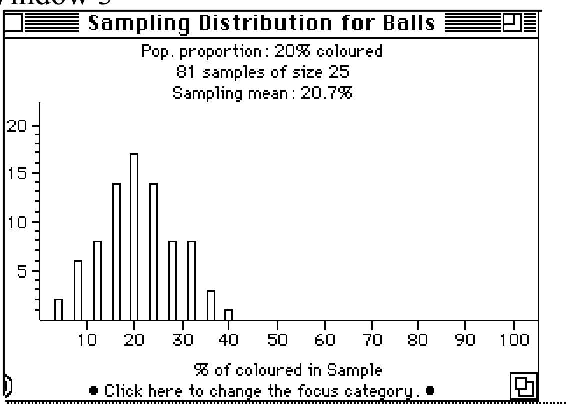
Window 1 shows as a bar chart the probability distribution for population. While the sampling takes place dynamically, this window remains static, reinforcing the notion of the population parameter remaining constant throughout the sampling process.

Window 2



Window 2 shows that, as each sample is selected the actual contents of the sample are displayed. The graphical representation is a stack of triangles, chosen so that the individual outcomes can be observed, and also to differ from the representation for the population. This window changes with each sample chosen.

Window 3



Window 3 shows the empirical sampling distribution of the sampling proportion ($\hat{p}$) building dynamically as the sampling takes place. The value of the sample proportion from the current sample is shown in black on the screen (not shown here), and the current mean value of the sample proportions is also shown.

The object of *Sampling Laboratory* is to illustrate both the sample to sample variability as well as the formation of a predictable pattern. Static representations illustrate the product (pattern) only, while dynamic linked representations illustrate the product and process and how they relate. Thus, the use of dynamic linked displays has the potential to considerably reduce the cognitive load involved in understanding text based explanations of the idea of a sampling distribution by illustrating both product and process, and should lead to greater understanding.

It is also possible in many computer analysis and simulation packages to superimpose a theoretical probability distribution (such as a normal distribution) on an empirical distribution, such as a histogram (*SAS*, 1990; Stirling, 1991). Whilst not dynamic, such a representation explicitly connects the theoretical and empirical representation of the sampling distribution, again possibly facilitating the creation of a link between these mathematical presentations in the cognitive structure of the student.

3.3.4 Maintenance of representations

In his work on concept images, Vinner (1983) pointed out that the concept image can be complex and multifaceted and that a particular task may evoke a particular aspect of the concept image. As a consequence, if students are engaged only in tasks that draw on a particular aspect of the concept image, one cannot assume that all representations, which were once available to the student, will continue to be available. As stated by Vinner (1983):

> Elements (of the concept image) which are not constantly reinforced have a good chance of being forgotten and thus the concept image is distorted. (p. 305)

Vinner's argument can be considered in terms of links which exist between concepts in the student's schema. Whilst considerable effort can be made on behalf of and by the student to establish links between representations, if all subsequent work reinforces only a subset of these linkages, then some of the links will be lost over time.

Thus, in general, teaching strategies must consider not only the necessity of multiple representations to the establishment well connected schema, but also the need to maintain those representations within the students' schemas which are going to assist the student in their understanding and application of the concept. This is important when trying to build an understanding of such complex and multi-faceted concepts such as the sampling distribution.

3.4 Teaching the sampling distribution

The establishment and maintenance of all the necessary links in the students' schemas for sampling distribution is undoubtedly a challenge in teaching introductory inference. The logic and procedures of hypothesis testing, for example, are clearly difficult for most students (see for example, Williams, 1995). Consequently, efficient experience based scripts, or superprocedures, have been developed over time for the solution of such problems, which students are encouraged to follow. For example, most introductory statistics texts will include a step-by-step procedure for the solution of hypothesis testing problems – almost a recipe (see, for example, Bluman, 1997; Devore & Peck, 1986; Goldman & Weinberg, 1985; Weiss & Hassett, 1987). As stated by Shaughnessy (1992) the current courses in statistics offered at most universities

> …continue to be either rule-bound, recipe type courses for calculating statistics, or overly mathematical introductions to statistical probability that were the norm a decade ago.
> (p. 466)

Thus, even whilst most courses and texts spend a great deal of time developing the idea of the sampling distribution, in one or both of its mathematical representations, once the bridge to inference is formed there is generally no further reference to the empirical representation of sampling distribution, although the theoretical representation (the probability density curve) continues to be used. Thus, according to Vinner (1983), it is likely that many students lose that aspect of their concept image of sampling distribution all together, or at least find that it is not evoked by the script they are using for the problem at hand.

Thus, an effective teaching strategy for statistical inference, which has understanding as its aim, must not only facilitate the development of a schema with key concepts appropriately linked. It must also continue to relate to the empirical representations of sampling distribution throughout the development of subsequent topics, thus maintaining the link between the algorithms of statistical inference and the sampling process.

The theory presented in this thesis suggests that students would benefit from being exposed to a range of representations of a complex concept, in order to facilitate the links between representations which support understanding of that concept. As well as exposing students to a variety of representations, students also require educational experiences which encourage the appropriate reorganisation of the relevant cognitive structures, and which support the establishment of links between concepts both within and, where necessary, between these structures.

To this end, a multitude of computer activities have been developed to help promote an understanding of sampling distribution (for example Glencross, 1986; Pedlar, 1991; Rennolls, 1991; Soon, 1990). Many of these activities stand alone, but some educators have recognised the need to have students focus on particular elements of the computer activity and have supplemented these activities with appropriately structured worksheets (for example Martin, Roberts, & Pierce, 1994). Supplementation of computer-based activities with worksheets is consistent with the view of Salomon and Globerson (1987), who noted that mindfulness (that is, mindful engagement with the task) as a necessary pre-requisite for a fruitful learning experience. However, despite their popularity, overall the value of these computer-based activities is as yet unclear, with little research based evidence as too their effectiveness. According to Shaughnessy (1992):

> To date, there have only been a few studies on the effects that computers or computer simulations have on students' learning and understanding of probability and statistics. (p. 484)

As a part of the current study it is planned to develop a teaching strategy which is designed to support the development of understanding of the concept of sampling distribution and its pivotal role in understanding statistical inference. This teaching strategy will recognise the multiplicity of symbolic representations possible for both the empirical and theoretical representations of the sampling distribution, and endeavor to make explicit the links between such representations.

3.4.1 Developing the concept of sampling distribution

A teaching strategy for the development of the concept sampling distribution based on recognition of the various representations of sampling distribution, and the need to make explicit links between representations is outlined in this section. In addition, further justification for the approach taken comes from Vygotsky's concept of the '*zone of proximal development'* (Vygotsky, 1978). Brown et al (1993) interpreted the zone of proximal development as:

> …the distance between current levels of comprehension and levels that can be accomplished in collaboration with people or powerful artefacts. The zone of proximal development embodies a concept of readiness to learn that emphasises upper levels of competence. (p. 191)

According to Vygotsky, to move students to their potential zone of development a teacher should firstly establish a level of difficulty for the student which is challenging but not too difficult and then provide scaffolded instruction, where scaffolding is gradually removed (McInerney & McInerney, 1994). Thus, a theoretical development is chosen in which students are exposed firstly to representations of sampling distribution which are closest to their own experiences, and then moves gradually to the more abstract representations.

The development of the concept of sampling distribution which follows moves through several levels of abstraction, in order of complexity, from the physical to the empirical and finally the theoretical.

Level 1: Physical Sampling

The first step in this teaching strategy designed to build a well-connected schema for sampling distribution starts with an activity based on physical sampling.

The need to begin discussions concerning the sampling process with physical sampling has been recognised by many statistics educators, amongst them Shaughnessy (1992) who notes:

> Although I have no formal research evidence, it seems to be very important for many students to have experience of actually generating and gathering their own data *physically*...before they can understand or accept computer simulations. (p.485)

Since a sampling distribution is the end product of a sampling process, students could benefit from the physical sampling experience; that is, drawing samples from an actual population and investigating the results. This intuitive recognition of the importance of the role of physical sampling is supported by a consideration of Piaget's view of an experience as either assimilated into an existing schema, or as a catalyst for the modification of a schema (accommodation). The act of selecting (sampling) and observing with interest the nature of the result is well within the experience of most children and adults and is part of the informal knowledge accumulated by many. Most children have selected a handful of sweets and noted their colour, or been chosen in a sports team and observed its constitution. There are reasons to believe then that a physical sampling activity may convey notions sampling which are assimilated into the student's existing schema, or be accommodated with minimal modification. Indeed, physical sampling experiences have been used successfully to investigate the sampling process with quite young children (for example, Rubin, Bruce & Tenny, 1990).

Level 2 Computer generated sampling (Empirical)

On the basis of physical sampling a sampling distribution can be generated for a sample statistic using a display such as a histogram or stemplot. However, due to the time and effort involved in undertaking the sampling and constructing the display, these sampling distributions are usually limited to the display of a relatively few values of the sample statistic. Furthermore, since the amount of data displayed is small, the resulting display may not adequately represent the form of the sampling distribution.

In order to establish patterns of variability in the sampling statistic a computer can be used to generate large numbers of samples. However, computer simulation of the sampling process is a higher level of abstraction for the student in that it involves representations no longer clearly connected to the process of sampling. In line with the theory of Vygotsky, the idea of the zone of proximal development, one could conject that by experiencing the physical sampling situation first, students who subsequently carried out computer simulation exercises with appropriate reference back to the physical sampling process are able to make these connections for themselves. That is, the physical sampling experience has the potential to help scaffold the development of the students' understanding, and provide an opportunity for the student to connect the computer sampling activities to the actual physical process of sampling in their schema.

Appropriate computer experiences include the use of the specialised computer packages mentioned previously, such as *Sampling Laboratory*, which use dynamic linked computer images to encourage and support the development of such linkages in the cognitive structures of the student. Through computer based sampling exercises, and appropriate summaries of their results using statistical displays such as histograms and stemplots, there is the potential for key ideas to be explored by the students. These include the form of the sampling distribution, the relationship between the sampling distribution and the distribution of the parent population, and the effect of sample size on the sampling distribution.

Level 3 Traditional Exercises (Empirical to Theoretical)

In most introductory statistical inference courses, students are expected to formalise their knowledge of the sampling distribution by using an appropriately chosen theoretical probability distribution as a model for it. It is this theoretical model which forms the basis of further development of statistical capability. A probability distribution, as mentioned earlier, is a mathematical representation of sampling distribution. It is not readily connected to the process of sampling or even to the empirical sampling distribution.

If, for example, the appropriate probability model is the normal distribution (which it often is), then the student will have earlier in their experience been introduced to the probability density curve which describes this distribution, and have used the properties of it to solve probability problems. As a result, the students will have already constructed in their cognitive structure a schema for the normal distribution. At this level, then, for enhanced understanding of one form or another, links between the student's schema for the normal distribution and the student's schema for empirical sampling distribution are important. If these links are not formed, then the notion of theoretical sampling distribution, as arising by observation of the behavior of the empirical sampling distribution is lost. The procedures of hypothesis testing and estimation become just applications of the normal distribution, unlikely to be related to the process of sampling for many students.

Designing and implementing instructional programs which encourage the formation of links between the empirical and theoretical sampling distributions, and the maintenance of the links during the development of the logic of hypothesis testing and estimation is clearly a challenge for the statistical educator. In the next section a strategy for teaching statistical inference without any reference to the theoretical sampling distribution, and the possible educational advantages of this strategy, are discussed.

3.5 Computer Intensive Methods

The developing area of Computer Intensive Methods provides an alternative path to inference to the one outlined in the last section. This approach theoretically reduces some of the complexity of teaching statistical inference, in that there is no need to introduce the theoretical sampling distribution in any of its representations. All of the analysis takes place using only the empirical sampling distribution.

These methods are called computer intensive because they involve the computation of the statistic of interest for many data sets obtained by repeated sampling, and consideration of the observed value of the statistic by comparison with this computer generated empirical distribution. According to Diaconis and Efron (1983):

> The payoff for such intensive computation is freedom from two limiting statistical factors that have dominated statistical theory since its beginnings: The assumption that the data conform to a bell-shaped curve and the need to focus on statistical measures whose theoretical properties can be analysed mathematically. (p. 96)

For the practising statistician, there are often definite advantages of using a computer intensive method rather than a classical method when restrictive assumptions about the nature of the data do not hold or traditionally used test statistics are not appropriate. From the pedagogic point of view there may also be advantages to the student of statistics.

Essentially, applying a computer intensive method means that the entire inference problem is dealt with from an empirical perspective using a computer generated distribution. Students are not required to make the conceptual connection between the empirical and theoretical sampling distributions, but actually use an empirical sampling distribution as the basis for the inference. For example, consider the following example, taken from a typical introductory statistics text by Weiss and Hassett (1987 p.427):

> A highway official wants to compare two brands of paint used for striping roads. Ten stripes of each paint are run across the highway. The number of months that each stripe lasts is given below.

Brand A		**Brand B**	
35.6	36.1	37.2	36.4
37.0	35.8	39.2	37.5
34.9	34.9	37.2	40.5
36.0	38.8	38.8	38.2
36.6	36.5	37.7	36.6

> Based on the sample data, does there appear to be a *difference* in mean lasting time between the two paints? Use $\alpha = 0.05$.

3.5.1 Traditional approach

Let us consider how this problem would be tackled by students using a traditional method of solution. From the question the students must recognise that they are required to perform a hypothesis test for the equality of means with a two-sided alternative thus:

$$H_0: \quad \mu_1 = \mu_2$$
$$H_1: \quad \mu_1 \neq \mu_2$$

If we assume that the two samples are taken independently from two normally distributed populations with means and respectively, and the respective standard deviations are unknown, then a hypothesis test for comparing the means can be carried out by consideration of the test statistic

$$\boldsymbol{t} = \frac{\overline{\boldsymbol{x}_1} - \overline{\boldsymbol{x}_2}}{\sqrt{\dfrac{\boldsymbol{s}_1^2}{\boldsymbol{n}_1} + \dfrac{\boldsymbol{s}_2^2}{\boldsymbol{n}_2}}}$$

which has a distribution which can be approximated by a *t*–distribution with degrees of freedom given by

$$\boldsymbol{df} = \frac{\left(\boldsymbol{s}_1^2 / \boldsymbol{n}_1 + \boldsymbol{s}_2^2 / \boldsymbol{n}_2\right)^2}{\dfrac{\left(\boldsymbol{s}_1^2 / \boldsymbol{n}_1\right)^2}{\boldsymbol{n}_1 - 1} + \dfrac{\left(\boldsymbol{s}_2^2 / \boldsymbol{n}_2\right)^2}{\boldsymbol{n}_2 - 1}}$$

where $\boldsymbol{n}_1$ and $\boldsymbol{n}_2$ are the sizes of the samples, $\overline{\boldsymbol{x}_1}$ and $\overline{\boldsymbol{x}_2}$ the means of the samples, and $\boldsymbol{s}_1^2$ and $\boldsymbol{s}_2^2$ the sample variances. Traditionally, substituting in these formulae gives value of the test statistic which is then compared with a value from the tables[1].

Rather than using a calculator to perform these calculations these days computations can be automated and tables made obsolete through use of statistical packages such as *Minitab* and *SPSS*, or the graphics calculator. The screen setup required to carry out an

[1]The formulae given are used when we cannot assume the variances of the two populations are equal. The assumption of equal variances was often made in the past because the distribution of the test statistic had otherwise only an approximate t-distribution. The common use of computers has meant that this is no longer an issue. Whether or not the assumption of equal variances is made the classical methods of solution are very similar.

independent group's *t*-test discussed here is shown in Figure 3.7, and the results of the test in Figure 3.8.

Figure 3.7 Screen of the TI83 graphics calculator illustrating the conduct of a *t*-test for two independent samples

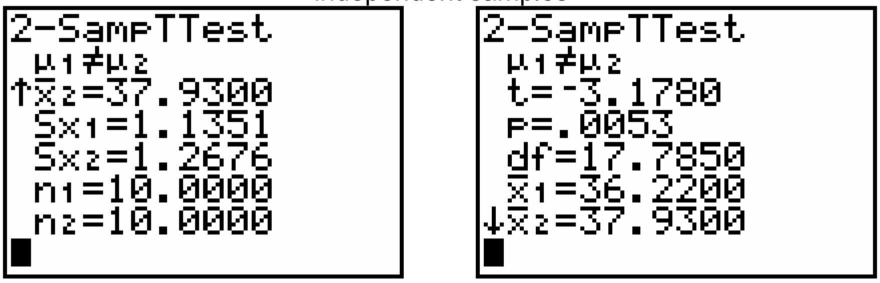


Figure 3.8 Screen(s) of the TI83 graphics calculator showing the results of a *t*-test for two independent samples

Of course the student must now relate the output generated by the computer or calculator to the original question, which would be, on the basis of the P–value (0.0053), to reject the null hypothesis and conclude that there is a difference in mean lasting time between the two brands of paint.

3.5.2 Computer intensive approach

Now, consider how students would proceed using a computer intensive method of solution. The observed mean difference in time the paint lasts for the two brands is 1.76 months. This difference may be due to a real difference in the lasting time of the paints, or it might well have occurred by chance. The null hypothesis is that there is no relationship between the brand of paint and the number of months it lasts. If this is so, all twenty observations can be regarded as coming from the same population. If this is true, the observed difference in the means, 1.76 months, would not be particularly unusual when compared to many values of this mean difference that could be found by randomly selecting two groups of these sizes from the same population. To proceed, the computer intensive procedure divides the twenty observations into two equal sized groups at random, designating one as Brand A and the other as Brand B, and computing the difference in means is computed. This is repeated many times, and as a result an empirical distribution for the difference in means is generated. The probability that there is a real difference in the means may then be estimated by determining the proportion of the randomisation samples that give a difference as far from zero as the actual difference observed. Further details regarding this procedure, the permutations test, can be found in Noreen (1989).

Computer intensive methods are just that, and cannot be performed easily without the aid of a computer. There are many computer packages readily available which will support computer intensive methods. One such package is *Models'n'Data* (Stirling, 1991, 1995), a Macintosh based computer package designed for teaching and learning statistics which offers the facility of using various computer intensive methods.
The empirical distribution of the difference in group means for the example under consideration shown in Figure 3.9 was generated by Models 'n' Data.

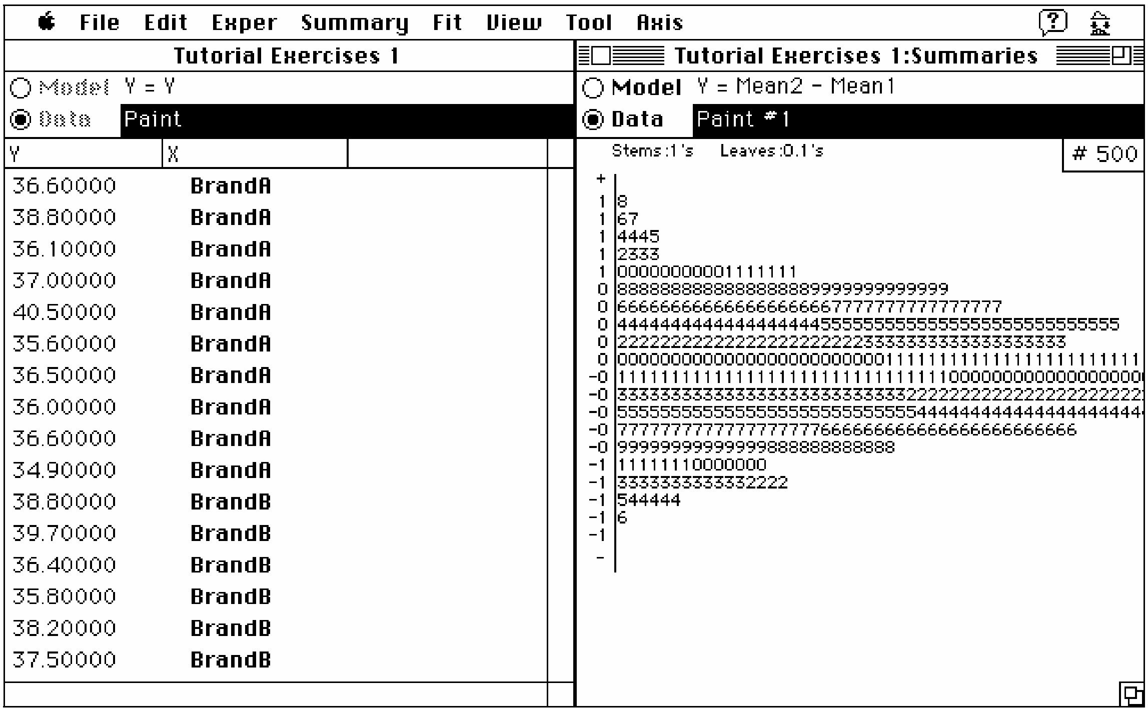


Figure 3.9 Empirical distribution of the difference between the sample means generated and displayed by *Models'n'Data*

In this example, an empirical distribution of 500 observations was generated, and displayed as a stem and leaf plot. From the plot it may be seen that there was only one value of the mean difference as far or further from zero than 1.76, giving our estimated P–value as 1 in 500 or 0.002. Thus we see that such a difference in means is very unlikely to have been caused by chance alone, and conclude that there is a real difference in the mean lasting times of the two brands of paint.

3.5.3 Comparison of methods

The essential components of a hypothesis test have been described by Noreen (1989) as follows:

> Three ingredients are usually required for a hypothesis test: A hypothesis, a test statistic, and some means of generating the probability distribution of the test statistic under the assumption that the hypothesis is true. (p.2)

It is in this process of defining a test statistic, and determining its distribution, where the essential difference between the traditional and computer intensive methods lies. That is, in the determination of the sampling distribution. A theoretical model for the sampling distribution is used in the traditional method. As such, the essential purpose of the process, to determine the likelihood of such a value of the test statistic from the sampling process, is easily lost. In traditional hypothesis testing the fundamental logic of the hypothesis test is complicated by the theory underlying the form of the theoretical sampling distribution of the test statistic.

The computer intensive method which is described here retains explicitly the link between the test statistic and the sampling process. Using a computer intensive method, an intuitively obvious test statistic is selected for which an empirical representation of the sampling distribution, based on the sampling process, is generated. The student remains linked to the overriding purpose of the exercise – the inference. Using technology in this way amounts to forming a partnership between the student and the computer in which the computer takes on the lower level tasks of performing the numerous calculations whilst the student undertakes the higher order tasks of actually applying the logic of inference to the particular situation. While the technology is also playing an important role in the case of the traditional solution, the nature of that role is very different, with the student possibly quite unaware of how the P-value is calculated, and what exactly is it measuring.

In fact, using the computer intensive described here, the student always must keep in mind the overall purpose of the exercise in order to carry it out successfully. Whilst the student of traditional inference needs to recognise certain situations, and make the appropriate assumptions, the students learning inference through computer intensive methods have been experienced a methodology which has application in most situations without further need for distribution theory.

Because the sampling process has remained explicit throughout the procedure, if students are able to be introduced to hypothesis testing via a computer intensive method such as that previously outlined, then the schema constructed by such students for hypothesis testing have a greater potential to be directly linked to their schema for sampling distribution. Thus, it seems possible that using a computer intensive method may enable beginning students of statistics, especially those without strong mathematical backgrounds, to connect the important principles of statistical inference to the process of sampling. Whilst at this stage little formal research exists to support this theory, what does exist seems encouraging (Simon, Atkinson, & Shevokas, 1976 for example).

One might then argue from a theoretical perspective that once the student has a functional schema for hypothesis testing, developed through learning introductory statistical inference through computer intensive methods, the theoretical model for sampling distribution when introduced would be appropriately linked both to the empirical representation and to hypothesis testing. Construction of a schema for statistical inference in which both the empirical and theoretical representations of sampling distribution were linked to each other and to hypothesis testing and estimation would suggest that students could invoke either representation of the sampling distribution depending on the nature of the external stimulus. This degree of connectedness could be seen to lead to the development of conceptual and procedural understanding in statistical inference, whether traditional or computer intensive methods were used, although one might ask why, in this computer age, anything but computer intensive methods is necessary. However, it remains necessary to teach students the traditional theories and methods of statistical inference, and these are standard strategies used in current statistical practice.

3.6 Concept maps

As has been discussed earlier in Chapter 2, concept maps may be used for many different purposes across a broad spectrum of knowledge domains. What are the potential roles for concept maps in a study investigating the relationship between the sampling distribution and the development of understanding in statistical inference? Three distinct roles have been recognised:

1. To facilitate expert analysis of the content domain of introductory statistical inference.
2. To promote reflective thinking concerning statistical inference by students.
3. To provide external representation of the students' schema for analysis by the researcher.

Earlier in this chapter the concept mapping exercises carried out by experts in order to deconstruct the structural knowledge implicit in a study of statistical inference were presented and discussed. In this section the other roles played by concept mapping exercises will be presented in more detail.

3.6.1 The concept map as vehicle for promoting cognitive reorganisation

As been argued earlier in this chapter, there are two distinct mathematical representations for sampling distribution and many symbolic and computer based representations for each of these. If a student is engaged in a metacognitive activity which requires them to reflect on the relationships between the various representations, it is believed that links between the representations may be formed (for example Schoenfeld, 1987; Novak, 1990b).

A well-documented strategy to promote the organisation and linking of a students' cognitive structure is the preparation of concept maps (for example Novak, 1990b; White & Gunstone, 1990; Jonassen, Beissner & Yacci, 1993). In prior studies it has been found that the actual process of constructing the concept map engages the student in metacognitive thinking, encouraging them to find structure and identify links within and between their schemas. According to Novak and Gowin (1984)

> Students and teachers constructing concept maps often remark that they recognise new relationships and hence new meanings (or at least meanings they did not consciously hold before making the map). (p. 17)

Thus, having students construct concept maps in the process of learning about statistical inference and, in particular, the sampling distribution, has the potential to encourage the learner to focus on the organisation of his or her cognitive structure and recognise the links between concepts within that structure. This activity, as an integral part of the teaching and learning strategy, is an overt strategy for facilitating links between the sampling process, the various representations of sampling distribution, and the logic and procedures of statistical inference, which are not explicit in the more traditional approaches to teaching introductory inference.

3.6.2 The concept maps as external representation of cognitive structure

The schema which an individual builds to accommodate a body of knowledge is a hypothetical construct. It is not possible to know such a structure directly. However, various methods of representing an individual's mental representation have been suggested by researchers including White and Gunstone (1990) and Jonassen, Beissner and Yacci (1993). Suggestions include concept maps, interviews, relational diagram, semantic maps, causal interaction maps, graphic organisers and many more.

As previously mentioned, concept maps have been used extensively in science education as a means of externalising the cognitive structure held by an individual (Novak, 1990b; Wallace & Mintzes, 1990). As stated by Novak and Gowin (1984)

> Because concept maps are explicit, overt representations of the concepts and propositions a person holds, they allow teachers and learners to exchange views on why a particular propositional linkage is good or valid, or to recognise missing linkages between concepts that suggest a need for new learning. (p. 19)

Thus, as argued in Chapter 2, concept maps are an accepted means of making an observable representation of a students internal mental representation at a given point in time and there is evidence from previous studies that they convey as much information as would be gained from other methods of externalising that structure, such as interview (Laturno, 1994).

Since this study is concerned with the role of the sampling distribution in understanding statistical inference, a method of externalising the student's conceptual structure for statistical inference is necessary. In particular, whether or not links are present between theoretical and empirical representations of sampling distribution and between the sampling distribution and statistical inference needs to be established. Thus, it seems that a sequence of concept mapping exercises can provide the necessary documentation concerning the structure and form of the students' conceptual structures for statistical inference, which can subsequently be analysed in order to address the research goals in this study.

3.7 The Research Hypothesis

This study aims to investigate the process of the construction, organisation and reconstruction of the schemas associated with learning statistical inference and the consequences of this process in terms of learning outcomes for a group of students studying introductory inferential statistics. In particular, it was hypothesised that, with regard to students' schemas for statistical inference:

> Students whose schema for sampling distribution demonstrates links to the sampling process and whose schema for statistical inference includes links to the sampling distribution, will show evidence of both conceptual and procedural understanding, whilst those who have not integrated these concepts into their cognitive structure may still demonstrate high levels of procedural understanding, but not of conceptual understanding.

3.8 Summary

In this chapter the theoretical framework which presented in Chapter 2 was operationalised in the context of introductory statistical inference. The present study aims to extend the existing body of research into the learning of introductory inferential statistics by establishing differences in the nature of the conceptual structures constructed by students who exhibit conceptual understanding from those who do not. As a consequence, this research may point to refinements in current teaching and learning strategies which facilitate this deeper level of understanding.

Chapter 4
The Design and Conduct of the Study

4.1 Introduction

The purpose of this study is to chart the changes in the cognitive structures of students as they participate in a learning journey. The knowledge domain through which they will travel is introductory statistical inference, and the student voyage will be navigated by the researcher and one of their teachers.

This study recognises the role that instruction has in the learning process for the student. However, the main purpose of this study is not to make hypotheses concerning the effect a particular instructional strategy has on the student knowledge structure. Rather, the purpose is to investigate how the cognitive structures constructed by each of the students in the study differ in qualitative terms, and whether the nature of these differences will enable in some way the prediction of learning outcomes.

4.2 Methodology

Schoenfeld (1994) in his article, *Some notes on the enterprise (Research in collegiate mathematics education, that is)* recommends a research genre where:

> …one makes explicit a set of theoretical assumptions about the nature of mathematical thinking and learning, about cognitive structures...The data–what then one sees–are then matched up against the theory. In the case of a good fit, the theoretical notions are substantiated and may be refined; in the case of a not–so–good fit, one may have to go back to the metaphorical drawing board, either theoretically (are the ideas wrong? not adequately specified?) or pragmatically (did the empirical treatment, or the curriculum and the way it was carried out, really reflect the theoretical ideas? If not, how can they be modified, and tried once again?). (p.9)

The current methodology in educational research, which is in accord to this description of research genre by Schoenfeld, is called a *teaching experiment*. The teaching experiment evolved as a legitimate research methodology in mathematics education in the 1970's (Romberg, 1992). Using this methodology, the researcher (who is often the teacher), attempts to ascertain the students' level of understanding in an area, intervenes instructionally and then investigates the effect of the intervention on the students' learning. In this sense, the teaching experiment allows the researcher to study learning at an individual student level. According to Romberg (1992), in a teaching experiment:

> ...hypotheses are first formed concerning the learning process, a teaching strategy is developed that involves the systematic intervention and stimulation of the student's learning, and both the effectiveness of the teaching strategy and the reasons for its effectiveness are determined. (page 57)

A teaching experiment usually consists of a sequence of teaching episodes which include a teaching agent, one or more students, and a method of recording what transpires during the episode (Steffe & Thompson, 1996). Traditionally, these records are extensive, and often involve videotape of individual student experiences which are analysed retrospectively in detail. The teaching experiment begins with exploratory teaching, where the teacher/researcher builds a good rapport with the students and, at the same time, gains some understanding of the nature of the ways in which the students are dealing with the mathematics. The researcher must also undertake an analysis of the knowledge involved in the teaching (e.g. Steffe, 1991). As a result of these or other experiences, the teacher/researcher formulates a hypothesis to be tested during the teaching experiment. Other hypotheses may be formulated within the time frame of the teaching experiment, perhaps between teaching episodes. By analysing the students conceptualisations, and how they have changed, the researcher then "constructed a model for mathematical learning that included the teaching action of the teacher" (Steffe & Thompson, 1996, p.20).

The methodology of the teaching experiment is appropriate for this study, in which an instructional strategy was designed based on an approach which recognised the experiential nature of the learning process, together with a detailed analysis of the structure of the knowledge domain concerned. An objective of this study was to document the development of the students' cognitive structures during and after a course of instruction specifically developed to facilitate the construction of important relationships between concepts in the students schema, in order to determine the existence and nature of qualitative differences between student cognitive structures.

Based on a theoretical analysis of the relationships in the mental structure necessary to support the development of understanding in introductory inference, predictions were made concerning the future abilities of students, and then the validity of these predictions was addressed.

However, the purpose of this study was to document conceptual development when a teaching and learning strategy was applied in a standard post-secondary educational context, therefore it made no sense to study individual students who were not undertaking the normal, group, student experience. Thus, a teaching experiment was undertaken in this study which was conducted in the post-secondary teaching environment under normal classroom conditions. This imposed certain constraints on the way in which the teaching experiment could be conducted. The time that the lecturer was able to spend with individual students was limited, as was the possibility for extended individual student-lecturer interaction. Consequently, the teaching experiment methodology needed to be modified in a way that retained the key elements which made it a legitimate research methodology, but which, at the same time, recognised the limitations imposed by the environment. With this in mind, the teaching experiment methodology was modified so as to retain the fundamental principles of the teaching experiment but be applicable in a post-secondary lecturing environment. Thus the methodology used may be summarised in the following set of key steps:

- Conduct an analysis of the knowledge involved in the teaching area.
- Develop a teaching strategy which is theoretically consistent within the conceptual model of schema theory of learning, and the development of connections within and between schemas.
- Conduct the teaching strategy developed with a group of students.
- Investigate the students' schemas at regular intervals throughout the conduct of the teaching experiment.
- Predict, on the basis of their cognitive structures, which students would be considered as successful in terms of the depth of their understanding of statistical inference after participating in the teaching experiment.
- Develop appropriate measuring instruments and use these to assess statistical understanding in the students.

The first of the key steps, the analysis of the knowledge domain, *introductory statistical inference* that was used in developing the instructional strategy, was carried out and reported on in Chapter 3. The remainder of the key steps is discussed in later sections of this chapter. Sections 4.3 and 4.4 detail the development of instruments that were used to assess understanding in statistics. The use of concept maps to record and facilitate of analysis of changes in students' conceptual structures is discussed in Section 4.5. The development of the instructional strategy used is documented in Section 4.6. Finally the setting and conduct of the teaching experiment are reported in Section 4.7.

4.3 Assessing understanding in statistical inference

An integral step in the teaching experiment outlined in the previous section was the assessment of the level of understanding of statistical inference attained by the students taking part in the study. In particular, the notions of procedural and conceptual understanding are important in this study and ways of measuring procedural and understanding need to be determined.

Traditionally, courses in statistics have trained students to carry out complex calculations and perform sophisticated statistical procedures from first principles, as can be seen from existing textbooks (see, for example, Johnson & Bhattacharyya, 1992; Kirk, 1990; Moore & McCabe, 1993). Mastery of this set of calculations and procedures, together with the relevant statistical vocabulary, has formed the basis of much assessment in statistics (Hawkins, Jolliffe & Glickman, 1992). However, it has been shown that students who are able to correctly solve a standard statistical problem may not understand either the question or the solution produced (Jolliffe, 1991). In fact, it is the belief of many statistics educators is that a large proportion of students understand little of what they are doing (for example: Konald, 1991; Well, Pollatsek, & Boyce, 1990; Williams, 1993).

The tendency for introductory courses to focus on procedural knowledge has long been noted by Pollatsek, Lima, & Well (1981), who state:

> One pedagogical point seems clear. In many introductory courses, students are taught to use formulas in a rote manner with the justification that a thorough understanding of the material can wait until the second course (or later). While it is undeniably true that students can solve some problems this approach, our data suggest that the range of problems that can be solved with only instrumental knowledge is vanishingly small.
> (p. 202)

While many introductory courses do aim to achieve conceptual understanding, their general failure to do this may be due in part to the style of assessment commonly used, which focuses on the assessment of procedural outcomes. And, such tasks may give a biased evaluation of student understanding. As stated by Vinner (1983)

> Students can succeed in examinations even when having a wrong concept image. (p. 305)

As suggested by Novak (1990b), students will actively avoid commitment to meaningful learning (Ausubel, 1978). This is because, from the student perspective, a depth of understanding is not required in their studies as it is not what is being tested.

Another reason why teachers fail to assess conceptual understanding is suggested by Demitrulias (1988), who believes that teachers concentrate on the computational aspects of statistics, as they fear students exploring statistics ideas beyond the routine. In order to emphasise conceptual learning and understanding rather than just computational proficiency, Demitrulias suggests some novel (more creative) ways to introduce conceptual ideas in class, such as:

> Suppose the measures of central tendency were applying for a job at a local hamburger joint. Prepare a short resume of their credentials (strong points, skills, job experiences, etc). (p. 169).

Similar ideas are expressed by others (for example: Hawkins 1986, Moore 1992a, Hogg 1992) who recommend that the amount of time devoted to repetitive calculation be reduced, and more time devoted to application of statistical concepts.

Researchers into teaching and learning of probability and statistics have recently seen as a high priority for research the development of appropriate instruments for the measurement of understanding (Garfield, 1993a; Gal & Ginsberg, 1994; Garfield & Ahlgren, 1988; Shaughnessy, 1992). As stated by Garfield (1994):

> Traditional forms of assessment of statistical knowledge provide a method for assigning numerical scores to determine letter grades but rarely reveal information about how students actually understand and can reason with statistical ideas or apply their knowledge to solving statistical problems. (p. 1)

Thus, considerable concern has been expressed concerning the need for appropriate instruments which measure both procedural and conceptual understanding. Whilst there exist a variety of tasks which can be used to measure procedural understanding, at the time of the conduct of the study, very little was available which had been trialed and validated that covered the content domain under investigation and measured conceptual understanding. Thus, it became apparent that such instruments would need to be developed for the purpose of the study.

If the view is taken that a student's cognitive structure, as it relates to statistical inference, reflects their level of understanding in the area, it follows that determination of a student's understanding would include methods for externalising that structure. In particular, assessment tasks should include means for identifying the links between concepts that have been established by the student. Since, as demonstrated in Chapter 3, the key concepts of statistical inference are complex and multifaceted, it also follows that the cognitive structures associated with these concepts would also be complex and multifaceted. Because of this complexity, it is unlikely that a single assessment task can reflect all aspects of understanding, and students need to engage in several different tasks in order to reveal, at least in part, their cognitive structure.

Consider once again the framework proposed by Putnam, Lampert and Peterson (1990), in which understanding was described around five themes: understanding as representation, understanding as knowledge structure, understanding as connections between types of knowledge, understanding as the active construction of knowledge and understanding as situated cognition. This framework allows the development of tasks that can be considered to measure aspects of either procedural or conceptual understanding. Tasks that fall within the classification of understanding as representation (the superprocedures described by Hiebert and Lefevre (1986)), are considered to measure procedural understanding in this study. Tasks which fall within any of the other four dimensions of understanding are considered to contribute to the measurement of conceptual understanding.

In order to develop a range of appropriate tasks, Nitko and Lane (1990) related the descriptions of understanding given by Putnam, Lampert and Peterson (1990) discussed earlier in Chapter 2 to understanding in statistics, and suggested the following assessment framework for statistics:

Procedural Understanding	
Understanding as representation	Tasks which involve application of standard notation, representation and algorithms to solve statistical problems. This would include standard applications of the *t*-test or chi-square test for example.
Conceptual Understanding	
Understanding as knowledge structure	Tasks which give insight into the knowledge structures of students. That is, tasks that demonstrate that the student has made a connections between concepts, such as hypothesis testing and confidence intervals for example.
Understanding as connections between types of knowledge	Tasks that require students to integrate formal knowledge with informal knowledge developed outside the class. This would include tasks requiring the interpretation of statistical concepts.
Understanding as the active construction of knowledge	Tasks that enable the teacher to monitor the development of knowledge over time, such as concept maps.
Understanding as situated cognition	Tasks which require the student to apply their knowledge in a variety of contexts, different from those previously seen and discussed in the classroom.

Using this framework as a guide, known work on assessment at this time was expanded and supplemented by the researcher which resulted in a set of tasks which covered the full range of aspects of understanding over the full content domain.

The known works used consisted of an instrument to assess conceptual understanding in probability and statistics was by Konald & Garfield (1993). Called the “Statistical Reasoning Assessment. Part 1: Intuitive Thinking”, this series of multiple choice questions build on earlier work of Konald and others (for example Falk, 1993; Kahneman & Tversky, 1972; Konald, 1991).

Specific details concerning the nature and purpose of these assessments are given in the next section.

4.3.1 Assessment tasks

In this section the assessment tasks developed for use in the study are presented, together with their purpose. The tasks were designed to measure procedural and conceptual understanding in statistical inference, as well as the range of the content domain that is the focus for the study. Thus, to ensure all content areas are covered, there are several tasks that pertain to the measurement of procedural understanding, each one concerned with a different sub-section of the content domain.

The tasks are named in order to facilitate discussion of the results of the study later in this thesis.

Task name

Sampling

Task

A survey is conducted with a random sample of 282 university students, in order to find out how far they travel to university each day. One student questions the validity of the study, noting that there are 4000 students at the university, not just 282. Read each of the statements listed below carefully, and select the ONE response that sounds the most reasonable to you.

A Agree, 282 is too small a percentage of the 4000 (7%) to allow us to draw conclusions.

B Agree, you should have a sample that is at least 50% of the population in order to make inferences.

C Agree, they should get all the students to participate in the survey.

D Disagree, 282 is a large enough number to use for these purposes if the sample was a random sample of students.

E Disagree, if the sample is random, the size doesn't matter.

Category

Understanding as knowledge structure (Conceptual).

Purpose and rationale

Adapted from the Statistical Reasoning Test (Konald & Garfield, 1993), the purpose of this task was to investigate student understanding of key aspects of the sampling process. In particular, to determine

- if the students accept sampling as a valid method of obtaining information about a population and
- if the student realises that the size of the sample is an important consideration.

Explanation

The best of the alternatives offered is D, which requires that the student has conceptually linked samples and populations, and appreciated that size of the sample is not related to the size of the population. Alternatives A and B suggest that the student erroneously believed that population size was important. Alternative C rejects sampling as a valid method of investigating a population all together, whilst alternative E suggests that sample size is not important.

Task name

Weather

Task

The Weather Bureau wanted to determine the accuracy of their weather forecasts. They searched their records for those days when the forecaster had reported a 70% chance of rain. They compared these forecasts to records of whether or not it actually rained on those particular days. On what percentage of these days would you expect to see rain had occurred if their forecast is very accurate?

A 95% – 100% of those days
B 85% – 94% of those days
C 75% – 84% of those days
D 65% – 74% of those days
E 55% – 64% of those days

Category

Understanding as connections between types of knowledge (Conceptual).

Purpose and rationale

Again adapted from the Statistical Reasoning Test (Konold & Garfield, 1993), the purpose of this task was to investigate whether or nor the students' reasoning was based on the *outcome* interpretation of probability (Konold, 1989). That is, that when asked to interpret a probability, the student does not do so in terms of a distribution of possible values but rather attempt to predict the result of a single trial. This instead of interpreting this forecast as "rain occurs on about 70% of such days" the student using the outcome interpretation predicts that rain will occur, since the probability quoted in more that 50%. For these students, any event associated with a probability of more than 50% should occur all the time, whilst those associated with probabilities of less than 50% should occur only rarely. This is relevant to the current study, since students need to be able to correctly interpret probabilities in order to interpret a P-value.

Explanation

In a study of 119 students conducted by Konold (1995), about 32% of students selected the correct alternative (D) but the most common response was A (36%). According to Konold:

> For these students, there seems to be little quantitative information conveyed in a probability value – a probability of 80% would seem to communicate no additional strength of belief over one of 70%. (paragraph 12)

Alternatives B, C and E given merely serve as distracters, and would most often be selected by students who were guessing their answer.

The task requires the students to relate their formal knowledge of probability to experiences outside the classroom (informal knowledge), indicating *Understanding as connections between types of knowledge*, an aspect of conceptual understanding

Task name

Hospital

Task

Half of all newborn babies are girls and half are boys. Hospital A records an average of 50 births a day. Hospital B records an average of 10 births a day. On a particular day, which hospital is more likely to record 80% or more female births?

A Hospital A (with 50 births a day)

B Hospital B (with 10 births a day)

C The two hospitals are equally likely to record such an event.

Category

Understanding as connections between types of knowledge (Conceptual).

Purpose and rationale

This task concerns the relationship between sampling variability and sample size, and was used in earlier research by Tversky and Kahneman (1982).

Explanation

Tversky and Kahneman found that 56% of undergraduate students incorrectly gave the answer C, suggesting that the majority of students believe that the variability of the sampling distribution is independent of the sample size. Obtaining the correct answer B, implies that the student appreciates that the variability in the sampling distribution of the sample proportion is larger when the sample size is smaller. Response A indicates that the variability of the sampling distribution is seen to increase with the sample size, and may indicate confusion between the sampling distribution and the distribution of the sample.

This task again requires the student to integrate formal knowledge with informal knowledge developed outside the class, an example of *Understanding as connections between types of knowledge* and contributing to the measurement of conceptual understanding.

Task name

t-test

Category

Understanding as representation (Procedural).

Purpose and rationale

This task is a routine, textbook type problem, concerned with the measuring the students' ability to carry out a standard *t*-test from first principles.

Task

According to a Census held in 1956, the mean number of residents per household in an inner suburb, Richthorn was 3.6. In 1995, a student randomly sampled 11 households from the suburb and recorded the number of residents in each with the following results

2	2	5	1	1	3	4	2	4	3	1

Can the student conclude that the mean number of residents per household in Richthorn has decreased since the 1956 Census.

Explanation

Students are given the necessary formulae and, assisted by a statistical calculator, determine the value of the test statistic for a *t*-test. They then use t-tables to determine the P-value which they will use to make a decision in this situation. Whilst these calculations are considered complex and challenging by some students, students are generally able to learn this procedure by repetition of similar tasks which are discussed in class and presented in written subject materials. Tasks of this type, where students are asked to carry out formal statistical tasks using the recognised structure, language and notation of the discipline fall into the category of *Understanding as representation*, and the ability to carry out this task successfully can be considered as evidencing procedural understanding.

Task name

Confidence interval

Task

Using a computer package, the student finds the 95% confidence interval for the mean number of residents to be (1.626, 3.465). Is this confidence interval consistent with what you found in part (a)* Explain.

*In the terminology used here, part (a) refers to Task *t*-test.

Category

Understanding as knowledge structure (Conceptual).

Purpose and rationale

The purpose of this task was to establish whether or not the students were able to see that hypothesis testing and confidence intervals are alternate ways of looking at the same problem. In this instance that is achieved by determining whether or not the hypothesised value for the mean lies within the given confidence interval, and thus identifying any inconsistency in the two sets of conclusions.

Explanation

Ability to do complete this task successfully, and explain the relationship between the conclusions based on the hypothesis test and the confidence interval, indicates that the students has linked the schema associated with these concepts in their mental structure.

Task name

Sample size

Task

Assuming that the number of residents in a household is normally distributed with a standard deviation of $\sigma = 1.4$, determine the minimum sample size needed to be 95% confident that the population mean μ will be within 0.5 units of the sample mean.

Category

Understanding as representation (Procedural).

Purpose and rationale

The content domain of the study included the affect of sample size on the width of a confidence interval, and calculations to determine sample size.

Explanation

To successfully complete this task the student was required to substitute into a given formula, a task which had been previously practiced in classes, and in doing so is considered to be demonstrating procedural understanding.

Task name

Two group *t*-test

Task

The following data was generated in a study of the effectiveness of in-service training of nurses. In this study, 15 nurses were given a test of their knowledge of cancer and cancer nursing prior to attending a one-day workshop on the topic. They were then retested on the same topic at the conclusion of the workshop. The results are displayed in the table below.

Nurse	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Pre	29	20	24	32	33	19	17	32	16	28	35	28	18	45	19
Post	35	41	33	41	39	20	29	42	36	37	36	33	35	42	19

The course organiser wanted to know whether there had been an increase in the nurse's knowledge of cancer and cancer nursing after attending the course. Two assistants, Roger and Annette, were given the task of analysing the data using *Minitab*.

Roger carried out a two-sample *t*-test with the following results:

```
TWOSAMPLE T FOR posttest VS pretest
           N      MEAN     STDEV   SE MEAN
posttest  15     34.53      7.14       1.8
pretest   15     26.33      8.29       2.1

TTEST MU posttest = MU pretest (VS GT): T= 2.90  P=0.0036  DF=  27
```

Annette converted the two columns of data into a single column of difference scores and carried out a one-sample *t*-test with the following results:

```
TEST OF MU = 0.00 VS MU G.T. 0.00

             N       MEAN     STDEV   SE MEAN     T    P VALUE
diffs       15       8.20      7.15      1.85   4.44    0.0003
```

(a) Which analysis is the more appropriate and why?

(b) Using the information generated in the analysis that you think is most appropriate, test the hypothesis that there has been an increase in the test scores for the nurses who attended the workshop. Quote the values of all relevant statistics.

Category

Understanding as representation (Procedural).

Purpose and rationale

As well as carrying out hypothesis test from first principles, the students in the study were expected to be able to carry out such tests using information generated by a standard statistical package.

Explanation

In carrying out this task the student chooses between the paired and independent *t*-test solutions offered as alternatives in part (a), giving the experimental design used as the reason for their choice. Part (b) of the task requires the student to carry out a *t*-test for paired data, using the given output with which they are familiar from computer laboratory sessions. The task involves the recognition and interpretation the language and notation of statistical inference as represented by a standard statistical package. This is a routine task in the study of statistics, concerned with *Understanding as representation* and contributing to the measurement of procedural understanding.

Task Name

Modelling

Task

A sample of 100 primary school children were asked which type of protection they preferred to use to protect their faces from the sun, a hat or sunscreen. Of the 100 children, 61 preferred to use a hat, and 39 preferred to use sunscreen.

(a) Use the computer generated sampling distribution given to test if there is a significant difference in the proportion of students who show preference for one type of protection over the other.

```
        Stems:0.1's   Leaves:0.01's
+ |
7 |
7 |
7 |
6 |
6 | 6
6 |
6 | 2 2 3
6 | 0 1
5 | 8 8 8 9 9 9 9
5 | 6 6 6 6 6 6 6 6 6 6 6 6 7 7 7 7 7
5 | 4 4 4 4 4 4 4 4 4 4 5 5 5 5 5 5
5 | 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3
5 | 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
4 | 8 8 8 8 8 8 8 8 8 8 8 8 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9
4 | 6 6 6 6 6 6 6 6 6 6 6 6 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7
4 | 4 4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
4 | 2 2 2 2 3 3 3 3 3 3
4 | 1 1
3 | 8
3 |
3 | 4
3 |
3 |
2 |
2 |
- |
```

Stemplot showing 200 values of the sample proportion obtained from drawing samples of size 100 from a population with proportion p = 0.5.

(b) Use the normal model for the sampling distribution to test if there is a significant difference in the proportion of students who show preference for one type of protection over the other.

Categories

Understanding as knowledge structure (Conceptual)

Understanding as representation (Procedural)

Understanding as connections between types of knowledge (Conceptual).

Purpose and rationale

A main focus of this study is the consideration of the students' mental structures associated with sampling distribution and in particular whether or not the students have available to them both the empirical and theoretical representations of sampling distribution when dealing with problems in statistical inference. This modelling task was specifically designed to investigate whether or not the students were able to solve the same hypothesis testing problem using both an empirical sampling distribution, generated by repeated sampling, and the normal model for the theoretical sampling distribution which can be used in the same situation.

Explanation

Ability to carry out the hypothesis test using the empirical sampling distribution is taken as evidence of a link between the hypothesis testing procedure and the sampling process in the students' mental structures. This aspect of the task can be considered as concerned with conceptual understanding in the category *Understanding as knowledge structure*. Carrying out the hypothesis test using the normal model, however, is a more procedural task involving the use of formulae and tables, disassociated with the sampling process and evidencing *Understanding as representation*, which is procedural.

However, recognizing the equivalence of the two parts of the task, and in particular the expectation of consistency between the results of the two analyses falls into the category of *Understanding as connections between types of* knowledge, again as aspect of conceptual understanding. It was considered possible that some students would not be able to complete one part of the task but successfully complete the other, as they are measuring different aspects of understanding.

Task name

Correlation

Task

The following data was collected in 1970 in an investigation to determine if the amount spent by a school district on public education is dependent on the per capita income of the school district.

Income ($)	4198	4008	6197	5343	5928	3764	3244	4612	4918	2194
Expenditure ($)	871	850	1210	1188	1547	834	623	1052	1256	476

(a) Assuming that the sample of states has been drawn at random from a very much larger population of states, can we conclude from this data that there is a relationship between expenditure on education per student and per capita income for school district in general? Justify your response by carrying out a hypothesis test on *r*.

(b) The value of r may or may not be statistically significant, but is it of practical significance? Justify your answer by calculating the value of an appropriate statistic.

Category

Understanding as representation (Procedural)

Understanding as knowledge structure (Conceptual)

Purpose and rationale

This task is designed to assess the student's ability to carry out an hypothesis test for the correlation coefficient, and then to evaluate the practical significance, or real-world importance, of the relationship identified.

Explanation

Part (a) is a standard hypothesis test for Pearson's *r*, requiring the use of previously rehearsed standard language and algorithms of hypothesis testing. This part measures *Understanding as representation*, and contributes to the overall measure of procedural understanding. Successful completion of part (b) of the task is taken as evidence of a cognitive link between the concepts of statistical significance, as measured by the P-value, and practical significance, which in this instance is best measured by the coefficient of determination. The ability to exhibit the formation of this link demonstrates *Understanding as knowledge structure*, a feature of conceptual understanding.

Task name

Chi-square

Task

A study was conducted to determine if there is a relationship between socioeconomic status and attitudes to an urban-renewal program. The results are as shown.

	Urban renewal project		
Socioeconomic status	Disapprove	Approve	Total
Middle	90	60	150
Lower	200	100	300
Total	290	160	450

Does the data support the contention that there is a relationship between attitude to the urban renewal project and socioeconomic status? Carry out an appropriate hypothesis test.

Category

Understanding as representation (Procedural).

Purpose and rationale

To assess the students' procedural understanding in another area of the content domain, that of relationships between categorical variables.

Explanation

Here the student must to carry out a routine chi-square test of significance, concerned with the relationship between two categorical variables, from first principles. The formulae necessary for the calculation of the expected frequencies and the test statistics were given, and similar examples previously discussed. Requiring the use of standard notation and algorithms categorises the task as measuring procedural understanding, under the sub-heading of *Understanding as representation*.

Task Title

Unknown test

Task

> A researcher wishes to know whether blood pressures became more variable after a particular treatment. To determine this she carries out an F-test (which you have not been taught) which can be used to test for the equality of variance in two independent samples.
>
> For the treatment group (n =10) the sample variance was found to be = 76.44 while for the placebo group (n = 11) the variance was = 34.82 giving an F statistics of 2.20, and a P-value of 0.26.
>
> Write down appropriate null and alternative hypotheses and use the P-value to draw a conclusion about the variability of blood pressure in the two groups.

Category

Understanding as knowledge structure (Conceptual).

Purpose and rationale

Conceptual understanding is considered by some (for example Skemp, 1987) to be demonstrated by the ability to assimilate new situations into the appropriate existing schema. Successfully identifying a situation as a specific example relating to a generalised schema, and relating the key features of the new situation correctly to those previously studied, gives evidence as to the existence and nature of that generalised schema in the students mental structure. This task, which can be considered as measuring *Understanding as knowledge structure*, required the student to recognise a novel scenario as a hypothesis testing situation, and identify the key components of the problem. The task was concerned with statistical testing as applied to the variance. Whilst the students were familiar with the notation for population variance (σ^2), they had not to the researcher's knowledge carried out any testing concerned with variance, nor seen an F-statistic before.

Explanation

Students who were unable to successfully complete this task may have constructed separate schemas for various examples of hypothesis testing, but not successfully synthesised these schemas into a generalised schema for hypothesis testing where each individual testing scenario is seen as an example of the over-riding principle. Such integrative reconciliation reflects conceptual understanding of this aspect of statistical inference.

Task name

Explanation

Task

As a part of your research, you are investigating the relationship between the intelligence of a child and the intelligence of their mother. To this end, you administer intelligence tests to the mother and eldest child of a randomly selected sample of 30 families. A scatterplot of the data obtained indicated the presence of a moderate linear relationship between the intelligence test score of the mothers and their children and the value of Pearson's r was found to be r = 0.5135. You carry out a hypothesis test as shown below and conclude that the data you have supports your long held contention that there is a relationship between the intelligence of children and the intelligence of their mother in the general population.

A friend, who is very interested in your research, but who understands little about statistics, asks you to explain to him how you have come to this conclusion. They are happy that there is a relationship between the mother's and the eldest child's intelligence in the sample, but can't understand how you can generalise your result to include mothers and eldest children in general. You explain that this is the purpose of the hypothesis test you performed.

In the space provided in the table give a brief explanation of each step in the hypothesis test which you have carried out, so that your statistically illiterate friend is able to understand what you have done and how you were able to draw your conclusion. You can assume that your sample is properly representative of the general population.

Steps in your hypothesis test	Explanation
Hypotheses: H_0: $\rho = 0$ H_1: $\rho \neq 0$ non-directional test	
Significance level: $\alpha = 0.05$	

Test statistic r = 0.5135 for n = 30 pairs of data values	
P-value P-value = 2 x P(r > 0.5) = 2 x 0.0025 = 0.005 or 0.5%	
Decision & conclusion As p< 0.05, reject H_0 and conclude that there is a relationship between the intelligence of children and their mothers in the general population.	

Category

Understanding as connections between types of knowledge (Conceptual).

Purpose and rationale

This task was designed to elucidate further the students' conceptual understanding of statistical inference, by requiring them to interpret the (procedural) steps in the hypothesis test in their own language.

Explanation

By linking the formal notation and algorithms with which they are familiar with informal knowledge which can be understood by most people without specialist statistical training the student in demonstrating *understanding as connections between types of knowledge,* an aspect of conceptual understanding

Task name

Radio

Task

A radio station claims to its advertisers that 20% of 18–25 year olds listen to this station between 6.00 pm and mid-night on weeknights. A market research company carries out independent research on behalf of an advertiser and finds that only 15% of their sample of 18–25 year olds listen to the radio station in this time period. The advertiser concludes that the radio station is misleading them. What do you think? Try to include all the relevant reasons for your answer.

Category

Understanding as situated cognition (Conceptual).

Purpose and rationale

This task is designed to ascertain which of the relevant statistical schema are activated when the students are asked to consider this real world context. The question is quite intentionally open ended, and contains insufficient information for an exact answer to be obtained using a standard algorithm.

Explanation

A complete discussion would consider both issues concerned with the sampling process, and issues concerned with the sample size. All students should consider the sampling process as a potential problem, as this is certainly a valid possible explanation. Consideration of the sampling process alone, however, explanations implying that there must be a problem with the data collection, suggests that the student has not recognised the link between sampling and the sampling distribution in this scenario. Consideration of the sampling distribution explanation suggests that the student recognises the plausibility of the difference between the population parameter and the sample statistic. Further identifying the important role of sample size suggests that the student's conceptual structure for sampling distribution includes recognition of the role of sample size in explaining sampling variability.

In summary, a range of tasks was developed which are designed to measure aspects of either procedural or conceptual understanding in the content domain under investigation. Completion of these tasks enabled each of the students in the study to be allocated two scores, one measuring their level of procedural understanding, and the other measuring their level of conceptual understanding. In Chapter 6 the relationship between the nature of the students' schema for statistical inference, and in particular the role of the sampling distribution in that schema, to the student scores on these measures of understanding will be investigated.

4.4 Preliminary study and validation of concept maps

Before embarking on the main study, some preliminary studies were undertaken. The details of these studies, their purpose and how they relate to the main study are discussed in this section.

4.4.1 The potential of the concept map

Before embarking on the main study, preliminary studies were undertaken to explore the potential of concept mapping as a tool for both externalising a student's cognitive structure and monitoring any changes occurring in that cognitive structure during the period of instruction. To this end a study was carried out to investigate the ability of the concept map to reflect qualitative changes in a student's schema which developed during a computer based activity designed to reinforce the relationship between key concepts related to the sampling distribution. In this activity students were introduced to the sampling distribution of a proportion via a physical sampling activity using a box of coloured beads. They then participated in a computer-based activity designed to illustrate key features of the sampling distribution, such as its shape, centre and spread, and the relationship between spread and sample size. The study was conducted with a group of 31 students undertaking a part-time graduate course in statistics in the year prior to the main study (1994).

Methodology

In this investigation the students were asked to construct a concept map before the computer session, using the following terms: *centre, constant, distribution, estimate, normal distribution, population, population parameter, population proportion, sample, sampling distribution, sample proportion, sample statistic, sampling variability, shape, spread, variable*. The list of words was selected according to the objectives of the teaching sequence. The software used was *Sampling Laboratory* (Rubin, 1990) which is discussed in detail later in this Chapter. Students were advised that these words were merely a suggested list, and any words could be omitted, or others added. (This was the usual way in which these students had constructed the maps in the past.) The words were listed down the left-hand side of a large sheet of paper, and the students were requested to construct their map using pencil only. When the students had completed the maps, they were collected and photocopied. After the completion of the computer sessions, the maps were returned to the students, who were requested to modify their maps in any way that they felt was appropriate. The resulting maps were again recorded.

Scoring concept maps

There are many ways in which a concept map may be interpreted. The original way, as proposed by Novak & Gowin (1984), involves a scoring system where points are allocated on the basis of the number of concepts involved, the number of levels in the hierarchy used to map the concepts, and the number of cross links to form propositions which are made between separate strands of the concept map. Whilst this method was successful for their purposes, it is an inappropriate method in a situation where the students are given a list of terms to use, as it would not give emphasis to the structure of the map which reflects the relationships between concepts. The relevant issues here were if the terms had been linked together to form propositions, whether or nor the linkages were correct and thus formed valid propositions, and if important links had been recognised. Hence, an alternative technique for evaluating concept maps suggested by Kirkwood (1994) was used in this study. This approach is based on the inclusion in the student's concept map of the 'key' propositions that have been earlier identified by consultation with 'experts' in the knowledge domain. Table 4.1 lists the propositions identified by content experts, the researcher and a colleague to be key to this particular topic.

Table 4.1 Key propositions identified by experts.

Key Propositions	
A	Populations give rise to samples.
B	A population has a distribution.
C	Attributes of population distributions are described by parameters.
D	Parameters are constant.
E	Attributes of sample distributions are described by statistics.
F	Statistics are variable, and can be described by a distribution.
G	The sampling distribution of $\hat{p}$ is approximately normal.
H	The sampling distribution of $\hat{p}$ is characterised by shape, centre, spread.
I	The spread of the sampling distribution is related to the sample size.
J	The sample statistic can be used to estimate the population parameter.

Results

The following results were recorded for each student: the number of the key propositions proposed which were present in their first map (A-J), the number of key propositions which were present in their second maps and any change in the number of propositions included. The results are summarised in Table 4.2.

Table 4.2 The number of key propositions identified in the students' concept maps, before and after the computer sessions.

Student	Map 1	Map 2	Change	Student	Map 1	Map 2	Change
1	3	3	0	17	4	4	0
2	4	4	0	18	3	4	1
3	2	2	0	19	5	9	4
4	4	6	2	20	7	7	0
5	4	5	1	21	2	7	5
6	6	6	0	22	0	0	0
7	0	5	5	23	4	6	2
8	3	3	0	24	4	5	1
9	4	4	0	25	2	4	2
10	2	6	4	26	4	6	2
11	7	7	0	27	6	6	0
12	5	7	2	28	3	6	3
13	4	5	1	29	2	4	2
14	6	7	1	30	7	7	0
15	3	6	3	31	4	4	0

16	7	9	2	

From Table 4.2 it can be seen that the number of propositions added to the students' concept maps during the experiment varied considerably. To give some indication of the extent to which the students changed their maps, a frequency distribution of the number of new propositions the student added to their second map was constructed. This is given in Table 4.3.

Table 4.3 Frequency distribution of the number of key propositions added to the second map (N=31).

Number of propositions	0	1	2	3	4	5
Frequency	13	4	7	2	3	2

Also of interest was the frequency with which different propositions were included in the students' concept maps. This is shown in Table 4.4, which summarises the number of students including each proposition, before and after the computer session.

Table 4.4 Frequency tables of propositions identified in the students' concept maps, before and after the computer sessions (N=31).

	Proposition									
Total	A	B	C	D	E	F	G	H	I	J
Before	24	5	24	3	19	14	11	6	-	12
After	24	6	29	5	22	17	15	19	3	18
Change	**0**	**1**	**5**	**2**	**3**	**3**	**4**	**13**	**3**	**6**

Discussion

The data shows that the concept maps enabled conclusions to be drawn on the basis of this study about the ability of the concept map to reflect even subtle changes (one or two propositions) in the students' mental structures. The results showed that participating in the computer workshop was associated with a change in mental structure as evidenced by the maps for many students, with 18 out of 31 (58%) students adding one or more propositions to their maps, and 7 of the 31 (23%) students, adding three or more. After the workshop 27 of the 31 (87%) of students had included at least 4 of the listed key propositions on their concept maps. Thus, the concept maps allowed the investigator to observe growth in the students' conceptual understanding. The understanding of a concept exhibited by a student will often be incomplete or faulty, and yet, using the concept map, even the most minor change in understanding could be observed and identified. At the same time, it was clear that each student had a unique concept map, and where changes were made, students changed their maps in different ways. This individual evolution of knowledge is consistent with the theory that understanding is constructed by the student on the basis of their current cognitive structure, rather than received from an instructor on an information transfer basis (Cobb et al.1992).

If we assume that concept maps do in some way serve to externalise a student's mental structure in a particular knowledge domain, then the previous results show that the concept map can also be used to chart the changes in their mental structures as they undergo instruction. Thus, concept maps allowed the researcher to document changes in the students' schemas and to identify links and connections to other related schema. On the basis of concept maps, qualitative differences in the schemas which students constructed for sampling distribution were investigated, as was the extent to which the links to sampling distribution had been established and maintained by an individual student when ideas of statistical inference were introduced.

4.4.2 Investigating the assessment tasks

Prior to the main study, a preliminary study was undertaken with a similar group of 35 post-graduate students as described in Section 4.4.1. The aim of this study was to trial some of tasks which had been developed to be used in the main study to measure procedural and conceptual understanding. In particular it was important to establish that the students understood what was required of them by the task, that they were able to complete the task with some degree of confidence, and that the results of the tasks adequately reflected the variation between students, that is that they were able to discriminate between students of different ability. More importantly, the relationship between the concept maps produced by these students, and their scores on a procedural and a conceptual task was investigated, in order to explore the nature of the information generated by the concept maps.

The study was designed to investigate student performance on a concept mapping task, a task designed to measure procedural understanding (see Section 4.3.1, *t*-test) and a task designed to measure conceptual understanding (see Section 4.3.2, explanation), and further to examine the relationships between performances on each task.

Methodology

The students were asked to prepare a concept map which summarised their concept of statistical inference, using a specified list of terms. These were: Confidence interval, Decision and Conclusion, Estimation, Hypotheses, Hypothesis Testing, Inferential Statistics, Interval Estimates, P-value, Point Estimates, Population, Sample, Sample Statistics, Sampling Distribution, Significance Level, Statistical Significance, Test Statistic. Students were instructed that not all terms needed to be used, and that other terms not on the list could be included if required. The maps were scored on the basis of the number of valid propositions that were included in the maps. The students were given two weeks over which to prepare their maps, and were able access any resource material they desired.

The second task required the students to explain the steps in a hypothesis test in non-technical language. The actual task was similar to that described in Section 4.3.1, except that the test to be explained was a *t*-test. The students were required to carry out this task in class, and without access to texts or notes. Their answers were scored numerically, with a maximum mark of 12 using a scoring scheme given in Appendix 1A. Answers which simply rephrased or described in words the statement given scored poorly, while those which attempted to describe the step in the hypothesis test in conceptual terms scored well.

The third task was a standard *t*-test task, which students have seen many times before, carried out in class, again without access to written materials. These answers were scored numerically, with a maximum available mark of 10. The marking scheme for this task is also given in Appendix 1B.

Results

Each of the students completed and submitted the concept map for statistical inference. The distribution of the number of valid propositions shown in the student concept maps for statistical inference is shown in a histogram in Figure 4.1.

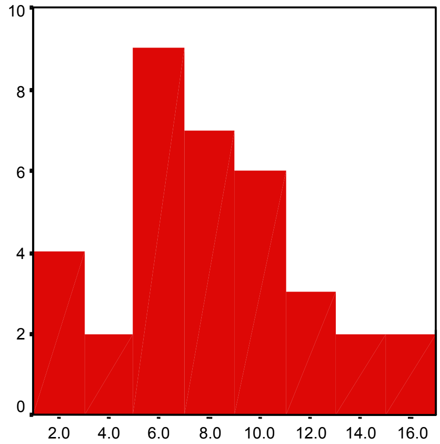


Figure 4.1 Distribution of the number of valid propositions shown in the student maps

The distribution of the number of valid propositions in the student maps was approximately symmetric, and ranged from 1 to 15, with a mean of 7.5, and a standard deviation of 3.8. (It may be of interest to note that the number of valid propositions in the 'expert' map was 17.) If we assume that the variation in the number of valid propositions included in the map is indicative of variation in the students' conceptual structures then this is again evidence that even though students have participated in the same teaching sequence they have experienced individual learning journeys.
All students completed the explanation task, and the distribution of the scores achieved for this task are shown in the histogram in Figure 4.2.

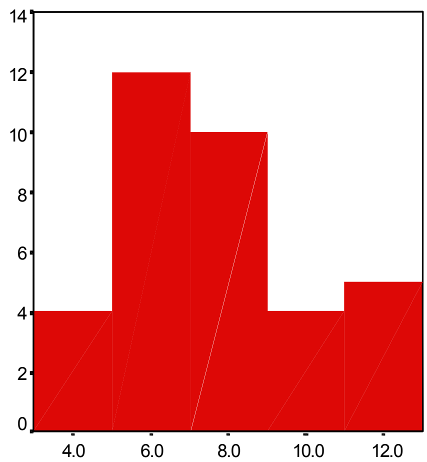


Figure 4.2 Distribution of the scores achieved by students for the explanation task

This distribution of scores is also approximately symmetric, ranging from 3 to 12. The mean score on the task was 7.1, with a standard deviation of 2.4.

As can be seen from the histogram of scores obtained by the students for the *t*-test task shown in Figure 4.3, the distribution is clearly negatively skew, with the marks achieved on this task ranging from 0 to 10. There were a few students (5 or 14%) who score 4 marks or less, but the majority (30 or 86%) scored 8 or more marks, with 24 students (69%) achieving full marks.

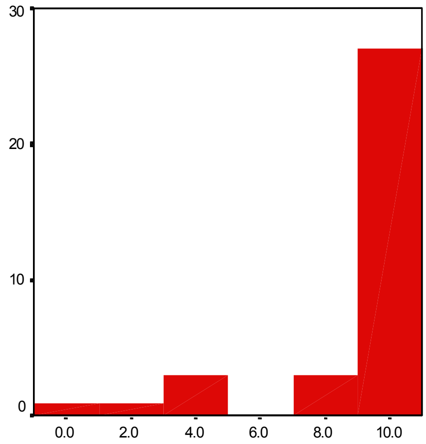


Figure 4.3 Distribution of the scores achieved by students for the *t*-test task

Discussion

The distribution of scores for the *t*-test task is markedly different from the distributions achieved for the other two tasks. This is not unexpected since the *t*-test task is primarily measuring procedural understanding, whilst the explanation task and the concept mapping task are both measuring conceptual understanding. To further investigate these relationships, Pearson product moment correlation coefficients were calculated and these are given in Table 4.5.

Table 4.5 Correlation between assessment tasks (N=35)

	Explanation	Standard Test
Concept map	0.455*	0.381*
Explanation		0.097^{ns}

* indicates $p<0.05$

As may be seen from Table 4.5, there was little correlation between the explanation question and the *t*-test question, indicating that students were able to carry out successfully standard tests without understanding the procedure, and vice-versa. This is consistent with the general feeling of those involved in teaching statistical inference who have long believed that it is possible to carry out the algorithms required to perform a hypothesis test without any understanding of the underlying principals (for example Konald, 1991; Well, Pollatsek, & Boyce, 1990). In terms of the theoretical model of understanding used in this study, these correlations reaffirm the contention that the procedural understanding needed to correctly carry out a hypothesis test may or may not be indicative of the possession of the deeper conceptual understanding that is required for students to be able to explain what they were doing in terms of sampling and sampling variability when carrying out an hypothesis test.

Also of interest is that the correlations show that the students' concept map scores are significantly positively correlated with both the explanation task (taken in this study as a measure of conceptual understanding) and the standard test (taken in this study as a measure of procedural understanding). That is, those students able to identify more valid propositions between the terms given for the construction of the concept map for statistical inference scored more highly on both the explanation task and the *t*-test task. This result in consistent with model for understanding which equates conceptual understanding with more highly organised and linked schema, but still retaining the relationships consistent with procedural understanding.

In summary, it has been assumed that the concept map is a viable tool for externalising at least part of the students' conceptual structure, enabling qualitative differences between these conceptual structures to be examined. Changes in these conceptual structures over time can also be studied by analysis of a series of maps for each student, and the presence or absence of important conceptual links ascertained. The preliminary studies confirm that the explain task and the *t*-test task are measuring different dimensions of student understanding, as predicted by the theoretical arguments. Finally, they establish that the complexity of the students' conceptual structure, as represented by the concept map is related to the student's achievement in tasks measuring both procedural and conceptual understanding.

4.5 Instructional intervention

It is not the purpose of this study to evaluate the effectiveness or otherwise of the teaching and learning model used to develop the instructional intervention. The instructional intervention is described here so that any changes that might occur in the students' procedural and conceptual knowledge can be interpreted in the light of the learning experience.

4.5.1 Introduction to concept maps

It was important that the students were able to construct concept maps before the intervention took place, otherwise there would be a possible confounding effect with students' maps reflecting both their knowledge structure and their concept map construction ability.

To ensure this at the beginning of the instructional sequence the students were introduced to concept mapping as a technique for summarising and externalising their knowledge structure. Prior to the instructional intervention, all students had spent some weeks studying descriptive statistics. At regular intervals during this period, the students constructed concept maps for various aspects of this material. Initially, maps were constructed by the instructor and students together with the instructor modelling the map construction process. Later maps were completed by the student alone, but discussed with the instructor after completion. By the time the students commenced the section of work which forms the basis of this study, they were able to construct concept maps quickly and confidently.

4.5.2 A physical sampling activity

The sampling distribution was introduced through a physical sampling activity, as the physical representation can be considered the most concrete and accessible representation of sampling available. This activity was conducted as a class exercise and the worksheet, which was used by students in the conduct of this exercise, is given in Appendix 2. In this activity a box containing both coloured and white beads was shown to the class and students were asked to guess the proportion of coloured beads in the box and to write down this guess. In order to gain some information about the reasonableness of the guessed value, the idea of taking a sample was introduced. Each student then used a sampling shovel (a flat wooden tray with a twenty-five holes drilled in it of the same diameter as the beads) to select a random *sample* from the *population* of beads in the box, and the *proportion* of coloured beads in the sample was calculated for each sample. The proportion of coloured beads in the box was defined as a *population parameter*, fixed for this box at this time, and the appropriate symbol introduced to represent this quantity (p). The proportion of coloured beads in the sample was defined as a *sample statistic*, clearly varying from sample to sample, and the appropriate symbol to represent this quantity introduced ($\hat{p}$). Particular care was taken to differentiate between p and $\hat{p}$ and what they represent throughout the class. Each value of $\hat{p}$ observed by a student was added to a histogram which was constructed as the samples were taken.

At the completion of the sampling exercise, the histogram showed that the values of $\hat{p}$ had formed a symmetric, single peaked distribution, which was named as a *sampling distribution*. Students were then asked if they wished to revise their guess for p on the basis of this sampling distribution. Class discussion of the sampling distribution led to the consensus that the "best" guess for the value of the population parameter would be the centre of the sampling distribution. It was also pointed out that, whilst the sampling distribution showed there were several possible values for the population parameter, some values which were observed sat well away from the sampling distribution, and were clearly unlikely values of p. The value of the population parameter was then given ($p = 0.2$ in this instance), and students acknowledged that the sampling process was a reliable method for gaining information about a population parameter, as this value was very close to the value agreed on by the class as the centre of the sampling distribution (which was in fact also 0.2 in this case).

4.5.3 The empirical sampling distribution

To serve as a bridge between the physical sampling activity and computer representations of the sampling process which were planned to be used extensively throughout the instructional strategy, the computer sampling package *Sampling Laboratory* (Rubin, 1990) was demonstrated next. This package provides dynamic linked representations was used to replicate the physical sample exercise, with coloured and white balls and a population proportion of coloured balls the same as that in the box (20%). The package requires no programmer as such by the user, who simply makes the appropriate entries as requested. The screens shown by *Sampling Laboratory* during the computer-based sampling exercise can be seen in Table 3.7. The move to a computer based sampling process enabled more samples to be drawn in a short time, and there are many computer packages available that will facilitate such sampling exercises (for example, *Minitab* or *Excel*).

A particular advantage of *Sampling Laboratory* over many other computer based sampling packages is that it creates a working model of the sampling process which has as its product both a histogram displaying the resulting sampling distribution (Window 3), and also the process by which the sampling distribution is obtained (Window 2). With *Sampling Laboratory* the sampling process can be observed in real time and students see the sampling distribution form as more and more samples are taken. Using *Sampling Laboratory* it is possible to illustrate both the sample-to-sample variability of the sample proportion ($\hat{p}$) as well as the formation of a (predictable) pattern when the distribution of the resulting values of the sample proportion are examined. It is important pedagogically for the students to appreciate that, while the values of the sampling proportion ($\hat{p}$) differ in an unpredictable way from sample to sample, the form of the sampling distribution which is produced by these values is quite predicable.

The use of dynamic linked computer representations to depict the sampling process, with all calculations automated and remaining very much in the background, is in accord with the theories of Kaput (1992), Pea (1987) and others concerning the use of computer software to facilitate the development of conceptual understanding. The analysis of the content domain undertaken in Chapter 3 established that an understanding of the sampling distribution and its relationship to the sampling process was central to the learning of inferential statistics. Static representations illustrate the product (pattern) only, while the dynamic linked representations illustrate the product and process and how they relate. Thus, the use of dynamic linked representations has the potential to considerably facilitate the formation of conceptual links between the empirical sampling distribution as an entity and sampling as the process by which the empirical sampling distribution is formed. At the same time, the similarities and differences between the sample proportion and the population proportion are explicit at all times.

Again a computer sampling package was used to generated dynamically an appropriate sampling distribution. This time, however, the package *Models'n'Data* (Stirling, 1991) which was described in Section 3.5, was chosen. When set up to undertake repeated sampling, *Models'n'Data* shows the results of each sample at each selection, and builds a sampling distribution of the sampling proportion at the same time. The disadvantage of *Models'n'Data* is that is does not show the distribution of the population at the same time as is possible with *Sampling Laboratory*, but it does have the advantage of offering many graphical representations for the sampling distribution, including histogram and stemplot.

4.5.4 Introduction to hypothesis testing

Hypothesis testing was introduced in the context of a real world example, based on a newspaper article (Fuller, 1992) headlined "Doubt on Letters Promise". In the article the reporter recounts his attempt to test the claim of the postal authority that 96% of letters are delivered on time. The reporter posts 59 letters and discovers that only 52 are delivered on time, a proportion of 88.1%. He then goes on to claim that this is evidence that the postal authorities are overstating their claim that 96% of letters are delivered on time. The question posed to the class is this: On the basis of the evidence, do we doubt the postal authority claim?

Shown in Figure 4.4 are the results of the 79th sample, shown as a bar chart, and a stemplot of the sampling distribution after 79 samples have been drawn. The value on the stemplot corresponding to the sample shown is covered by a black box, which is flashing when viewed on the computer screen.

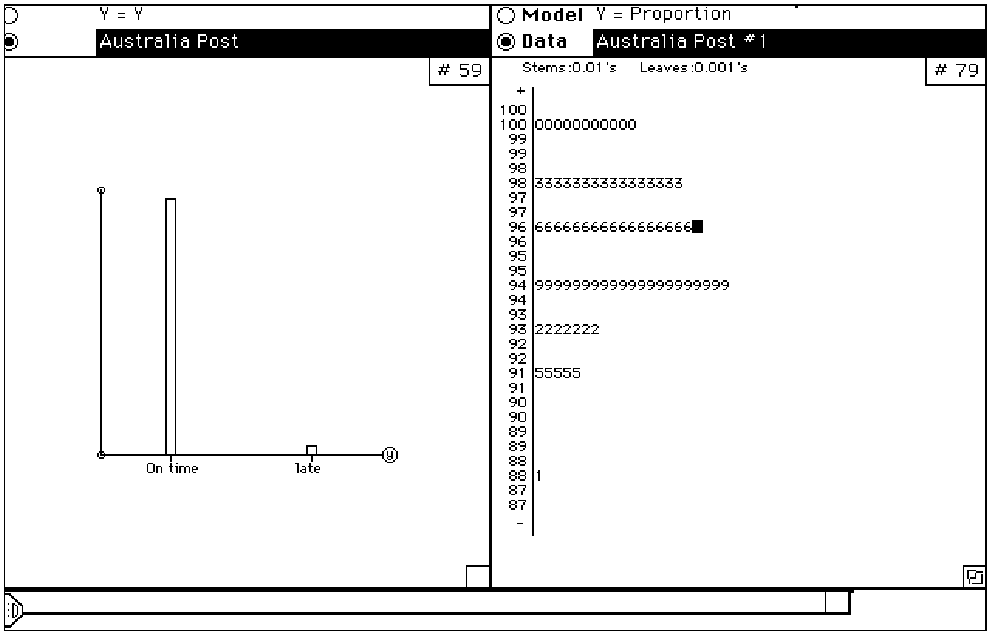


Figure 4.4 *Models'n'Data* screen showing the results of the 79th sample in a histogram, and a stemplot of the sampling distribution after 79 samples have been drawn.

The stemplot was selected as the preferred graphical representation because each data value in the sampling distribution is clearly identifiable. This enables the dynamic growth of the sampling distribution to be clearly seen, and also allows the students to determine how often values as or more extreme than the one actually observed occurred in the simulation. As a part of the teaching sequence, *Models'n'Data* was used to draw two hundred samples of size 59 from a population where 96% of letters were delivered on time, and a sampling distribution for the sample proportion was formed. Only twice was a sample proportion as low as 88.1% (52/59) observed. This led to the conclusion that, on probabilistic grounds, it is unlikely that the sample has come from the hypothesised population. On this basis, one has legitimate cause to doubt the postal claim.

Argued consistently here, and right throughout the teaching sequence, was that underlying the decision in a hypothesis test was a consideration of the sampling variability of the sample proportion. That is, that the sample proportion can differ from the population proportion hypothesised for either of two reasons:

- the difference is reasonably explained by sampling variability (do not reject the null hypothesis) or;
- there is evidence of a real difference (reject the null hypothesis).

Once the logic of hypothesis testing was introduced using the empirical sampling distribution and a practical, real-world example, it was necessary to formalise the terminology and notation of hypothesis testing for a proportion. At this stage of the instructional process the appropriate language were labels are introduced. These included null hypothesis, alternate hypothesis, test statistic, significance level and P-value. Each concept was defined in the context of the Letters Problem, which was used as a real-world reference when further scenario were introduced.

4.5.5 The theoretical sampling distribution

Whilst the advent of computer sampling packages such as *Models'n'Data* means that students are able to consider hypothesis testing problems using an easily generated empirical distribution, students generally need to learn traditional distribution based methods. Using theoretical probability distributions to carry out inference problems is standard statistical practice. Thus, it was necessary to introduce the theoretical sampling distribution as a step in the instructionally strategy.

Prior to doing this, several more examples were discussed in which the empirical sampling distribution generated by a computer package was used describe the distribution of the test statistic, and to determine empirically the P-value on which the decision is based. After generating and describing in statistical terms several sampling distributions it was evident that, under certain conditions, the sampling distribution had a symmetric, bell shaped distribution. It followed from this that perhaps the normal model was an alternative way of approaching the hypothesis testing problem, allowing the theoretical determination of the P-value.

Models'n'Data was used as a bridge between the theoretical and the empirical sampling distributions as it has the facility to superimpose a normal curve over the histogram of an empirical sampling distribution. This meant that a representation of the theoretical model of sampling distribution was visually displayed together with a representation of the empirical model. Figure 4.5 shows a sampling distribution for the sample proportion generated using *Models'n'Data*, with the same data displayed both as a stemplot and as a histogram with the theoretical sampling distribution shown.

This integration of the empirical and theoretical representations is be an important step in facilitating the cognitive development of the students, in that it makes explicit graphically the link between the theoretical sampling distribution and the empirical sampling distribution. It has been proposed by this researcher that maintaining the link between statistical inference and the sampling process enables students to demonstrate conceptual understanding of statistical inference. If the link between empirical sampling distribution and theoretical sampling distribution is not formed, then a student using traditional statistical methods will have difficulty interpreting their hypothesis testing in terms of the sampling process. Thus, the development of a link between these representations in a student's schema for sampling distribution and hence statistical inference is fundamental.

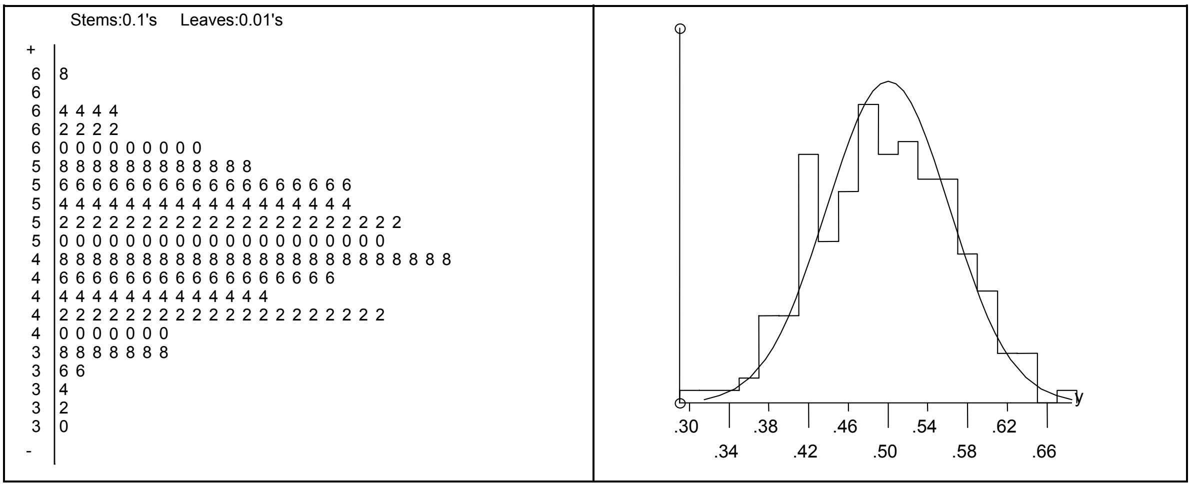


Figure 4.5 Sampling distribution for the sample proportion displayed both as a stemplot and as a histogram with the theoretical sampling distribution superimposed.

Once the normal model for the sampling distribution had been introduced, problems previously discussed were revisited using the normal model, and answers obtained using both the empirical sampling distribution and the theoretical sampling distribution compared and found to be remarkable similar in most instances. However, it was noted that the normal distribution is a very poor model for the sampling distribution in the Letters problem, where the population proportion *p* is 0.96 (too close to 1). It can be clearly seen in Figure 4.6 that the sampling distribution is in fact negatively skew. This reinforced for the students the idea that the normal distribution was only a model for the behavior of the sampling distribution – a model which was suitable in some instance but not in others.

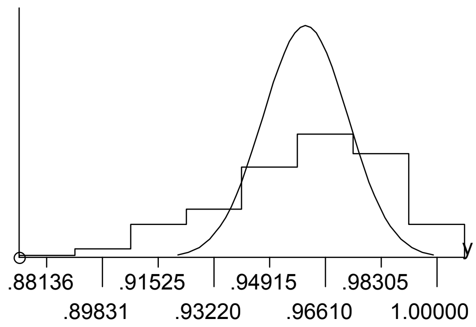


Figure 4.6 Sampling distribution for the sample proportion when the population proportion p=96% displayed a histogram with the theoretical sampling distribution superimposed.

Once the students had completed the discussion of the logic of hypothesis testing was extended to the mean (μ) using *Models'n'Data* to generate an empirical sampling distribution for the sample mean ($\bar{x}$) in a specific example. Then, again, the normal model for the empirical distribution was introduced as an alternative method for determining a P-value.

Models'n'Data was also able to be used to generate empirical sampling distributions for the correlation coefficient, Pearson's r. The students were able to see for themselves that the sampling distribution of Pearson's r was symmetric and bell-shaped, but not normal because it was constrained between the limits of –1 and +1. Students were informed that the theoretical model for the sampling distribution was known, and could be used for the purpose of hypothesis testing via tables or a computer package. For the remainder of the hypothesis tests discussed only the theoretical sampling distribution was used in the determination of the P-value. However, it was continually explained to students that when carrying out an hypothesis test:

- the choice they were making was based on whether or not they could reasonably expect to the sample to come from the population hypothesised;
- that this in turn depended on the sampling variability inherent to the problem;
- that the sampling variability was summarised in the sampling distribution;
- that models for the sampling distribution allowed a numerical value to be placed on this level of reasonableness.

4.5.6 Estimation

The teaching sequence that forms the basis of this study concluded with an introduction to estimation. Estimation was introduced as a parallel strand of statistical inference, with the form of the confidence interval determined by the nature of the sampling distribution, and the exact value of the confidence interval dependent on the value of the sample statistic. Only confidence intervals for the population proportion were discussed.

As well as a theoretical rationale based on the normal distribution explaining the form of the confidence interval 95% confidence intervals were represented schematically as random intervals which captured the population parameter 95% of the time, and missed the population parameter 5% of the time. Using *Models'n'Data* both static and dynamic computer based representations could be displayed which demonstrate this concept. Shown in Figure 4.7 is an example of the dynamic display of confidence intervals for the Letters scenario. The bar chart on the left of the screen shows the result of the results of the last sample of size 59 (the 51st sample taken), and this graphic changes with each sample selected. On the right of the screen is a graphic showing the confidence interval for the population proportion *p* which has been determined based on that sample (shown as a dotted line) as well as all of the confidence intervals determined from the previous fifty samples. Also shown as a vertical line is the value of the population parameter for the population from which these samples were drawn, clearly indicating that, whilst most (about 95%) of the confidence intervals contain the value 0.96, one clearly does not.

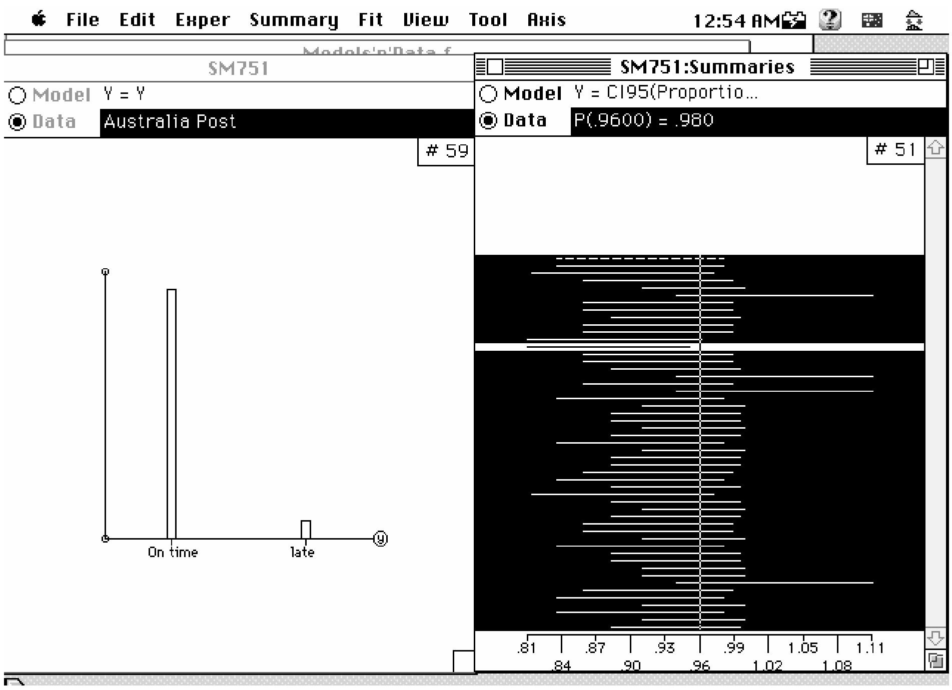


Figure 4.7 *Models'n'Data* screen of a dynamic display illustrating the interpretation of a 95% confidence interval

The equivalence of the results of the hypothesis test and the confidence interval was demonstrated by consideration of several examples, rather than through a mathematical rationale. That is, it was noted that a sample statistic that did not provide sufficient evidence to reject the null hypothesis would generate a confidence interval that contained the hypothesised value for the population parameter and vice-versa.

4.5.7 Conclusion

The instructional intervention implemented in this study was developed to encourage students to modify, restructure and connect schemas for sampling, sampling distribution and statistical inference according to the principles of schema based instruction (see Section 2.6). The instructional strategy was also based on the analysis carried out in Chapter 3 that clarified the key concepts in the knowledge domain of statistical inference and the representations available for these concepts, particularly those involving new dynamic media which are designed to make explicit important links between concepts.

A set of student materials was developed to support this teaching and learning strategy (Lipson & Jones, 1995).

4.6 The setting of the study

The study was conducted at Swinburne University of Technology, Melbourne, Australia. The students chosen for the study were undertaking either the Graduate Diploma in Social Statistics or Health Statistics, a part-time, evening program of study. The students who enrol in the program are generally graduates in some other discipline, such as Business or Nursing, who have determined that knowledge of statistics would be helpful for them in their careers. Some of those who enroll are unemployed and hope that the qualification will lead to employment. Others who enroll have no prior formal studies, but qualify for entry to the course because of their relevant work experience. The study took place during the conduct of a subject called An Introduction to Statistics which was taught over one 13 week semester, and classes were held one evening a week for 3.5 hours. The subject was taken by all students in the program, with no pre-requisite knowledge assumed. There were no assumptions made concerning the mathematical knowledge of the students, which was in many cases minimal. The primary aim of the subject was clearly to engender in these students a qualitative understanding of the concepts of statistical inference, which are assumed as they proceed through the Graduate Diploma. Very little was expected of the students in terms of calculation in subsequent courses as it is assumed that they have access to a computer package such as *SPSS* when working statistically. However, at later stages in the course they are expected to carry out sophisticated statistical analyses using a statistical package, which means that they need to be able to formulate a problem in statistical terms, select the appropriate statistical procedure, and interpret the results accordingly. In addition, because many of the students have employment in areas where statistics has application, there is an expectation from the students themselves that they will be able to transfer the knowledge they have gained from the course into their workplace.

The study took place with a group of 23 mature age students, most at least graduates and well as self-motivated in the study of statistics. As stated by Dubin and Taveggia in their comparative study of college teaching methods (Dubin & Taveggia, 1968):

> At least three factors characterise the teaching-learning situation at college level: (1) voluntarism on the part of the student in choosing the subjects of instruction; (2) a knowledge base possessed by the student for making judgements about the content and quality of instruction received, judgements which, in turn, influence the voluntary choices made; and (3) the complex of culturally derived expectations and behaviours which comprise what we loosely summarise as the motivation to learn. (p.7)

In these terms, the students in the study can be considered as mature, motivated and with prior post-secondary experience. Whilst they are not in any sense a randomly chosen sample of students taking and introductory statistics course, the purpose of the study was not to generalise to other similar groups but to document the changes in the conceptual structures constructed by a group of students, and to investigate the relationship between the nature of these conceptual structures and subsequent performance on a variety of statistics tasks.

In this sense, the group chosen was suitable, for the following important reasons. Firstly, the size of the group (23) is small in comparison to the size of classes often encountered in post-secondary teaching, where lecture groups of 200 are not unusual. This means that the scope of the task, in terms of managing the data collection and analysing the subsequent data was manageable.

More importantly, however, in order to document changes in the conceptual structure it is necessary for the students to be engaged in a meaningful learning collaboration. That is, there needs to be some intention on behalf of the student to proceed beyond the instrumental level of understanding (Skemp, 1978). This motivation to seek a deeper understanding is termed by Salomon & Globerson (1987) as mindfulness, defined by them as:

> ...the volitional, metacognitively guided employment of non-automatic, usually effort demanding processes. (page 625)

Salomon & Globerson go on to suggest that there are three clusters of reasons why students do not perform as well as they could have, given the knowledge that they have acquired. These are cognitive factors, motivational factors and personality variables. In this study, the relationship between the cognitive structure and performance are of primary interest. The student's presence in this group established that he or she was motivated, either for personal or professional reasons. Similarly, previous academic and workplace experiences established that each student had demonstrated a reasonably high level of cognitive ability. Whilst is was not possible to control for personality variables, the fact that two of the three clusters of variables (which were not of particular interest to the research question) could be to some extent controlled by the use of this group of students was a good reason to choose these students for this study.

4.7 Conclusion

This study was formulated to investigate the research hypothesis that:

> Students whose schema for sampling distribution demonstrates links to the sampling process and whose schema for statistical inference includes links to the sampling distribution, will show evidence of both conceptual and procedural understanding, whilst those who have not integrated these concepts into their cognitive structure may still demonstrate high levels of procedural understanding, but not of conceptual understanding.

To test this hypotheses a teaching experiment was undertaken, which involved the participation in an instructional intervention developed by a group of students studying introductory statistical inference. Concept maps were used as a means of externalising the students' conceptual structures throughout the study, and a range of tasks was developed to measure the students' levels of procedural and conceptual understanding after instruction. The data gathered during the teaching experiment form the basis for the analyses undertaken in the next chapter.

Chapter 5

The Results of the Study

5.1 Introduction

As described in Section 4.6, the subject used in the conduct of this study is called An Introduction to Statistics. The students enrolled in the subject were exposed to an instructional treatment that was designed to encourage the development of a rich concept image for sampling distribution and to facilitate the formation of links between the sampling distribution and hypothesis testing. At various stages during the study, the students constructed concept maps using sets of terms that were provided by the instructor. These maps were collected and form part of the data for the study.

Several tasks which had been designed as a result of both an analysis of the concepts involved in the knowledge domain, and a cognitive model of understanding, were used to measure both procedural competency and conceptual understanding in statistical inference for these students (see Chapter 4). These measures were then used to address the relationship between the concept image of sampling distribution and its relationship to inference, and the students' subsequent performance.

The data collected in this study has been analysed at two levels. At a group level, the results of the analysis shows general patterns indicating what may be concluded from the study. At the individual student level, case study analysis gives some insight into why the relationships evidenced at a group level might occur. The group level analysis is undertaken in Chapter 5, and the case study analysis is carried out in Chapter 6.

In this chapter the series of maps constructed by each student were compared in order to ascertain whether there were links between concepts that were recognised or not recognised by the majority of students. On the basis of the schemas evidenced by the concept maps, the students were allocated into one of three groups according to certain general features. Group 1 students showed evidence of the development of the concept of sampling distribution and subsequently integrated sampling distribution appropriately into their schema for statistical inference, while Group 2 students showed evidence of the development of the concept of sampling distribution but did not relate this to their schema for statistical inference, and Group 3 students did not at any stage show evidence of the development the concept of sampling distribution.

Student performance was assessed using the various tasks designed to measure of understanding discussed in Chapter 4. As well as investigating the group results on each of these tasks separately, a factor analysis was undertaken verified the conceptual/procedural dichotomy. This analysis enabled the results for the individual tasks to be combined to give each student two numerical scores, one which measured their procedural understanding and the other which measured conceptual understanding.

The analysis in this chapter explores the relationships between the three groups formed on the basis of the form of the students' schemas and the students' scores on the two measures of understanding. This enables the researcher to address the research hypothesis that the degree of integrative reconciliation of the concept of sampling distribution, and the extent of to which links have been formed between the sampling distribution and statistical inference will have some relationship to the students' level of conceptual understanding, but not to their level of procedural understanding.

5.2 Student characteristics

The participants in the study consisted of twenty-three postgraduate students, of whom 19 were female and only 4 were male. The age of the students ranged from 24 to 54 years, with a median age of 34 years. The students in the study were all graduates, but came from diverse discipline areas as shown in Table 5.1

Table 5.1 Students' prior major discipline area

Discipline area	Number of students	Discipline area	Number of students
Accounting	1	Mathematics	1
Anthropology	1	Medical Admin	3
Business	1	Medicine	1
Economics	1	Microbiology	1
Geography	1	Nursing	2
Health Education	1	Physiotherapy	2
Marketing	2	Psychology	5

Many of these students have previously studied statistics as a component of their post-secondary studies. The level at which the students had completed their previous studies ranged from two-year post-secondary TAFE qualifications, to a Masters degree. The highest level attained by each student in previous study is shown in Table 5.2.

Table 5.2 Students' highest level of academic study

Highest level of study	Number of students
TAFE	2
Graduate	17
Post graduate	4

The students also differed in their level of relevant work experience prior to commencing the course with 11 of the 23 students, or 48%, in statistically related employment.

To obtain a measure of student prior knowledge students were asked to complete a questionnaire in which they assessed their own knowledge in statistics. (This questionnaire is given in Appendix 3) The questionnaires, which assessed their knowledge of 27 statistical terms, were scored and students subsequently classified as low, medium or high in terms of prior statistical knowledge. The questionnaires were scored by allocating 0 points to 'never heard of it', 1 point to 'heard of', two points to 'some knowledge' and 3 points to 'understand'. The points for statement were then summed to determine a total score for each student. On the basis of these score the students were then allocated to groups. Students scoring 0-27 were designated as Low prior knowledge, those scoring 28-54 as Medium, and those scoring 55-81 as High. The results for the study group are summarised in Table 5.3. Individual responses ranged from students who claimed to have never heard of most of the terms before, to those who believed that they have an understanding of all or almost all of the terms.

Table 5.3 Students' prior statistical knowledge (self assessed)

Level	Number of students
Low	9
Medium	10
High	4

As can be seen from the table, most students rated their prior knowledge as Low or Medium, with only 4 of the 23 students rating themselves as High.
In summary, the participants in the study were predominantly female post-graduate and mature-age, who, although from a variety of discipline areas, were particularly motivated in their study of statistics because of its potential relevance to their careers.

5.3 Student concept maps

The regular completion of concept maps was an integral part of the students' coursework. However, only those concept maps relevant to the research question form part of the data for this study. There were six of these, completed over a six-week period, as follows:

Map 1 Concerned with the sampling distribution of the sample proportion.
Map 2 Concerned with the sampling distribution of the sample mean.
Map 3 Concerned with the sampling distribution.
Map 4 Concerned with the hypothesis testing.
Map 5 Concerned with the estimation.
Map 6 Concerned with the statistical inference.

The list of terms used as a starting point for each of these maps is given in Appendix 4. Map 2 was set to be completed out of class, as a homework assignment, and as such was only completed by 9 students. After this poor result, all further maps were completed within the class time. However, because of interruptions to their study due to illness or work commitments, not all students completed all other maps.

The purpose of the concept mapping exercises was to document the students' schemas at particular points in time. This would enable the researcher to identify in the maps the propositions formed by relating the terms given and which indicate understanding of particular statistical concepts by a student. Over a period of time, any changes in the nature of this understanding could be documented by analysis of the sequence of maps.

Analysis of the students' maps allowed general trends to be identified concerning relationships between concepts that were generally recognised or generally not recognised, and any overall group features. On the basis of this analysis students could be categorised according to certain general features of their schemas. This was done so that the student schemas could be related to their subsequent level of performance on the measures of procedural and conceptual understanding. At the individual level, the analysis addresses the hypothesis that student conceptual structures are highly individual, showing varying degrees of cognitive reconciliation, and also point to some misconceptions which may be held by individual students.

5.3.1 Group analysis of propositions

In Table 5.4 is given the percentage of students who included particular key propositions (those developed by the content experts) in each of the concept maps.

Table 5.4 Summary of propositions

Key Propositions	Map 1 n=21	Map 2 n=9	Map 3 n=23	Map 4 n=21	Map 5 n=22	Map 6 n=23
Populations give rise to samples.	95	78	87	71	77	91
Population distributions are described by parameters.	90	89	91	62	9	26
Parameters are constant.	19	44	91	10	9	0
Sample distributions are described by statistics.	71	67	83	71	73	74
Statistics are variable.	33	44	61	48	0	13
Sample statistics form a distribution known as a sampling distribution.	67	100	43	43	27	48
The sampling distribution of the sample statistic can be modelled by a known probability distribution.	43	33	9	5		
The sampling distribution of the sample statistic is characterised by shape, centre, spread.	52	56	48			
The spread of the sampling distribution is related to the sample size.	10	67	35			
The sampling distribution is centred at the population parameter.	5	22	9			
Hypothesis testing is about populations				43		39
Hypotheses are about parameters				43		13
The test statistic is formed from the sample statistic				57		
The exact sampling distribution depends on the null hypothesis				5		
The test statistic and its sampling distribution together help to determine the P-value.				19		30
The P-value depends on the alternate hypothesis				0		
A decision is based on comparing the P-value to the significance level				48		70
The decision is concerned with the null hypothesis				24		4
Estimation is concerned with population parameters					41	22
Point estimates for a parameter is the sample statistic					64	39
Knowledge of the sampling distribution enables us to calculate an interval estimate.					14	13
This interval estimate is called a confidence interval.					68	26
Statistical inference is concerned with both hypothesis testing and estimation						78
Both aspect of inference are concerned with knowing more about a population parameters						30
Consideration of the confidence interval is an equivalent act, leading to the same conclusion.						26

Key:
Map 1 Sampling distribution for the sample proportion
Map 2 Sampling distribution for the sample mean
Map 3 Sampling distribution for the sample proportion
Map 4 Hypothesis testing
Map 5 Estimation
Map 6 Inference

From Table 5.4 it can be seen that there were some key propositions concerning samples and populations which were included by most students. For example, consider Map 3, the map for the general sampling distribution, which required students to synthesis the general features of sampling distribution. Most students understood that there was a relationship between populations and samples (87%), that parameter is the term used for a measure which describe population distributions (91%), and statistic is the term used for a measure which describe sample distributions (91%). As well, 61%

of students went further and correctly identified samples statistics as variable, a fundamental key concept when building a schema for sampling distribution.

For some students knowing that sample statistics were variables led to propositions concerning distributions and their features. However, it was here that the students tended to divide into three groups. These were:

- Those who correctly linked the sampling distribution as the distribution of the sample statistic (43%). These students indicated in their maps that the sample statistic was determined from the sample, and that the variability of the sample statistic could be described by the sampling distribution. (See Figure 6.7 for example)
- Those who incorrectly designated the distribution of the sample (the sample distribution) as the sampling distribution (22%). For these students the sampling distribution described the distribution of the sample, and sample statistics were calculated to summarise features of the sampling distribution. (See Figure 6.23)
- Those who did not clearly indicate whether the sampling distribution was describing the distribution of the sample or the sample statistic (35%). This points to possible confusion on their part over the basis of the sampling distribution. (See Figure 6.30)

This confusion between sample distribution and sampling distribution, which occurred early in the development of the concept of sampling distribution for many students, proved to be startlingly common. Whilst earlier research has identified student misconceptions with sampling (see Chapter 2), this confusion as a source of the problem has not previously been specifically stated, and thus provides a new (and perhaps unexpected) insight into student thinking.

What are the potential consequences of confusing the sampling distribution for the sample distribution? Consideration of the other propositions related to sampling distribution shown in the expert maps indicates that confusion between the sample distribution and the sampling distribution will lead to difficulties in the construction of a linked concept image for sampling distribution. This is because, if the students do not actually recognise that it is the distribution of the relevant sample statistic that is being discussed, then a lot of the subsequent discussion of the properties of sampling distribution will be meaningless. For example, the variability of the sample distribution does not reduce with increasing sample size, whereas the variability of the sampling distribution does. This misconception was not one which the researcher anticipated and has repercussions for this approach to teaching about the sampling distribution which will be developed more fully in Chapter 7.

The key propositions identified in the concept maps are summarised in Table 5.4. Since the students prepared these maps over a six-week period, they evidence changes in the students' conceptual structures over time. Whilst the percentage of students including some propositions remained reasonably constant, others showed large increases or decreases over time. For example, the percentage of students recognising that the sampling distribution is characterised by shape centre and spread was fairly constant at around 50% on maps 1, 2 and 3. However, the percentage of students knowing that the spread of the sampling distribution is related to the sample size was only 10% on map 1, rose to a high of 67% on map 2 and then dropped back to 35% on map 3. This observation supports the contention of Vinner (1983) that the student concept image is dynamic, and does not necessarily retain all the desired features over time unless attention is paid to these features. For example, a necessary concept for the interpretation of a P-value as obtained from a standard hypothesis test such a *t*-test is that the sampling distribution can be modelled by a known theoretical probability distribution. The students in this study participated in an instructional sequence which was designed to establish and reinforce this concept, and thus from an educator's viewpoint it can be said that attentions was continued to be paid to important features of the student schema. However, it can be clearly seen that whilst 43% of students included the appropriate proposition in their maps after the computer based session in which this relationship was recognised, only 9% included this key concept in the maps constructed later in the semester.

As described in Section 4.5, during this instructional strategy emphasis was given to the nature of the sampling distribution. From Table 5.4 it may be seen from the Figures for Map 3 that, whilst 48% of the students noted that the sampling distribution was characterised by shape, centre and spread, only 35% related the spread of the sampling distribution to sample size and a very small 9% noted that the sampling distribution was centred at the value of the population parameter. Identification of the centre and spread of the sampling distribution, and the relationship of these to both the population and the sample size are fundamental for both the application of the theoretical distribution model and the interpretation of the results of inference. However, these results suggest that for many students these ideas were not understood.

The theoretical analysis of the content domain carried out in this study suggested that formation of conceptual links between the empirical sampling distribution and the determination of P-values and confidence intervals was necessary to facilitate conceptual understanding in statistical inference. The analysis of the concept maps in Table 5.4 shows that only 30% of the students explicitly linked the sampling distribution to the determination of the P-value, whilst a very small 13% of the students explicitly linked the sampling distribution to the determination of the confidence interval. This lack of connections suggests due to inadequate or incomplete schemas, many students may later exhibit a lack of conceptual understanding in these areas.

Finally, it may be seen from Table 5.4 that, while 78% of the students knew that statistical inference was concerned with both hypothesis testing and estimation, only 26% of the students were able to directly relate these threads and recognise their possible equivalence.

5.3.2 Group analysis of schemas

Whilst the presence or absence of certain key propositions certainly gives valuable information concerning the students concept image, it tells us little about the underlying structure of the knowledge. To investigate this, we need to look more closely at the concept maps, with particular focus on the links between various propositions. In particular, it is important in this study to ascertain whether or not the student formed an appropriate link between the sampling distribution and hypothesis testing, estimation or both.

A theoretical analysis of the key propositions to be identified in the students' final maps concerned with statistical inference enabled several stages of increasing development of schema and relationships between schema to be identified. These are described in Table 5.5. Stage 1 indicates poorly formed schemas, and increasing levels are associated with the formation of schema becoming more and more similar to those of the content experts, and could be said to be indicative of increased understanding.

Table 5.5 Stages of schema development identified in the concept maps for Statistical Inference

Stage	Description
1	No clear structure, confused
2	Sampling distribution absent. Two strands, one hypothesis testing and one estimation. No link between conclusions reached.
3	Sampling distribution absent. Two strands, one hypothesis testing and one estimation (or equivalent). Link between conclusions reached indicating recognition of equivalence.
4	Sampling distribution not correctly specified, but present. Two strands, one hypothesis testing and one estimation. No link between conclusions reached.
5	Sampling distribution not correctly specified, but present. Two strands, one hypothesis testing and one estimation. Link between conclusions reached indicating recognition of equivalence.
6	Sampling distribution correctly summarising sample statistic, but not linked to either hypothesis testing or estimation
7	Sampling distribution correctly specified and linked only to hypothesis testing or estimation
8	Sampling distribution correctly specified and linked to both estimation and hypothesis testing
9	Sampling distribution correctly specified and linked to both estimation and hypothesis testing, and these linked to each other

Analysing the concept maps prepared by the students in this study with reference to these stages, the distribution shown in Table 5.6 was obtained.

Table 5.6 Number of students exhibiting each stage of schema development

Stage	Number of students
1	1
2	0
3	5
4	5
5	4
6	1
7	4
8	3
9	0

The results given in Table 5.6 show that for some students (Stages 1, 2 and 3), the sampling distribution has no place in statistical inference, and is omitted entirely from their concept map. That is, there is no link between the sampling distribution and statistical inference for these students. Other students recognise that there is a role for sampling students, but many are unclear as to the precise nature of that role.

There is certainly strong evidence from the series of maps that the students have created individual conceptual structures, with no two maps even slightly similar in content and/or structure. This variation in schema will be further discussed in Chapter 6. However, as can be seen from Tables 5.5 and 5.6, some general patterns are able to be distinguished. A qualitative analysis of all of the concept maps prepared by each student, including consideration of the concept maps prepared for sampling distribution as well as the stage of development evidenced by the student in Table 5.6 in their map for statistical inference was carried out. This analysis enabled each student to be allocated into one of three broad categories. These were:

Group	Schema Levels	Description
1	7-9	Students whose schemas showed evidence of the development of the concept of sampling distribution and subsequently integrated sampling distribution appropriately into their schema for statistical inference (7 students).
2	4-6	Students whose schema showed evidence of the development of the concept of sampling distribution but did not relate this to their schema for statistical inference (10 students).
3	1-3	Students who did not show evidence of the development the concept of sampling distribution (6 students).

Chapter 6 gives examples of the maps prepared by individual students in the case study analysis that clearly illustrates the differences between the groups.

One of the hypotheses addressed in this study concerns the relationship between the form of the student's schemas and procedural and conceptual understanding. This relationship, between the groups and students' subsequent performance on a variety of statistical measures will be considered later in the chapter. In the next section the level of student performance on the tasks developed to measure procedural and conceptual understanding will be discussed.

5.4 Measuring understanding

In this section the analysis of the results of each of the assessment tasks aimed at assessing a range of aspects of statistical understanding, and described in detail in Section 4.3, is reported. After the individual task results are discussed, a factor analysis is carried out in order to construct separate composite measures of conceptual and procedural understanding for each student.

5.4.1 Task Analysis

Several tasks were developed for assessing the students understanding following the instructional intervention. A range of tasks, which could be considered to together provide measures of both procedural and conceptual understanding in the students, was developed. The actual tasks, their source and rationale may be found in Chapter 4. What follows in this section, is analysis of the student performance on each of the individual tasks at a group level.

Task: Sampling

A survey is conducted with a random sample of 282 university students, in order to find out how far they travel to university each day. One student questions the validity of the study, noting that there are 4000 students at the university, not just 282. Read each of the statements listed below carefully, and select the ONE response that sounds the most reasonable to you.

A Agree, 282 is too small a percentage of the 4000 (7%) to allow us to draw conclusions.

B Agree, you should have a sample that is at least 50% of the population in order to make inferences.

C Agree, they should get all the students to participate in the survey.

D Disagree, 282 is a large enough number to use for these purposes if the sample was a random sample of students.

E Disagree, if the sample is random, the size doesn't matter.

The purpose of this task was to investigate student's beliefs concerning the sampling process as a method of obtaining information about a population. A bar chart of the responses to this question for the student group is given in Figure 5.1.

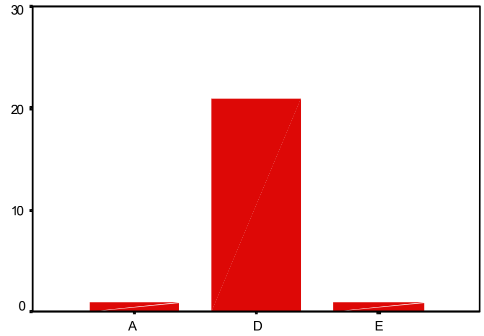


Figure 5.1 Distribution of student responses to the Sampling task

From the bar chart it can be seen that only two students gave an incorrect answer to this question, indicating that the question of sampling as a valid method of obtaining information about a population prompted a correct response for most students when asked directly.

Task: Weather

The Weather Bureau wanted to determine the accuracy of their weather forecasts. They searched their records for those days when the forecaster had reported a 70% chance of rain. They compared these forecasts to records of whether or not it actually rained on those particular days. On what percentage of these days would you expect to see rain had occurred if their forecast is very accurate?

A 95% – 100% of those days
B 85% – 94% of those days
C 75% – 84% of those days
D 65% – 74% of those days
E 55% – 64% of those days

This task was designed to explore student interpretation of probability. A bar chart of the students' answers to this question is shown in Figure 5.2.

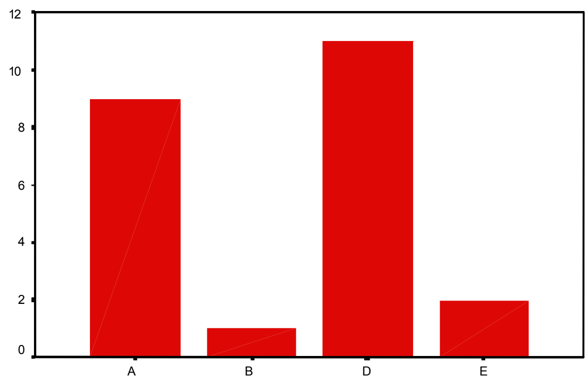


Figure 5.2 Distribution of student responses to the Weather task

The responses to this question show more variety, with 9 of the 23 students, or 39.1%, selecting response A. According to Konald (1995), the reasoning used by these students is based on an outcome interpretation of probability, discussed previously in Section 4.3.1. (These results are also consistent with those obtained by Konald who found 36% of the students in his sample chose response A). It is possible that for these students comparison and interpretation of small probabilities could be problematic.

Task: Hospital

Half of all newborn babies are girls and half are boys. Hospital A records an average of 50 births a day. Hospital B records an average of 10 births a day. On a particular day, which hospital is more likely to record 80% or more female births?
A Hospital A (with 50 births a day)
B Hospital B (with 10 births a day)
C The two hospitals are equally likely to record such an event.

This task is concerned with the relationship between sampling variability and sample size. A bar chart of the students' responses to this task is shown in Figure 5.3.

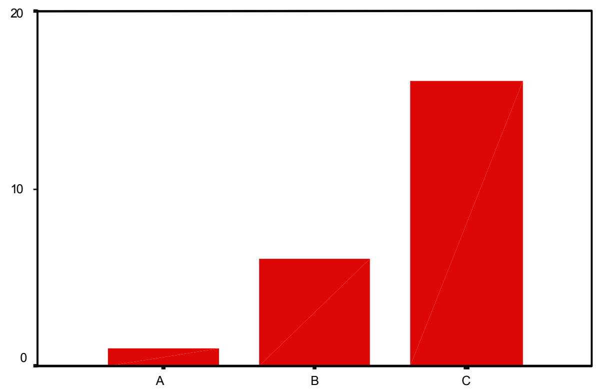


Figure 5.3 Distribution of student responses to the Hospital task

As can be seen from the bar chart, most students (16 out of 23, or 69.6%) selected response C, slightly higher that the 56% observed by Tversky & Kahneman (1982). This could indicate that the student has not made explicit in their conceptual structure for sampling distribution the link between the variability of the sampling distribution and the size of the sample. Alternatively, this result might mean that, whilst the student is aware of the link between sampling variability and sample size in a classroom context, the scenario used to illustrate sampling here does not evoke that aspect of the student's conceptual structure for sampling distribution.

Task: *t*-test

According to a Census held in 1956, the mean number of residents per household in an inner suburb, Richthorn was 3.6. In 1995, a student randomly sampled 11 households from the suburb and recorded the number of residents in each with the following results

2	2	5	1	1	3	4	2	4	3	1

Can the student conclude that the mean number of residents per household in Richthorn has decreased since the 1956 Census.

This task required the students to carry out a standard *t*-test. The task was scored out of a possible 10, and a detailed marking scheme may be found in Appendix 5A. A frequency distribution of the students' scores on this section of the question is given in the Table 5.7:

Table 5.7 Distribution of student score for the *t*-test task

Score	Frequency	Percentage
0	1	4.3
3	1	4.3
8	4	17.4
9	3	13.0
10	14	60.9

From the table it may be seen that most students scored extremely well on this task, with 21 out of the 23 students receiving a mark of 8 or more out of a maximum of 10. The mark of 0 was obtained by a student who omitted the question, whilst the student who obtained a mark of 3 exhibited several problems including incorrect choice of test, incorrect conclusion based on the P-value obtained, and calculation errors. The overall results however indicate that for most students, even those with little mathematical aptitude, use of formulae, calculators and tables are skills well within their ability.

Task: Confidence Interval

Using a computer package, the student finds the 95% confidence interval for the mean number of resident to be (1.626, 3.465). Is this confidence interval consistent with what you found in part (a)* Explain.

*In the terminology used here, part (a) refers to *t*-test task.

The question was designed to probe the link between the hypothesis testing and the confidence interval. These results were indicative of a poor general recognition of the relationship between these two facets of statistical inference, with only 8 students (35%) fully appreciating that the results were consistent because the confidence interval given did not include the hypothesised value of μ, the sample mean. The student who obtained 1 mark indicated that there was a contradiction, but then gave the correct reasoning. Of the students 14 who scored zero, three students either misinterpreted or omitted the question, whilst an alarming 11 students indicated that there was no inconsistency, as the sample mean of 2.545 lay within the confidence interval given. This answer would appear to indicate a general lack of conceptual understanding concerning the confidence interval and its interpretation.

The scores allocated to the group are summarised in Table 5.8.

Table 5.8 Student scores for Confidence Interval Task

Score	Frequency	Percentage
0	14	60.9
1	1	4.3
2	8	34.8

Task: Sample Size

Assuming that the number of residents in a household is normally distributed with a standard deviation of $\sigma = 1.4$, determine the minimum sample size needed to be 95% confident that the population mean μ will be within 0.5 units of the sample mean.

To successfully complete this task required the students to substitute into a given formula and use their calculator to determine the answer. The maximum score available was 3, and the marking scheme for the task is given in Appendix 5B. The results for the group are given in Table 5.9.

Table 5.9 Student scores for the Sample Size task

Score	Frequency	Percentage
0	1	4.3
1	0	0
2	6	26.1
3	16	69.6

Once again, the students performed extremely well on this procedural task. The student who scored zero omitted this question, whilst all the other who lost marks did so due to errors in calculation.

Task: Two group *t*-test

The following data was generated in a study of the effectiveness of in-service training of nurses. In this study, 15 nurses were given a test of their knowledge of cancer and cancer nursing prior to attending a one-day workshop on the topic. They were then retested on the same topic at the conclusion of the workshop. The results are displayed in the table below.

Nurse	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Pre	29	20	24	32	33	19	17	32	16	28	35	28	18	45	19
Post	35	41	33	41	39	20	29	42	36	37	36	33	35	42	19

The course organiser wanted to know whether there had been an increase in the nurse's knowledge of cancer and cancer nursing after attending the course. Two assistants, Roger and Annette, were given the task of analysing the data using *Minitab*.

Roger carried out a two-sample *t*-test with the following results:

```
TWOSAMPLE T FOR posttest VS pretest
           N      MEAN     STDEV   SE MEAN
posttest  15     34.53      7.14       1.8
pretest   15     26.33      8.29       2.1

TTEST MU posttest = MU pretest (VS GT): T= 2.90  P=0.0036  DF=  27
```

Annette converted the two columns of data into a single column of difference scores and carried out a one-sample *t*-test with the following results:

```
TEST OF MU = 0.00 VS MU G.T. 0.00

             N      MEAN     STDEV   SE MEAN     T    P VALUE
diffs       15      8.20      7.15      1.85  4.44     0.0003
```

(a) Which analysis is the more appropriate and why?

(b)	Using the information generated in the analysis that you think is most appropriate, test the hypothesis that there has been an increase in the test scores for the nurses who attended the workshop. Quote the values of all relevant statistics.

This task required the students to select either a paired or an independent groups *t*-test, and then use the appropriate computer output, which was given, to carry out the test. All students chose the correct test, although some did not obtain full marks for the question as their reason was either omitted or not clearly expressed. The scores for the conduct of the two group *t*-test (marking scheme given in Appendix 5C) were also very good, as can be seen from Table 5.10, with 78.2% of students scoring 7 or more marks from a maximum of 8.

Table 5.10 Student scores for Two group *t*-test Task

Score	Frequency	Percentage
5	2	8.7
6	3	13.0
7	5	21.7
8	13	56.5

Overall, it may be seen that students performed very well in this question. Marks were lost by students for incorrectly specified hypotheses, failure to mention the level of significance at which they were working, and clumsy wording.

Task: Modelling

A sample of 100 primary school children were asked which type of protection they preferred to use to protect their faces from the sun, a hat or sunscreen. Of the 100 children, 61 preferred to use a hat, and 39 preferred to use sunscreen.

(a) Use the computer generated sampling distribution given to test if there is a significant difference in the proportion of students who show preference for one type of protection over the other.

```
Stems:0.1's    Leaves:0.01's
+ |
7 |
7 |
7 |
6 |
6 | 6
6 |
6 | 2 2 3
6 | 0 1
5 | 8 8 8 9 9 9 9
5 | 6 6 6 6 6 6 6 6 6 6 6 6 7 7 7 7 7
5 | 4 4 4 4 4 4 4 4 4 4 5 5 5 5 5 5
5 | 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3 3 3 3 3 3 3 3 3 3
5 | 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
4 | 8 8 8 8 8 8 8 8 8 8 8 8 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9
4 | 6 6 6 6 6 6 6 6 6 6 6 6 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7 7
4 | 4 4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
4 | 2 2 2 2 3 3 3 3 3 3
4 | 1 1
3 | 8
3 |
3 | 4
3 |
3 |
2 |
2 |
- |
```

Stemplot showing 200 values of the sample proportion obtained from drawing samples of size 100 from a population with proportion p = 0.5.

(b) Use the normal model for the sampling distribution to test if there is a significant difference in the proportion of students who show preference for one type of protection over the other.

This task consisted of two parts, with the students required to carry out an hypothesis test for a proportion using both the empirical sampling distribution in part (a), and the normal model for the sampling distribution in part (b). Each part of the tasks was scored out of maximum of 7, and the marking scheme used for the allocation of scores is given in Appendix 5D.

An analysis of the marks obtained by the students revealed that the mean score for part (a), which required the used of the empirical sampling distribution was 5.0 marks whilst the mean mark for part (b) which required the use of the normal model was 5.7 marks. A paired t-test showed that the students performed significantly better in part (b) than in part (a), ($t(22) = 2.296$, $p = 0.032$). This difference is not large, but indicates that the students have on the whole performed better in the task that measure procedural understanding than that measuring conceptual understanding. However, most students scored quite well on both parts of the task, as can be seen from Table 5.11, with 17 students (74%) scoring 5 or more for part (a), and 20 students (87%) scoring 5 or more for part (b).

Table 5.11 Distribution of student scores for the Modelling Task

Score	Frequency (a)	Percentage (a)	Frequency(b)	Percentage (b)
0	2	8.7	2	8.7
1	0	0	0	0
2	1	4.3	0	0
3	1	4.3	1	4.3
4	2	8.7	0	0
5	8	34.8	4	17.4
6	2	8.7	4	17.4
7	7	30.4	12	52.2

The correlation between the scores was also quite high (r = 0.724) indicating that, in general, those students who attained the higher scores for part (a) also attained the higher scores for part (b). In later analysis, the two parts of the question are considered separately.

Task: Correlation

The following data was collected in 1970 in an investigation to determine if the amount spent by a school district on public education is dependent on the per capita income of the school district.

Income ($)	4198	4008	6197	5343	5928	3764	3244	4612	4918	2194
Expenditure ($)	871	850	1210	1188	1547	834	623	1052	1256	476

(a) Assuming that the sample of states has been drawn at random from a very much larger population of states, can we conclude from this data that there is a relationship between expenditure on education per student and per capita income for school district in general? Justify your response by carrying out a hypothesis test on *r*.

(b) The value of r may or may not be statistically significant, but is it of practical significance? Justify you answer by calculating the value of an appropriate statistic.

This task consisted of two parts and each was scored using the marking scheme given in Appendix 5E. In the first part of the task the students were required to carry out a standard hypothesis test for correlation, and the scores for part (a) are shown in Table 5.12.

Table 5.12 Distribution of student scores for the Correlation (a) Task

Score	Frequency	Percentage
0	2	8.7
3	1	4.3
4	3	13.0
5	17	73.9

Once again, the majority of students demonstrated that they were extremely competent when carrying out a procedural task, with 87% of the group scoring 4 or more out of a maximum 5. The two students who scored zero omitted this question. The student who scored 3 omitted to use the tables to find the P-value, and used the value of Pearson's r instead. Such a mistake is not rare in the researcher's experience, and seems to indicate that the student has not real understanding of the logic of hypothesis testing, but is following through an algorithm and has on this occasion omitted one step.

Part (b) of the correlation task required the students to interpret the correlation coefficient in terms of practical significance. Results for part (b) are given in Table 5.13.

Table 5.13 Distribution of student scores for the Correlation (b) Task

Score	Frequency	Percentage
0	6	26.1
1	3	13.0
2	4	17.4
3	10	43.5

Of the six students who scored zero for this task, 3 omitted the question, whilst the other three gave explanations which were incorrect and irrelevant. All of the remaining students calculated or attempted to calculate the coefficient of determination. Those scoring 1 mark either did not more that this or gave an incorrect interpretation. Those scoring two marks generally gave a clumsy but essentially correct interpretation of r^2, whilst those scoring three marks correctly calculated r^2, correctly interpreted this value, and commented appropriately on the practical significance of the result. The fact that

17 students (74%) calculated r^2 indicates that most students showed some sense of the difference between significance and importance.

Task: Chi square

A study was conducted to determine if there is a relationship between socioeconomic status and attitudes to an urban-renewal program. The results are as shown.

	Urban renewal project		
Socioeconomic status	Disapprove	Approve	Total
Middle	90	60	150
Lower	200	100	300
Total	290	160	450

Does the data support the contention that there is a relationship between attitude tot the urban renewal project and socioeconomic status? Carry out an appropriate hypothesis test.

This task required the students to carry out a standard chi-square test, which was marked according to a marking scheme given in Appendix 5F. Student scores are shown in Table 5.14.

Table 5.14 Distribution of student scores for the Chi squared Task

Score	Frequency	Percentage
3	1	4.3
4	1	4.3
7	3	13.0
8	2	8.7
9	16	69.6

As may be seen from the table, 21 of the 23 students scored 7 or more out of 9, again confirming the competency of most of the student group when carrying out tasks of a procedural nature.

Task: Unknown test

A researcher wishes to know whether blood pressures became more variable after a particular treatment. To determine this she carries out an F-test (which you have not been taught) which can be used to test for the equality of variance in two independent samples.

For the treatment group (n =10) the sample variance was found to be = 76.44 while for the placebo group (n = 11) the variance was = 34.82 giving an F statistics of 2.20, and a P-value of 0.26.

Write down appropriate null and alternative hypotheses and use the P-value to draw a conclusion about the variability of blood pressure in the two groups.

In this task the student was asked to write down the hypotheses, the P-value and give a conclusion for a test that they had not seen before. The responses to this task were evaluated using criteria given in Appendix 5G, and the scores given are shown in Table 5.15. The maximum available score was 4 marks.

Table 5.15 Distribution of student scores for Unknown Test Task

Score	Frequency	Percentage
1	3	13.0
2	15	65.2
3	0	0
4	5	21.7

From Table 5.15 it can be seen that only 5 students were able to present a correct solution to this problem, and of these 5, only two actually wrote their hypotheses in symbols, whilst the other three wrote them in words. This would perhaps indicate some uncertainty about which symbol to use, as almost all hypotheses in this teaching sequence are written in symbols. Of the rest, 11 wrote their hypotheses incorrectly in terms of s, and another four wrote their hypotheses in terms of μ. One student wrote hypotheses in terms of F, another ρ, and another felt unable to complete the hypotheses in this question. Regardless of the hypotheses, all but one student used the P-value correctly to make a decision regarding the appropriateness of the null hypothesis in this context.

Task: Explanation

As a part of your research, you are investigating the relationship between the intelligence of a child and the intelligence of their mother. To this end, you administer intelligence tests to the mother and eldest child of a randomly selected sample of 30 families. A scatterplot of the data obtained indicated the presence of a moderate linear relationship between the intelligence test score of the mothers and their children and the value of Pearson's r was found to be r = 0.5135. You carry out a hypothesis test as shown below and conclude that the data you have supports your long held contention that there is a relationship between the intelligence of children and the intelligence of their mother in the general population.

A friend, who is very interested in your research, but who understands little about statistics, asks you to explain to him how you have come to this conclusion. They are happy that there is a relationship between the mother's and the eldest child's intelligence in the sample, but can't understand how you can generalise your result to include mothers and eldest children in general. You explain that this is the purpose of the hypothesis test you performed.

In the space provided in the table give a brief explanation of each step in the hypothesis test which you have carried out, so that your statistically illiterate friend is able to understand what you have done and how you were able to draw your conclusion. You can assume that your sample is properly representative of the general population.

Steps in your hypothesis test	Explanation
Hypotheses: H_0: $\rho = 0$ H_1: $\rho \neq 0$ non-directional test	
Significance level: $\alpha = 0.05$	
Test statistic r = 0.5135 for n = 30 pairs of data values	
P-value P-value $= 2 \times P(r > 0.5)$ $= 2 \times 0.0025$ $= 0.005$ or 0.5%	
Decision & conclusion As $p < 0.05$, reject H_0 and conclude that there is a relationship between the intelligence of children and their mothers in the general population.	

This task asked students to explain each of the steps in a given hypothesis test in layman's language. Details of the criteria and marking scheme are given in Appendix 5H.

Marks were not awarded where the student merely re-phrased or described in words the step in the hypothesis test. The task was scored out of a maximum of 10, and the results achieved by the group are summarised in the histogram in Figure 5.4.

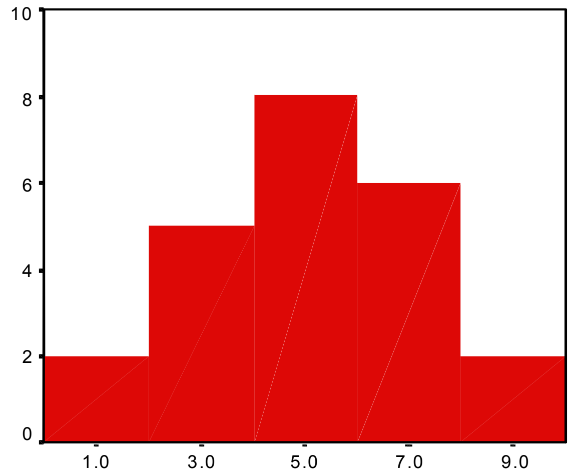


Figure 5.4 Distribution of student scores for the Explanation Task

From the histogram it may be seen that the distribution of scores is symmetric. The mean score is 4.7 and the standard deviation 2.2.

In order to gain a clearer understanding of the student performance, the student answers will be further discussed under the major subheadings given in the question

Hypotheses

The mean mark for this part was 1.65, with most students (16 out of 23) scoring 1 or 2 marks out of a possible 3 marks. No student was awarded full marks. In general, students were quite capable of interpreting the hypotheses given in terms of the scenario at hand, and were able to describe verbally what was meant by a non-directional test. However, they in general omitted to discuss the purpose of the hypothesis test, and the role the hypotheses played in the steps following.

Significance level

There were two marks available here, and the group scored an average of 0.98 marks. Seven students scored less than 1, nine students scored 1, and seven students scored more than 1. Many students wrote a procedural definition of the level of significance, such as "we will reject the null hypothesis of the P-value is less than this value", while others were capable of interpreting this value quite correctly.

Test statistic

Marked out of 1, the group average was 0.41, with only four students being awarded the whole mark. Eleven students were awarded half a mark for an informative discussion of Pearson's r and how it can be interpreted, but the key issue here was for the students to recognise and mention that this is a measure of the strength of the relationship in this sample, and it will be used as evidence of a relationship in the population.

P-value

This section was marked out of 2, with the group averaging 0.67 marks. Of the 23 students, 14 were awarded zero here, and seven were awarded full marks. Most students awarded zero offered an operational definition, such as "to get the P-value you look up the tables and then multiply the answer by two because it is a two tail test".

Decision and conclusion

The mean mark for the group for this section was 0.96, out of a possible 2 marks. Seven students scores zero, whilst 16 scored one or more mark. Many of these students mentioned that the hypothesis was rejected because sampling variability was not the preferred explanation for the difference between the sample statistic and the hypothesised value (because of the smallness of the P-value), an argument which had been used repeatedly in the lectures and student materials.

In summary, it seems that most students knew what they were trying to achieve by carrying out the hypothesis test, and were correctly able to interpret the conclusion, with 15 out of the 23 students scoring 2.5 or more out of 5 for these two section combined, and a group average of 2.6 marks. However, the general understanding of level of significance and P-value was lower, with an average of 1.6 marks out of 4 awarded, and only 10 of the 23 students scored 2 or more marks.

Task: Radio

A radio station claims to its advertisers that 20% of 18–25 year old listen to this station between 6.00 pm and mid-night on weeknights. A market research company carries out independent research on behalf of an advertiser and finds that only 15% of their sample of 18–25 year olds listen to the radio station in this time period. The advertiser concludes that the radio station is misleading them. What do you think? Try to include all the relevant reasons for your answer.

Based on the possible explanations that may be given, a marking scheme was devised according to a set of criteria, and is given in Appendix 5I.

The general themes of the students' answers may be summarised as shown in Table 5.16.

Table 5.16 Summary of students' answers to the Radio task

Theme of answer	Frequency
Sample selection problem only	6
Hypothesis test only	8
Both	9

The average mark achieved in this question was 3.1. Of the 23 students in the group, a total of 17 or 73.9% recognised that sampling variability was a possible explanation for the difference between the sample statistic and the population parameter and that a statistical procedure existed which would enable them to decide if the this was the preferred explanation. This would indicate that for these students, the hypothesis testing procedure was well enough understood to be recognised in a real world situation.

In this section the students' performances on each task has been analysed and discussed. However, it is not the outcome of individual tasks which is main purpose of the analysis. These tasks were designed so that between them they would provide composite measures of the student's procedural and conceptual understanding. The next step in the analysis process is thus the construction of the variables that will be used to measure the two facets of understanding identified in this study.

5.4.2 Measures of procedural and conceptual understanding

As discussed in Section 4.3.1, the tasks reported in the previous section were designed to measure a variety of facets of student understanding of statistical inference. The intention the researcher was to combine the variables represented by the individual tasks into two composite variables, enabling each student to be allocated two scores, one measuring procedural understanding and one measuring conceptual understanding.

To achieve these composite variables, the scores given to each of the individual tasks needed to be aggregated in a meaningful way. Factor analysis (Klein, 1994) is a statistical procedure used to reduce a large number of inter-related variables to a relatively small number of underlying factors which are conceptually meaningful.

An exploratory factor analysis was carried using all of the tasks, with the intention of allocating each task to one of two factors. Using the statistical package *SPSS*, and the statistical procedure Factor, the data was found to be suitable for factor analysis. The

Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was found to equal 0.732. Small values of the KMO suggest that factor analysis is not appropriate. Bartlett's Test of Sphericity test the null hypothesis that there are no correlations amongst the variables, and consequently we would hope to reject this hypothesis with a very small P-value. In this instance a P-value less than 0.0005 was returned. To allow for the two factors to be correlated, an oblimin rotation was chosen. The resultant factors attained through this procedure together explained 55.5% of the total available variance. Table 5.17 gives the factor loadings for this analysis, with the tasks sorted by factor, and loadings less that 0.1 blanked.

Table 5.17 Rotated factor loadings sorted and blanked when all are Tasks entered

	Factor 1	Factor 2
Chi square	.956	
Modelling (b)	.911	-.144
Correlation (a)	.900	.114
Modelling (a)	.875	
t-test	.845	.122
Sample size	.753	.249
Two group *t*-test	.629	.145
Sampling	.549	-.196
Weather	.510	-.230
Correlation (b)	.458	.307
Unknown	.453	
Explanation	.127	.774
Radio	.114	.704
Hospital	-.207	.704
Confidence Interval	.280	.348

At this stage of the analysis it can be seen that most of the tasks which have loaded onto Factor 1 could be considered measures of procedural understanding, whilst most of those loading onto Factor 2 were tasks designed to measure conceptual understanding. However, further examination of these factors revealed some unexpected results. Whilst it had been anticipated that the modelling question would give some insight into the conceptual understanding, both parts of this question were clearly highly correlated with other questions measuring procedural understanding. On reflection this result was understandable, as questions similar to the Modelling task had been regularly discussed during the instructional intervention, and given as practice exercises out of class. What had originally been a task designed to measure conceptual understanding had been altered to a task measuring procedural understanding by the considerable amount of previous practice experienced by some of the student with similar tasks.

The Weather, Sampling and Unknown tasks designed to measure conceptual understanding, were also loading onto Factor 1 which seemed to be measuring Procedural understanding, whilst the Confidence Interval task and the Correlation (b) task were not loading highly onto either factor. In order to resolve the tasks more clearly into distinct factors, the factor analysis was repeated with Weather, Sampling, Correlation (b) and Unknown omitted. The rationale for omitting these tasks is as follows:

- The Weather and Correlation (b) tasks were omitted because the concepts being examined through these tasks did not related directly to the inference process, the fundamental objective of the study.
- The Sampling and Unknown tasks were omitted for an entirely different reason. The unseen hypothesis test task was created in order to measure the student's ability to transfer their knowledge of hypothesis testing to a new and novel scenario. Detailed analysis of the students' responses to the Unknown task revealed that for this group of students, the scenario as presented in this question was not always novel. For some students, who had undertaken previous courses in statistics, this task was already known and could thus be classified as procedural understanding. For others, it falls into the domain of conceptual understanding, as they had not seen this before. On this basis, the decision was made to repeat the factor analysis without the Unknown task. Similarly, it was concluded that on the basis of the student's previous statistical experience, the Sampling tasks would be procedural for some students and conceptual for others, and for this reason this task was omitted from the final analysis. While these tasks were designed to measure conceptual understanding, but the data indicated that this was not the case. Since there was no theoretical basis on which to argue that the tasks were measuring procedural understanding, they were omitted as it was unclear to the researcher exactly what they were measuring.

After these tasks had been removed the data was found again to be suitable for factor analysis, with KMO = 0.801, and Bartlett's Test of Sphericity returning a P-value less than 0.0005. The resultant factors attained through the second analysis together explained 67.9% of the total available variance. Table 5.18 gives the factor loadings for this analysis, with the tasks sorted by factor, and loadings less that 0.1 blanked.

Table 5.18 Rotated factor loadings sorted and blanked when the Weather, Correlation (b) Sampling and Unknown tasks are omitted

	Factor 1	Factor 2
Chi square	.981	
Correlation (a)	.944	
Modelling (b)	.936	-.218
t-test	.885	
Modelling (a)	.833	
Sample size	.797	.188
Two group *t*-test	.718	
Explanation		.828
Radio		.776
Hospital	-.111	.555
Confidence Interval	.221	.524

The factor matrix now shows simple structure, and the resultant factors are clearly identifiable as measures of procedural understanding (Factor 1) and conceptual understanding (Factor 2). The correlation between Factor 1 and Factor 2 was 0.254, indicating a weak positive relationship between the two factors, due one might suggest to a general underlying ability factor.

These analysis confirm that, as suggested by the theoretical analysis, the tasks developed concerning assessment of aspects of understanding can be resolved into two variables measuring the underlying constructs of procedural and conceptual understanding.

Using the factor loadings given in Table 5.18, factor scores were calculated to give composite measures of procedural understanding and conceptual understanding for each student.

5.5 The relationship between students conceptual structure and understanding

This thesis addresses an hypothesis concerning the relationship between the content and form of the student's schemas for sampling distribution and statistical inference, and the level of understanding which he or she demonstrates.

Earlier in this chapter (Section 5.3.4) students were assigned to one of three groups on the basis of their schemas for sampling distribution and statistical inference as evidenced by a series of maps which they constructed over the period of the instructional intervention. These groups were as follows:

Group 1 Students whose schema showed evidence of the development of the concept of sampling distribution and subsequently integrated sampling distribution appropriately into their schema for statistical inference.

Group 2 Students whose schema showed evidence of the development of the concept of sampling distribution but did not relate this to their schema for statistical inference.

Group 3 Students who did not at any stage show evidence of the development the concept of sampling distribution.

It was hypothesised that those students whose schema for sampling distribution demonstrates links to the sampling process and whose schema for statistical inference includes links to sampling distribution, will show evidence of both conceptual and procedural understanding, whilst those who have not integrated these concepts into their cognitive structure may still demonstrate high levels of procedural understanding, but will not of conceptual understanding. On the basis of the clear separation of the students into the three groups given on the basis of the evidence of the nature of their conceptual structures, this hypothesis was slightly revised as follows:

> That those students who showed evidence of the development of the concept of sampling distribution and subsequently integrated sampling distribution appropriately into their schema for statistical inference (Group 1) would exhibit higher levels of conceptual understanding that those students who showed evidence of the development of the concept of sampling distribution but did not relate this to their schema for statistical inference (Group 2), who would in turn exhibit higher levels of conceptual understanding than students who did not at any stage show evidence of the development the concept of sampling distribution (Group 3). It was further hypothesised that there would be no difference between these groups in the levels of procedural understanding shown.

In order to test this hypothesis, factor scores were generated for each student on each of the factors, procedural understanding and conceptual understanding, using the coefficients obtained from the factor analysis. The means and standard deviations for each of these factor scores for each group are given in Table 5.19.

Table 5.19 Summary Statistics for the Procedural and Conceptual understanding factor scores for each group

Group		Procedural understanding		Conceptual understanding	
	N	Mean	SD	Mean	SD
Sampling distribution concept developed and linked to inference	7	.384	.205	.838	.968
Sampling distribution concept developed but not linked to inference	10	.004	1.010	-.007	.589
Sampling distribution concept not well developed	6	-.455	1.311	-.966	.741

Note that factor scores are standardised so that each variable has a mean of 0 and a standard deviation of 1. The data in Table 5.19 enables comparison to be made between the three groups separately for the measures of procedural and conceptual understanding[2].

Procedural understanding

It can also be seen from the table that in this sample students in Group 1 scored higher on average in procedural understanding ($\underline{M}$ = .384) than students in Group 2 ($\underline{M}$ = .004), who in turn scored higher than students in Group 3 ($\underline{M}$ = -.455), but the size of these differences in smaller than those observed for the conceptual understanding measure.

A oneway Analysis of Variance showed that the differences in the scores in procedural understanding between the three groups were not statistically significant, $F(2,20) = 1.154$, $p = 0.336$. Since the sample sizes are very small, however, we would be unlikely to find any difference unless these differences were quite large.

Conceptual understanding

From Table 5.19 it can be seen that the average score for students in Group 1 ($\underline{M}$ = .838) on the measure of conceptual understanding was higher than that for Group 2

[2] For equivalent analysis using the raw scores see Appendix 7

(M = -.007), which was in turn higher than that for Group 3 (M = .-.966). In fact, the Group 1 students scored about 1 standard deviation above the mean for the whole group, the Group 2 scores were about average, and the Group 3 students 1 standard deviation below the mean. The variation in scores was small for the Group 1 students compared to Group 2 and 3 students, although there was no significant difference between the variances.

A second oneway Analysis of Variance revealed that there was a significant difference in the scores in conceptual understanding between the three groups, $F(2,20) = 9.142$, $p = 0.002$. Planned comparisons showed that the mean score for conceptual understanding in Group 1 was significantly higher than that for Group 2, $t(20) = 2.259$, $p = 0.018$, and that the mean score for conceptual understanding in Group 2 was significantly higher than that for Group 3, $t(20) = 2.451$, $p = 0.012$.

Conclusion

The research hypothesis which has been investigated with these analysis, concerning the role of the sampling distribution in the student's conceptual structure and the level of conceptual understanding exhibited, has been confirmed by these analyses.

Given the size of the study group, it is not possible to conclude that there is no relationship between the role of the sampling distribution in the student's conceptual structure and the level of procedural understanding shown, as was also hypothesised. However, it can be said this relationship is weaker than that between the conceptual structure and conceptual understanding in this group of students. In order to establish whether or not such a relationship exists, the study would need to be repeated with a larger number of students.

5.6 Summary

In this chapter the data collected during the conduct of the teaching experiment was analysed, in order to address the research hypothesis:

> Students whose schema for sampling distribution demonstrates links to the sampling process and whose schema for statistical inference includes links to the sampling distribution, will show evidence of both conceptual and procedural understanding, whilst those who have not integrated these concepts into their cognitive structure may still demonstrate high levels of procedural understanding, but not of conceptual understanding.

In order to address these goals, the subjects in this study constructed a sequence of concept maps which provided external representations of the mental constructs created by individuals. From these maps it was possible to address the preliminary question explicitly. It was clear from the concept map analyses that the subjects differed in the ways in which their knowledge structures were organised. The number of concepts included in their maps in the form of propositions, and the number of interrelationships between concepts identified varied widely from student to student, with no two students constructing the same maps, and indicated that the schemas constructed by students for sampling distribution were highly individual.

It was possible to conclude from the students' concept maps that the role of the sampling distribution in statistical inference, evidenced by the links between the sampling distribution and the processes of estimation and hypothesis testing, was clearly recognised by some students, while other identified no relationship at all. That means some students did not even recognise that the sampling distribution has a place in the conduct of an hypothesis test, or in the construction of a confidence interval.

On the basis of the series of concept mapping exercises students were allocated to one of three groups. Students in Group 1 demonstrated schemas for sampling distribution which were similar to those constructed by experts, and which were linked to inference. Those in Group 2 also demonstrated schemas for sampling distribution which were similar to those constructed by experts, but were unable to link the sampling distribution to statistical inference. Finally, students in Group 3 demonstrated conceptual structures for sampling distribution which were confused and limited in the number of propositions and relationships included, and consequently were unable to be linked to statistical inference.

It was hypothesised that students in Group 1 would exhibit higher levels of conceptual understanding that those in Group 2, who would in turn perform better on tasks which measure conceptual understanding than those in Group 3. It was also hypothesised that these groupings would have little relationship to the students' procedural knowledge. The students, at the end of an instructional intervention designed to facilitate the formation and maintenance of the links between sampling distribution and statistical inference, undertook several tasks which were developed on the basis of current educational theory in order to give some insight into each student's level of understanding. At the completion of these tasks, the students could be assigned composite scores reflecting the levels of procedural and conceptual understanding demonstrated by the student. Subsequent analyses confirmed that indeed the nature of the relationship between the student's schemas for sampling distribution and statistical inference was as predicted. There was found to be no relationship between the form of the student's schema and procedural understanding on the basis of this group of subjects, but the small group size means that it is not possible to conclude that there is no relationship in general.

That there exists a relationship between the students' schemas and their level of understanding has been established in this Chapter. In the next chapter, detailed examination of the data collected from six students, two from each of Groups 1, 2 and 3, will be examined. From this examination more detail concerning the source of the relationship between the relevant schemas and the level of understanding demonstrated will be available, to further inform the research.

Chapter 6
Analysis of the case studies

6.1 Introduction

In this chapter case study analysis (Yin, 1989) for six of the students who participated in the study will be carried. The purpose of these analyses is to investigate the source of the relationship between the form of the students' schemas for sampling distribution and statistical inference that was established in the previous chapter.

Previously analysis of the concept maps for the students showed that the students could be grouped into three categories on the basis of the features of their conceptual structures evidenced by the series of concept maps constructed. These were:

Group 1	Students who showed evidence of the development of the concept of sampling distribution and subsequently integrated sampling distribution appropriately into their schema for statistical inference (7 students).
Group 2	Students who showed evidence of the development of the concept of sampling distribution but did not relate this to their schema for statistical inference (10 students).
Group 3	Students who did not at any stage show evidence of the development the concept of sampling distribution (6 students).

Two students have been selected from each of these groups for a more detailed descriptive analysis of the relationship between their schemas and levels of understanding demonstrated. The students have been labeled according to the Group to which they were allocated, so that G1S1 indicates Group 1, Student 1, and so on. For each of these six students the series of concept maps constructed during the course of the study was examined in detail. That is, 5 or 6 maps for each student, from the following list (some students omitted some maps due to absence from classes):

- Map 1 Concerned with the sampling distribution of the sample proportion.
- Map 2 Concerned with the sampling distribution of the sample mean.
- Map 3 Concerned with the sampling distribution.
- Map 4 Concerned with the hypothesis testing.
- Map 5 Concerned with the estimation.
- Map 6 Concerned with the statistical inference

Since the purpose of the concept maps was to provide an external representation of the student's schema at several stages during the study, an analysis of the sequence of maps prepared by a particular student allowed the researcher to trace the development of the student's schema over this period.

For each of the concept maps reported the list of terms suggested to the student at the time that the map was constructed is given, together with a detailed description of the concept map and a comment concerning the form of the student's schema as evidenced by the map. Following the analysis of all of the maps prepared by a student, these observations are drawn together to provide a description of the cognitive development of the student over the course of the study.

It is expected that the analysis will illustrate that students who exhibit certain conceptual links in their schemas will be able to access that link and thus successfully carry out conceptually related tasks, whilst students who have not formed such links will have problems with these tasks.

The concept maps illustrated in this chapter have been re-drawn for clarity. An example of an original map, together with the schematic diagram constructed from that map, is given in Appendix 6.

6.2 Analysis of Group 1 students

As previously reported, students who showed evidence of the development of the concept of sampling distribution and subsequently integrated sampling distribution appropriately into their schema for statistical inference were allocated to Group 1. Two students have been selected from Group 1 for detailed analysis. The first student from Group 1, (G1S1) was selected because she achieved the highest score overall for the composite measure of conceptual understanding, and thus would make an interesting case study. The second student (G1S2) was chosen at random from the other students in Group 1.

6.2.1 Student G1S1

Student G1S1 is a mature-age female undertaking her first formal course of study in statistics. She works full-time in the health science area. Prior to the teaching experiment she rated Low on the questionnaire which gave a self-assessed measure of prior knowledge. She achieved an outstanding score of 2.3 on the conceptual understanding measure factor score, and a marginally above average score of 0.4 on the procedural understanding measure factor score.

6.2.1.1 Concept map analysis for student G1S1

Map 1: Sampling distribution of the sample proportion

Suggested Terms: centre, computer generated, constant, distribution, normal model, p, $\hat{p}$, population, population parameter, population proportion, repetitions, sample proportion, sample size, sample statistic, sample(s), sampling distribution, sampling variability, spread, variable.

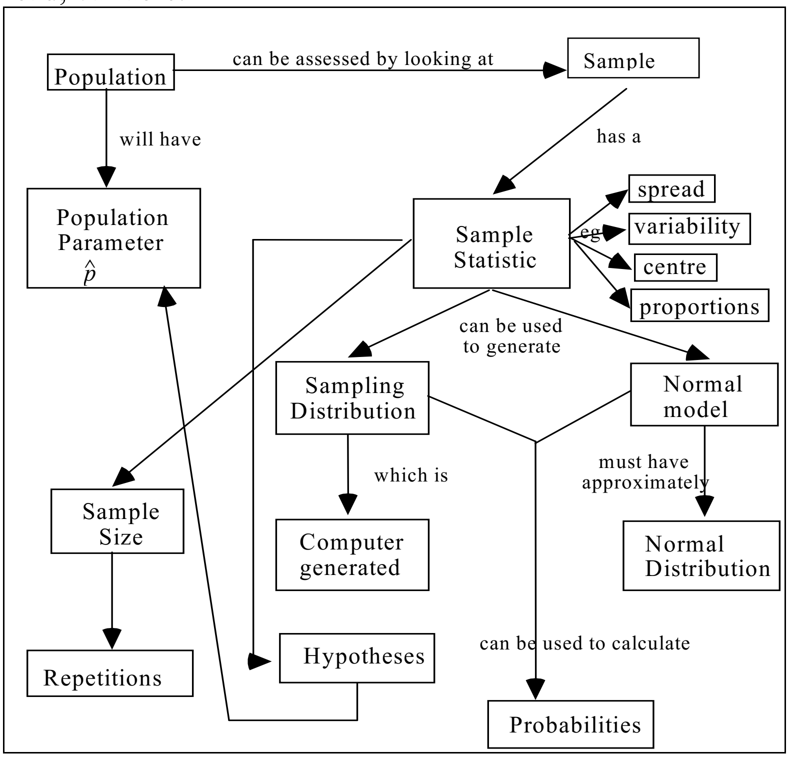


Figure 6.1 Map of sampling distribution for sample proportion for student G1S1 (Group 1)

Description

This map (see Figure 6.1) suggested that student G1S1 had made a promising beginning to understanding sampling distribution. Propositions indicating that the student recognised that samples enable information to be obtained about a population, and that populations are described by parameters and samples by statistics are present. However, the symbol $\hat{p}$ has incorrectly been included under population parameter. While this concept map was concerned with the sampling distribution of the sample proportion, this student has recognised that there are many sample statistics which can be calculated from the sample, but does not specifically link the sampling distribution to any one of them in particular. The sample statistic however was clearly indicated as the basis for the sampling distribution. It was recognised that the sampling distribution could either be empirical, based on repetitive sampling generated by a computer, or described by the normal model, the theoretical representation of the sampling distribution. Whilst the student has omitted both the terms *variable* and *constant,* from the map, there was a clear indication of the recognition of the sample statistic as a variable quantity. Also shown was a proposition indicating that the sampling distribution is the basis for determination of probabilities of particular sample outcomes, and the relationship between hypotheses and the population parameter. The role of sample size has not been addressed in this first map for sampling distribution.

Comment

In summary this concept map showed an emerging concept of sampling distribution which recognised both empirical and theoretical representations. Whilst there were some misunderstandings and omissions, the concept map suggested a developing well connected schema for sampling distribution, with the inclusion of five of the 10 key propositions contained in the expert map.

Map 2: Sampling distribution (general)

Suggested Terms:

centre, constant, μ, *p*, $\hat{p}$, population, population parameter, *r*, ρ, sample size, sample statistic, sample(s), sampling distribution, spread, variable.

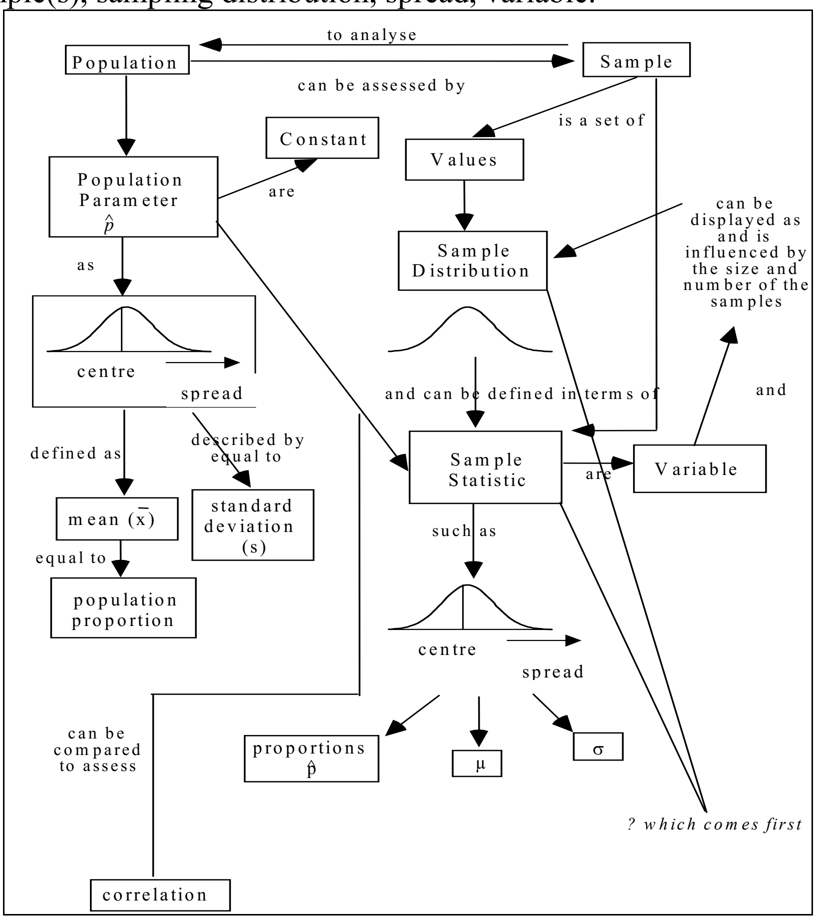


Figure 6.2 Map of sampling distribution (general) for student G1S1 (Group 1)

Description

This was a complex map (Figure 6.2) that included many cross-links between concepts. Close analysis reveals that student G1S1was undergoing a major cognitive reorganisation in regard to the concept of sampling distribution. The student had realised that there are in fact *three* distributions involved - the distribution of the population, the distribution of the sample, and the distribution of the sample statistic.

G1S1 also understood that both a sample distribution and a sampling distribution exist, but was not clear on the use of the terminology, and was attempting to use the term *sample distribution* for both the *sample distribution* and the *sampling distribution*. The confusion was evident from the comment "which comes first" shown on the map. Also shown is a proposition recognising that the sampling distribution of the sample statistic is influenced by the size of the sample selected, indicating that as part of her cognitive reorganisation the student has incorporated sample size into her schema for sampling distribution.

Comment

This map indicates that the student has undergone considerable conceptual growth since the last map was prepared, and on this map she has included 9 of the ten propositions identified on the expert map

Map 3: Hypothesis testing

Suggested terms: alternate, decision, hypotheses, null, P-value, population, population parameter, sample statistic, sample(s), sampling distribution, sampling variability, significance level, test statistic.

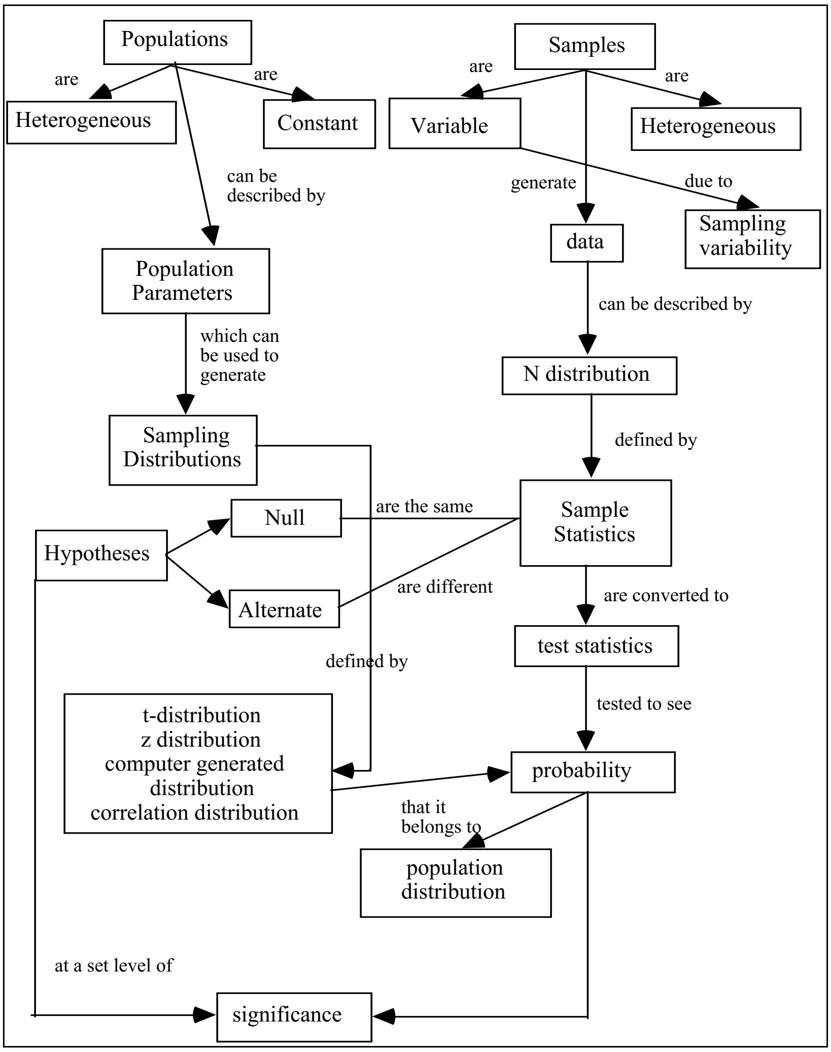


Figure 6.3 Map of hypothesis testing for student G1S1 (Group 1)

Description

In this map (Figure 6.3) there was a distinction made right at the start between populations and samples. Both populations and samples were recognised as heterogeneous, being made up of a variety of individuals. However, populations were labeled as constant while samples were recognised as variable, due to sampling variability. Populations were described by population parameters, which give rise to sampling distributions. With this proposition “population parameters can be used to generate sampling distributions” the student was recognising that the sampling distribution used to determine the likelihood of a particular sample outcome is generated under the null hypothesis. That is, a particular value for the population parameter must be assumed, otherwise the sampling distribution is not known exactly. This recognition of the link between the population parameter and the sampling distribution indicated a good understanding of the role of the sampling distribution in hypothesis testing. A high level of conceptual understanding is further supported by the inclusion of the t-distribution, the z-distribution, the computer-generated distribution and the correlation distribution as examples of sampling distributions. The fact that both the theoretical and empirical representations of the sampling distribution were included in this student's concept map also indicated that the student G1S1 recognised that the sampling distribution could exist in both forms and that these, for all practical purposes, were equivalent in hypothesis testing. This was also confirmed by the link between these sampling distributions and probability (P-value). Also inherent in this map was a clear indication of the procedure of hypothesis testing, which tracked through the map from sample statistics to test statistic, on to the P-value which was linked to significance and then back to the hypotheses. The only obvious problem in the map was the statement that the data can be described by the normal distribution, which is not always true, but has been true for most of the situations which the student has studied.

Comment

In this map many important relationships are evidenced by the cross-links between concepts. The propositions present in this concept map suggests that this student had reconciled the empirical and formal/theoretical representations of sampling distributions, and that G1S1 had also established a clear appreciation of the place of the sampling distribution in hypothesis testing. The map included 9 of the 14 propositions identified in the expert concept map for hypothesis testing.

Map 4: Estimation

Suggested terms: confidence interval, estimation, interval estimates, point estimates, population, sample, sample statistics, sampling distribution.

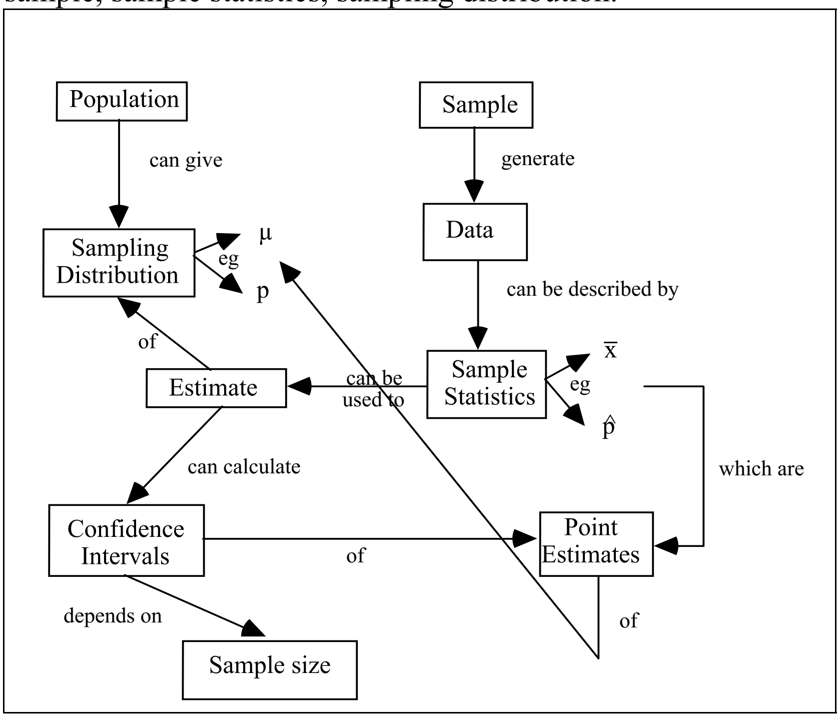


Figure 6.4 Map of Estimation for student G1S1 (Group 1)

Description

The role for the sampling distribution in estimation was less obvious in this concept map (Figure 6.4) than in the previous concept map for hypothesis testing. The sampling distribution was again formed from the population, but the relationship was not so clearly indicated in this map as it was in the last. The rationale here may be the same as it was in the student's map 3, where the sampling distribution was generated using knowledge of the values of the population parameters, particularly as μ and p are linked to the sampling distribution.

A very interesting link was that which formed the proposition "sample statistics can be used to estimate (of) sampling distribution". Given that the student's conception of the sampling distribution is based on knowing the parameter value, then this proposition makes sense. That is, one could estimate the sampling distribution on the basis of the sample statistics, and that is in fact what is done in the calculation of a confidence interval. The map also showed the relationship of sample size to the confidence interval and correctly linked the sample estimates to the appropriate population parameters. This student was by now consistently using the correct notation for population parameters and sample statistics.

Comment

In summary, in the concept map for estimation the precise role for sampling distribution was less clear than in the map for hypothesis testing, but the fact that sampling distribution was included in a key position on the map indicated that the student has perceived some role for the sampling distribution in estimation. The map included 4 of the ten key propositions identified in the expert map for estimation.

Map 5: Statistical Inference

Suggested terms: confidence interval, decision & conclusion, estimation, hypotheses, hypothesis testing, inferential statistics, interval estimates, P-value, point estimates, population, sample, sample statistics, sampling distribution, significance level, statistical significance, test statistic.

Description

This was again a complex map (Figure 6.5) including many links between concepts. How was the role that the sampling distribution played in statistical inference been identified in this concept map? The concept map showed that together the sample statistic and the sampling distribution led to the formation of a test statistic. Since knowledge of the appropriate sampling distribution for the sample statistic dictates the form of the test statistic, the sampling distribution has been correctly located. That the hypotheses are concerned with the population parameters was clearly shown, as was the link between the point estimates and the sample statistics. The logic of statistical testing was clearly and correctly indicated. There was no link shown between estimation and the sampling distribution in this map, nor was the potential equivalence of the confidence interval and the hypothesis test identified.

Comment

This map evidenced both the correct procedures of hypothesis testing as well as the conceptual link between the procedure and the sampling distribution. The link between estimation and the sampling distribution is less clear. The map includes 9 of the 14 key propositions identified in the expert map.

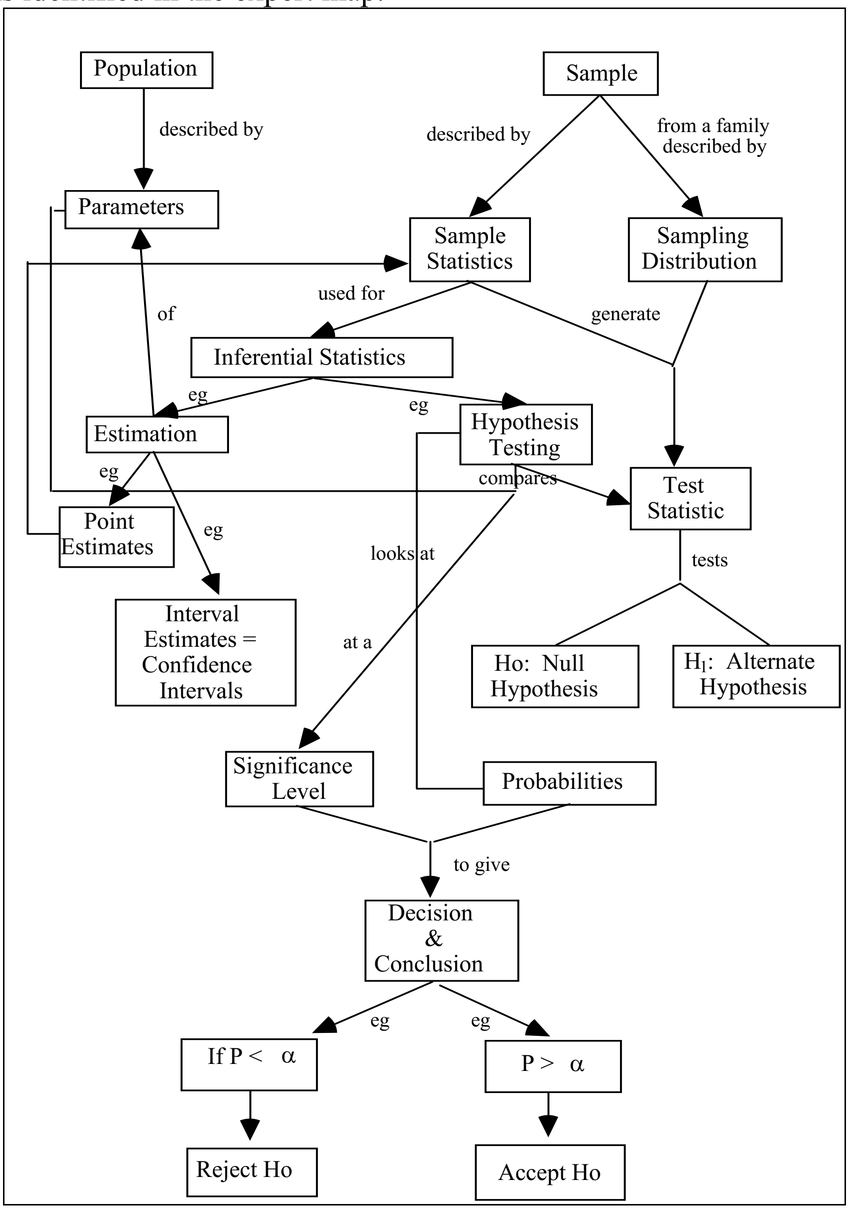


Figure 6.5 Map of Statistical Inference for student G1S1 (Group 1)

Summary of schema development for student G1S1

Student G1S1 has exhibited through this series of concept maps recognition of the existence of and the relationships between the distribution of the population, the distribution of the sample, and the sampling distribution. No other sequence of maps prepared for this study, or for the preliminary study, or even by the experts, indicated explicitly that there were in fact three distributions under discussion. This is an interesting pedagogical point, and leads one to wonder why the experts did not include the sample distribution. Perhaps the experts considered it too obvious, or not relevant to the study of the sampling distribution. However, in the light of this students' obvious early confusion over the three distributions, which was later clarified in her schema for inference, it may well be that this distribution is important and should be explicitly discussed. This issue is addressed in more detail in Chapter 7.

The concept maps showed that student G1S1 had recorded a rich network of valid relationships for sampling distribution which included both the empirical and the formal/theoretical representations, and recognition of the relationship between these representations in the conduct of hypothesis testing. The role of the sampling distribution was clearly and correctly identified in relation to hypothesis testing. In particular, this student consistently noted that the specification of the exact sampling distribution used in hypothesis testing depends on the value of the population parameter hypothesised (the only student to do so) and that the sampling distribution was the basis of the determination of the P-value (one of five students to do so).

6.2.1.2 Relating conceptual structure to understanding for student G1S1

The following table gives the scores allocated to each of the tasks for student G1S1. The maximum score allocation for the task is also given in brackets.

Table 6.1 Tasks scores for student G1S1

Task	Score
Chi square	8 (9)
Correlation (a)	5 (5)
Modelling (b)	5 (7)
t-test	10 (10)
Modelling (a)	7 (7)
Sample size	3 (3)
Two group *t*-test	8 (8)
Explanation	8.5 (10)

Radio	5 (5)
Hospital	2 (2)
Confidence interval	2 (2)

Student G1S1 received a factor score of 0.4 for the procedural understanding measure, a result slightly above the average for the group (ranking 11th out of 23 students on this variable). The student's result in the conceptual understanding measure was an outstanding 2.3, the highest achieved by any student in the study group (being 2.3 standard deviations above the mean).

Consistently G1S1's answers indicated knowledge of the sampling distribution and an ability to relate this knowledge in a variety of situations. For example, G1S1's answer to the Explain task (see Section 4.8), which required the student to explain the process of carrying out a hypothesis test to a non-statistically trained person, included the following description of the significance level:

> ...due to random sampling it is possible to get a result that suggests there is a correlation when there is not. Significance is the probability that the result is due to sampling alone. A smaller value of α increases the likelihood of the difference being significant. At
> $\alpha = 0.05$, we will accept a 5% chance that the result is due to sampling variability.

Whilst this statement is not entirely correct (actually a larger value of α increases the likelihood of the difference being significant), it is in fact an extremely good interpretation of the significance level and what is meant when it is set at 5%.

The explanation given by G1S1 of the P-value when explaining the process of hypothesis testing was as follows:

> ...is the probability that the result is due to random sampling. This is multiplied by 2 because it is testing in both directions. It is calculated from a family of curves known to represent the probability of results for given samples. The curves are density curves and the area under them is equal to 1 or 100%. From the tables the area in the tail is read off and represents the probability of getting the sample result due to sampling variation.

Interestingly, here G1S1 has chosen to give the formal/theoretical form of the sampling distribution (as a density curve). Whilst the P-value is not the probability of any particular result actually occurring, her explanation is consistent with the language used in instruction, clearly explained, and importantly, related to the sampling process. This answer confirms that student G1S1 has integrated both the empirical and theoretical representations of sampling distribution into the appropriate schema, and is able to move between them.

Under Decision and Conclusion G1S1 in her explanation wrote:

> The probability that the result was due to sampling is < 0.05 which was the predetermined level that we would accept as not acceptable to say it was due to sampling variability. ∴ The result we got is not due to sampling variability but is actually due to the fact that there is a relationship between the intelligence of mother and child.

G1S1 correctly interpreted the conclusion made in the Explain task, that we would not attribute the correlation observed to sampling variability, but rather conclude that there is a relationship between these variables in the population.

Further insight into the conceptual understanding of statistical inference can be gained from her response to the Radio task (see Section 4.8), in which students are asked to suggest reasons why the observed proportion of listeners might be different from the claimed proportion. G1S1's answer to the Radio task was as follows:

> Reasons for the difference between the two claims could be:-
>
> 1 The sample was not representative - ie - not random
> - questions not correctly asked
> ∴ claim still true
>
> 2 The sample was representative but the standard deviation is large
> ∴ claim still true
>
> 3 The sample was representative but due to random sampling error gave an unusual result.
> ∴ claim still true
>
> 4 The advertisers claims are false

In her answer G1S1 logically argues all possible explanations for the difference between the two values, showing an ability to relate the given scenario to schemas linking sampling, sampling distribution and hypothesis testing, and in particular the effect of sample size on the variability of the sample statistic.

The analysis of the tasks answers given by student G1S1 indicates a high level of conceptual understanding, consistent with the prediction made on the basis of the form of the conceptual structure created by G1S1 as evidenced by the concept maps, in which numerous key concepts were identified.

6.2.2 Student G1S2

Student G1S2 also is a mature-age female undertaking her first formal course of study in statistics, and also works full-time in the health science area. She scored a Medium rating on the self-assessed prior knowledge survey. She achieved a creditable score of 0.8 for the conceptual measure, and a slightly above average score of 0.5 for the procedural measure.

Student G1S2 was randomly selected from the students allocated to Group 1 for further analysis.

6.2.2.1 Concept map analysis for student G1S2

Map 1: Sampling distribution of the sample proportion

Suggested Terms:

centre, computer generated, constant, distribution, normal model, p, $\hat{p}$, population, population parameter, population proportion, repetitions, sample proportion, sample size, sample statistic, sample(s), sampling distribution, sampling variability, spread, variable.

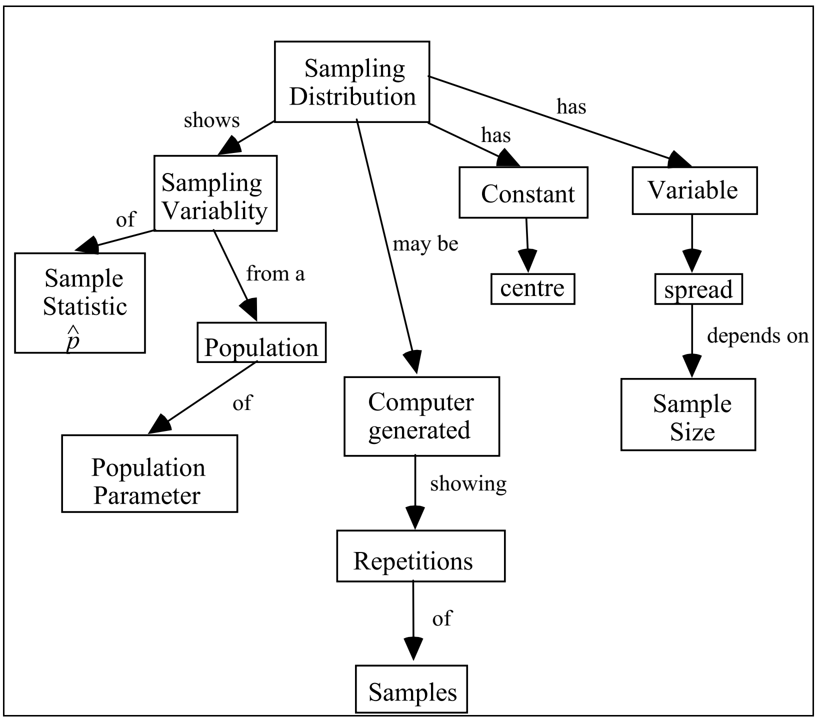


Figure 6.6 Map of sampling distribution for sample proportion for student G1S2 (Group 1)

Description

The concept map for the sample proportion constructed by G1S2 is shown in Figure 6.6. The map is clear, and shows many important propositions, including that the variability of sample statistic was summarised in the sampling distribution, and that the sampling distribution had a constant centre but its spread was dependent on the sample size. These were relevant concepts for the work that followed in statistical inference. It was not clear from this map whether or not the student had integrated the theoretical model for sampling distribution into this schema. The proposition "may be Computer generated" might have indicated that the student was aware that there was another form of the distribution, but it also might not.

Comment

The concept map shows four distinct strands, but no cross links, indicating that some of the important links were yet to be made. The map included 5 of the ten key propositions from the expert concept map.

Map 2: Sampling distribution of the sample mean

Suggested Terms:

centre, computer generated, constant, distribution, normal model, μ, $\bar{x}$, population, population mean, repetitions, sample mean, sample size, sample statistic, sample(s), sampling distribution, sampling variability, spread, variable

Description

Once again the sampling distribution was clearly indicated in this map as describing the distribution of the sample statistic. Also indicated (again) was the effect of sample size on the spread of the sampling distribution. This map showed increasing complexity from the last map in that, while retaining the key features from the last map, further relationships had been identified. These links indicated that the student had now recognised that:

- the centre of the sampling distribution was at the population parameter; and
- the sampling distribution may be either computer generated, or derived using the normal model.

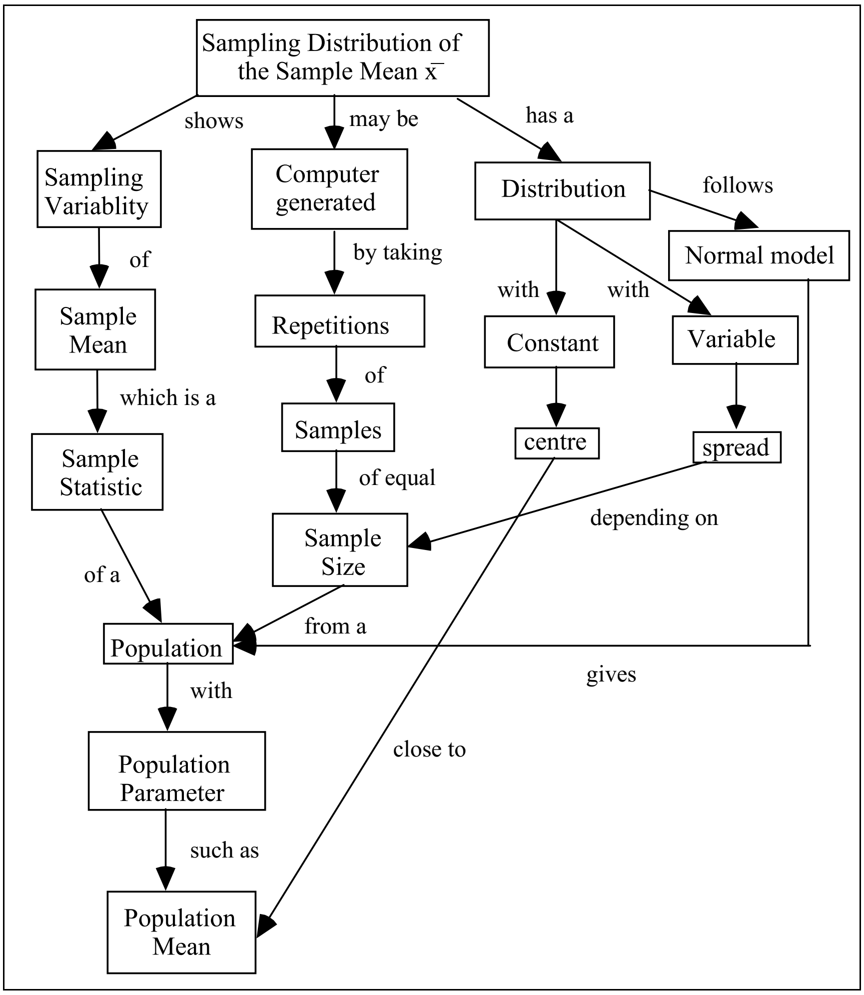


Figure 6.7 Map of sampling distribution for sample mean for student G1S2 (Group 1)

Comment

The map still consists four strands, but at this stage many more relationships between concepts have been identified, which is evidence by the cross-links which now appear in the map. The map is comparable to one which may have been considered to be from an expert, including 8 of the ten propositions identified in the expert map.

Map 3: Sampling distribution (general)

Suggested Terms:

centre, constant, μ, p, $\hat{p}$, population, population parameter, r, ρ, sample size, sample statistic, sample(s), sampling distribution, spread, variable.

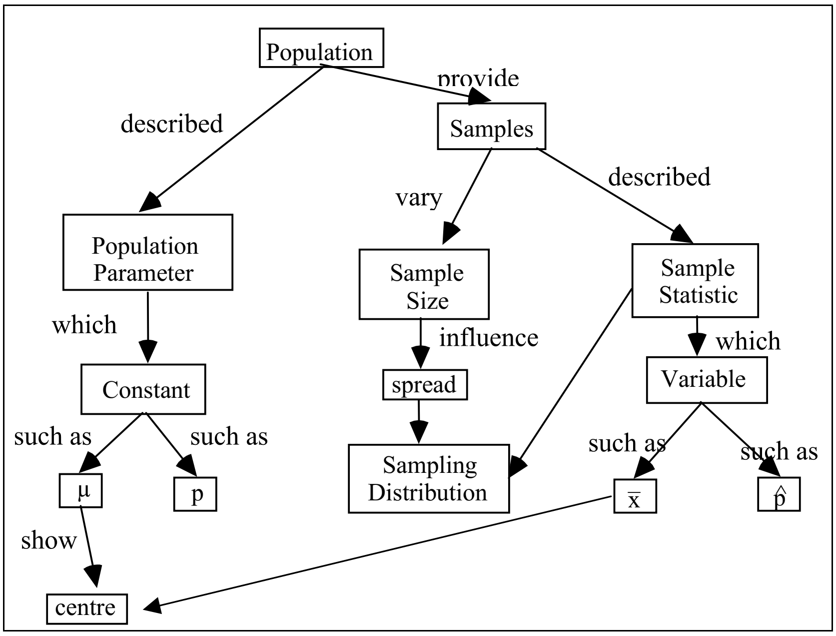


Figure 6.8 Map of sampling distribution (general) for student G1S2 (Group 1)

Description

The concept map prepared here by G1S2 (Figure 6.8) was quite different in structure to the two earlier maps. Whilst the map was again clearly structured, it showed fewer of the key propositions than the last. In particular, the centre of the sampling distribution was not shown as occurring at the value of the parameter, although a less sophisticated link was made between μ and $\bar{x}$. Adapting the earlier map to include the other statistics had been difficult. This could have been because the identification of the population proportion and population correlation as the *centres* of distributions seemed to be a contradiction. However, the other key features of earlier maps were still present, such as the clear distinction between population and sample, parameter and statistic. The sampling distribution was correctly seen as describing the sample statistic, and the influence of sample size on the spread of the distribution clearly noted. The correlation coefficient symbols (ρ and r) were also omitted. This was possibly because the student is not clear what they represent.

Comment

The map indicated that the process of integrative reconciliation of the examples discussed for sampling distribution into a generalised schema for sampling distribution was achieved for mean and proportion, but there was a reduction in the relationships recognised in the process. This map included 6 of the ten propositions shown on the expert map.

Map 4: Estimation

Suggested terms

confidence interval, estimation, interval estimates, point estimates, population, sample, sample statistics, sampling distribution.

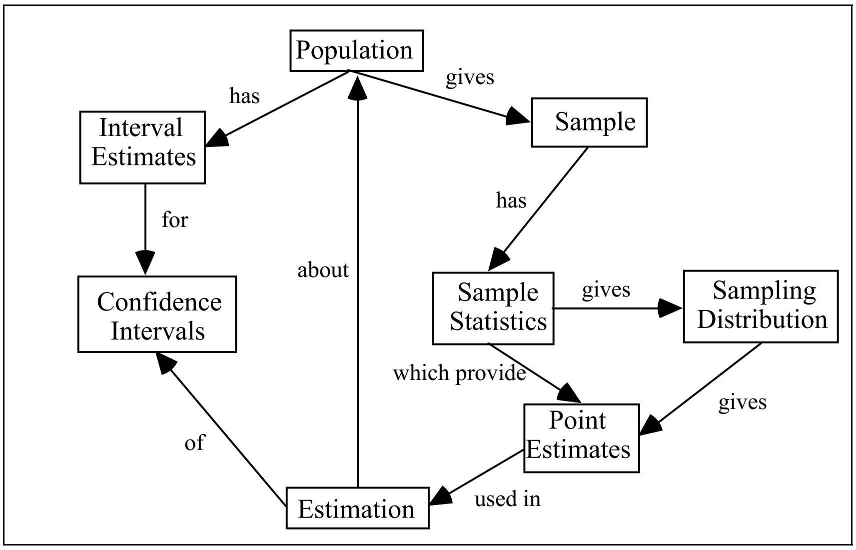


Figure 6.9 Map of estimation for student G1S2 (Group 1)

Description

Again, student G1S2 prepared a clear map with several cross-links indicating relationships between concepts. The proposition that estimation is concerned with the population was present in this map, although the term *population parameter* was been omitted. The sampling distribution was correctly describing the sample statistic, and was linked through point estimate to the process of estimation and then on to confidence interval.

Comment

Overall this is map pointed to a possible role for sampling distribution in estimation, although the exact nature of that role was not clear. This map contained 6 of the ten propositions included in the expert map for estimation.

Map 5: Statistical Inference

Suggested terms

confidence interval, decision & conclusion, estimation, hypotheses, hypothesis testing, inferential statistics, interval estimates, P-value, point estimates, population, sample, sample statistics, sampling distribution, significance level, statistical significance, test statistic.

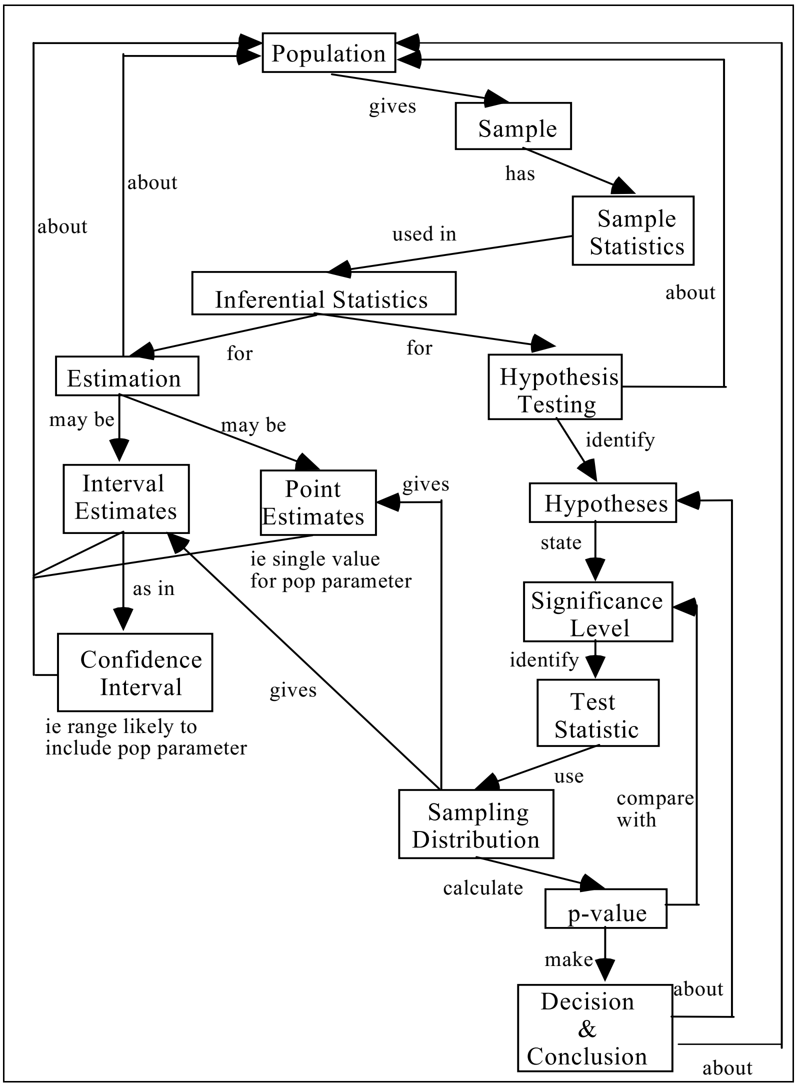


Figure 6.10 Map of Statistical Inference for student G1S2 (Group 1)

Whilst this concept map did not directly link the sampling distribution to the sample statistic, it was reasonable to assume that the sampling distribution was correctly specified as it was shown as summarising the sample statistic in all of this student's previous maps. The sampling distribution was linked to both the hypothesis testing and estimation strands of the concept map. In the hypothesis testing section, the sampling distribution was shown as the basis of determination of the P-value. In the estimation section, the sampling distribution was linked to both the point estimate and the interval estimate. The confidence interval was quite correctly explained as the range likely to include the population parameter.

Comment

This concept map included several cross-links, and indicated a role for sampling distribution in both hypothesis testing and estimation, but omitting the link between the hypothesis test and estimation as potentially equivalent actions. Eight of the 14 propositions shown on the expert map, were included.

Summary of schema development for student G1S2

Student G1S2 demonstrated throughout this series of concept maps all of the propositions that describe the concepts associated with sampling distribution, including both empirical and the formal/theoretical representations and that sample size affects the spread of the sampling distribution. The place of the sampling distribution in both hypothesis testing and estimation was also clearly recognised.

6.2.2.2 Relating conceptual structure to understanding for student G1S2

The following table gives the scores allocated to each of the tasks for student G1S2. The maximum score allocation for the task is also given in brackets.

Table 6.2 Tasks scores for student G1S2

Task	Score
Chi square	9 (9)
Correlation (a)	5 (5)
Modelling (b)	7 (7)
t-test	10 (10)
Modelling (a)	7 (7)
Sample size	3 (3)
Two group *t*-test	6 (8)
Explanation	8 (10)
Radio	3 (5)
Hospital	0 (2)
Confidence interval	2 (2)

Student G1S2 achieved a factor score of 0.5 for the procedural understanding measure, also a result slightly above the average for the group (ranking 8th of the 23 students in the group). Thus student's result on the conceptual understanding measure was 0.8, almost one standard deviation above the mean, and ranking 5th of the 23 students in the group.

A detailed analysis of the tasks undertaken to demonstrate conceptual understanding by student G1S2 indicated that her answers generally showed knowledge of the sampling distribution and an ability to relate this knowledge in a variety of situations. The exception to this was the hospital question, where option C (there would be no difference) was chosen.

However, in the Explain task the answers given by G1S2 were generally clear and insightful. For example, the significance level was interpreted as follows:

> 95% sure that the result we reach has not occurred by chance alone

Whilst this interpretation of the significance level is not theoretically correct, it gives some sense that what we about in a hypothesis testing situation is consideration that the observed results just may have occurred by chance, which as absolutely correct. This interpretation is one which is readily understood by a person who has not been statistically trained, the stated purpose of the exercise.

G1S2's explanation of the P-value was as follows:

> This shows the probability of getting that result for r in our sample if we took the sample from a population where $\rho = 0$ (ie where there was no relationship).

This explanation of the P-value is excellent as it relates directly the observed value of the test statistic and the likelihood of this occurring under the null hypothesis. The sampling distribution indirectly referred to here is the empirical representation, relating the P-value to the sampling process and indicating the connection that exists between the theoretical and empirical representations for this student.

In the section labeled Decision and Conclusion, G1S2 wrote:

> The chance of sampling variability alone giving the result that we found for r is very low (less than 5%) and $\therefore$ we conclude that the sample came from a population that has a relationship i.e. There is a relationship between the intelligence of children and their mothers in general.

G1S2 correctly interprets the result of the hypothesis test, that is the rejection of sampling variability as an explanation for the value of the sample statistic, and the acceptance of the explanation that the sample comes from a different population where there is a relationship.

The answer given by G1S2 to the Radio tasks, concerning the possible reasons for the difference between the observed and claimed proportion of listeners was as follows:

If knew the sample size could use z as test statistic.

If sample only 1

H_0: p=0.2

H_1: p≠0.2

$$z = \frac{0.15-0.2}{\sqrt{0.2\times0.8}}$$

$$= \frac{-0.05}{0.40}$$

$$= -0.125$$

$$P = 0.45 \times 2$$

$$= 0.9$$

P > 0.05 cannot reject null hypothesis.

With given information cannot conclude radio station is misleading them. Do not know sample size or how it was chosen.

This answer indicated that the student initially connected this scenario with the schema for hypothesis testing, correctly interpreting the situation with a hypothesised value of 0.2 (the radio station claim), and the observed sample statistic of 0.15. This hypothesis testing schema was linked to the theoretical sampling distribution, which the student attempted to use to answer the question, and then realised this was not going to be totally successful. Finally, the student included the sampling strategy used and the size of the sample as determinants of the validity of the claim. Thus the student had made a further conceptual link and connected the scenario with the sampling process, and the empirical sampling distribution.

An analysis of the tasks undertaken by student G1S2 indicate an above average level of conceptual understanding, consistent with the prediction made on the basis of the form of the conceptual structure created by G1S2 and evidenced by the concept maps.

6.3 Analysis of Group 2 students

Students who showed evidence of the development of the concept of sampling distribution but did not relate this to their schema for statistical inference were allocated to Group 2. Both of the students chosen for analysis in this section were selected at random from the group.

6.3.1 Student G2S1

Student G2S1 is a mature-age female undertaking her first formal course of study in statistics. She also works full-time in the health science areas. Prior to the teaching experiment she rated Low on the self-assessed prior knowledge questionnaire. She achieved a score of 0.3 on the procedural understanding measure, and 0.2 on the conceptual understanding measure.

6.3.1.1 Concept map analysis for student G2S1

Map 1: Sampling distribution of the sample proportion

Suggested Terms: centre, computer generated, constant, distribution, normal model, p, $\hat{p}$, population, population parameter, population proportion, repetitions, sample proportion, sample size, sample statistic, sample(s), sampling distribution, sampling variability, spread, variable.

Description

The concept map prepared by this student uses correct terminology and notation in relation to populations and samples. Whether the sampling distribution is based on the sample or the sample statistic is not clear which from this map. The effect of sample size on the sampling distribution has been noted. It can also be seen that the student has included both the empirical and the theoretical sampling distributions and recognised a relationship between them. There are some unusual links which form propositions (such as the normal model is used to calculate the constant) which are difficult to interpret.

Comment

In summary, this student has included many valid propositions in this map, some of which include terms were not on the suggested list such as *histogram*. G2S1's map integrates the key ideas that were developed in the teaching sessions which formed part of the teaching experiment. Whilst not all that has been proposed in the map is quite correct, G2S1 has included six of the ten key propositions identified in the expert map.

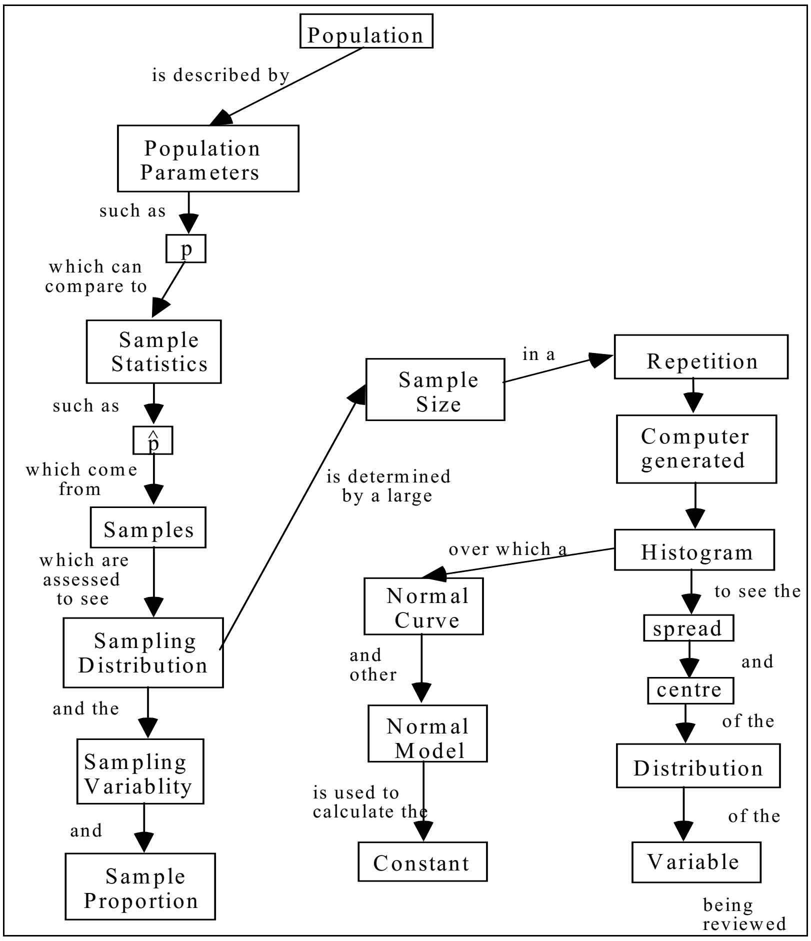


Figure 6.11 Map of sampling distribution of the sample proportion for student G2S1 (Group 2)

Map 2: Sampling distribution of the sample mean

Suggested Terms:

centre, computer generated, constant, distribution, normal model, μ, $\bar{x}$, population, population mean, repetitions, sample mean, sample size, sample statistic, sample(s), sampling distribution, sampling variability, spread, variable

Description

From this concept map the sampling distribution could be identified as arising from the sample statistic. It could also be seen from the map that the student was aware that the sampling distribution could be generated from repetitive sampling or modeled using the normal distribution. Also included was the proposition noting that the sampling distribution could be used to gauge sampling variability.

Comment

The student G2S1's concept of sampling distribution was less linear in this map than the earlier map, and indicated clarification of the concepts involved. In particular, there was clear recognition of the equivalence of the empirical and formal/theoretical representations for sampling distribution. This map included only four of the ten key propositions included in the expert map, since some of the propositions relating samples, populations, statistics and parameters had been omitted.

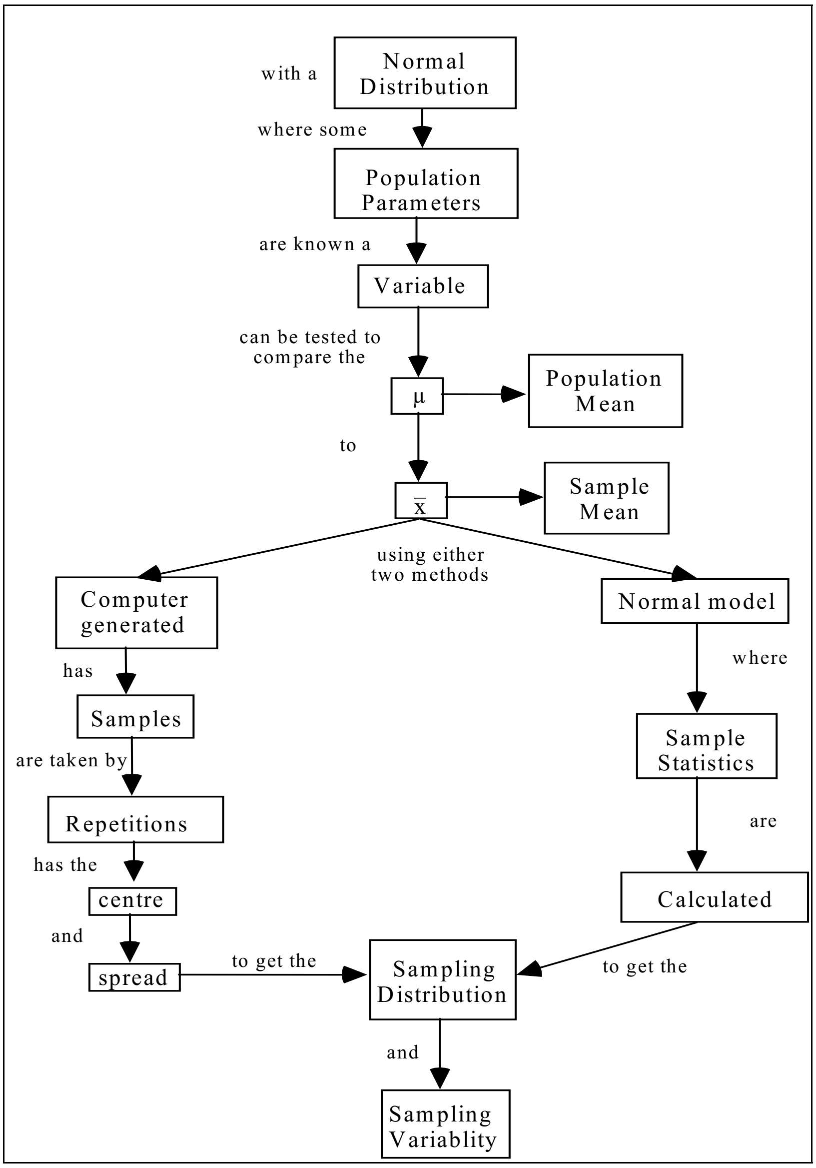


Figure 6.12 Map of sampling distribution of the sample mean for student G2S1 (Group 2)

Map 3: Sampling distribution (general)
Suggested Terms:

centre, constant, μ, p, $\hat{p}$, population, population parameter, r, ρ, sample size, sample statistic, sample(s), sampling distribution, spread, variable.

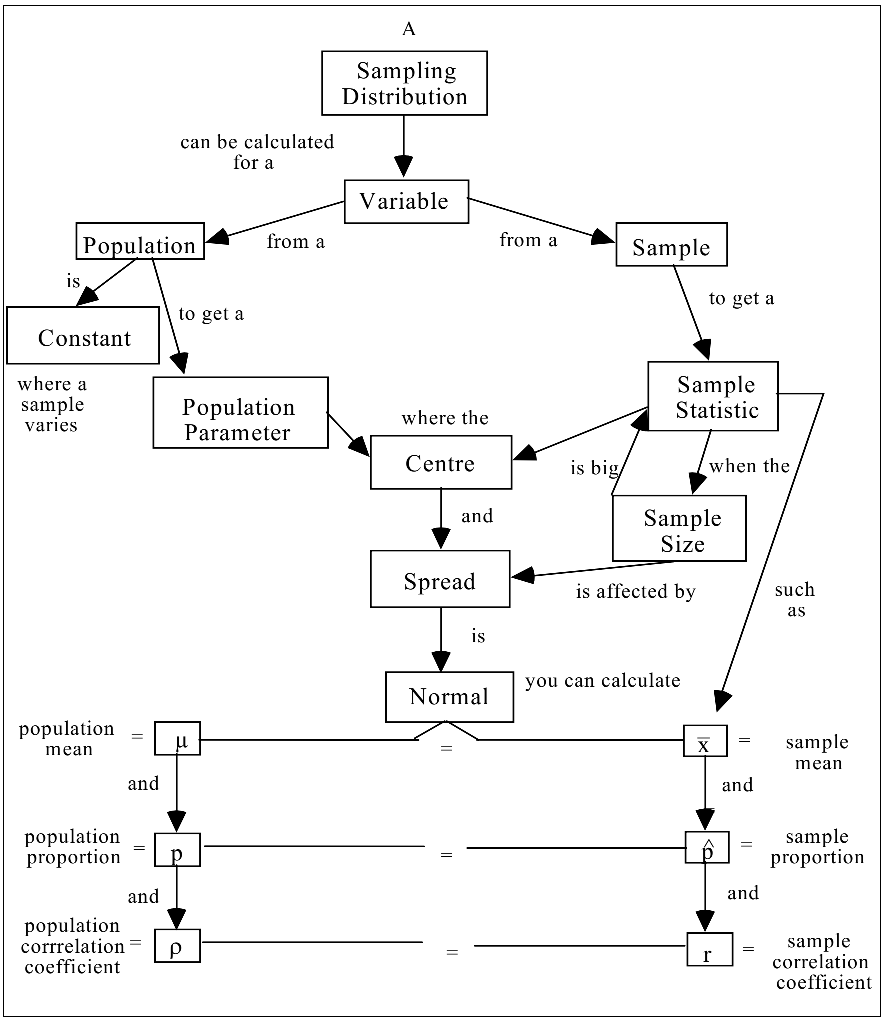


Figure 6.13 Map of sampling distribution (general) for student G2S1 (Group 2)

Description

This map clearly showed that the sampling distribution is based on the sample statistic, and also identified the relationship between the spread of the sampling distribution and the sample size. Also clearly shown is the idea that the sampling distribution may be derived from knowledge of the population, or from the sample. The left branch of the map may be considered to have described the formal/theoretical sampling distribution, whereas the right branch described the empirical sampling distribution. These came together and both were seen to embody the features of shape, centre and spread. Only the spread of the empirical sampling distribution has been seen to be affected by sample size in this map. G2S1 also demonstrated excellent use of notation, and the existence (and potential equivalence) of the various population parameters and sample statistics.

Comment

In this map G2S1 demonstrated integrative reconciliation of the sampling distributions which had been discussed in specific terms for proportion, mean and correlation, and was able to see each of these as examples of a general concept. The map included 8 of the ten key propositions identified in the expert map.

Map 4: Hypothesis testing

Suggested terms: alternate, decision, hypotheses, null, P-value, population, population parameter, sample statistic, sample(s), sampling distribution, sampling variability, significance level, test statistic.

Description

This map was remarkably linear in structure, suggesting that it had become more of a flow chart than a concept map. The map described the process of hypothesis testing rather than identifying the concepts underpinning hypothesis testing. The decision term was incorrectly related to the sample rather than the population, and the hypotheses were formed on the basis of the sample statistics, which was also an incorrect connection. The sampling distribution was seen here in its formal/theoretical representation. The student identified the step associated with the sampling distribution as concerned with the type of test to be carried out (z-test, *t*-test etc) so that the appropriate test statistic could be calculated.

Comment

This map described the steps undertaken when carrying out an hypothesis test rather in the form of a recipe. Many of the relationships between key propositions identified in earlier maps were omitted. This map included only two of the 14 key propositions identified in the expert map.

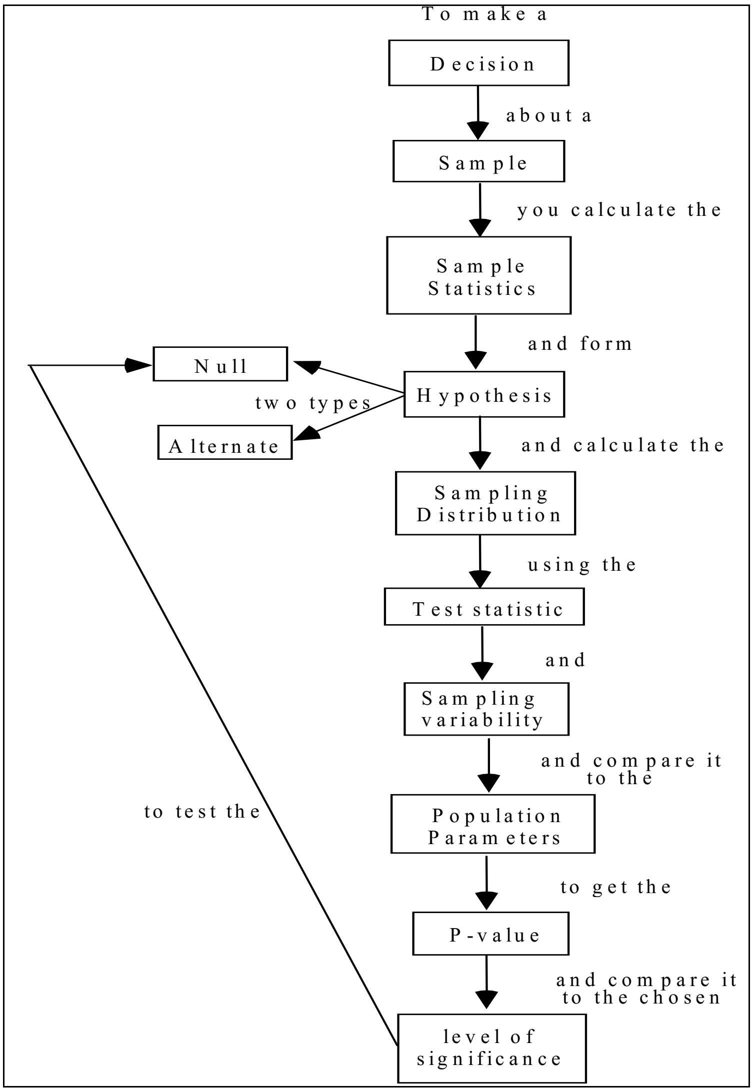


Figure 6.14 Map of hypothesis testing for student G2S1 (Group 2)

Map 5: Estimation

Suggested terms

confidence interval, estimation, interval estimates, point estimates, population, sample, sample statistics, sampling distribution,

Description

Once again the map prepared by G2S1 was quite linear, and written as a flow chart. There were no cross-links at all, and the map described the procedure of calculating a confidence interval. Even the language used to describe the links was written in the active voice, as if dictating a set of instructions, rather than attempting to link the concepts. The sampling distribution described the distribution used for the determination of the confidence interval, which was a correct application of the formal/theoretical representation.

Comment

This map described the process of estimation in the form of a recipe. Again, many of the relationships between key propositions identified in earlier maps were omitted. This map included only three of the 10 key propositions identified in the expert map.

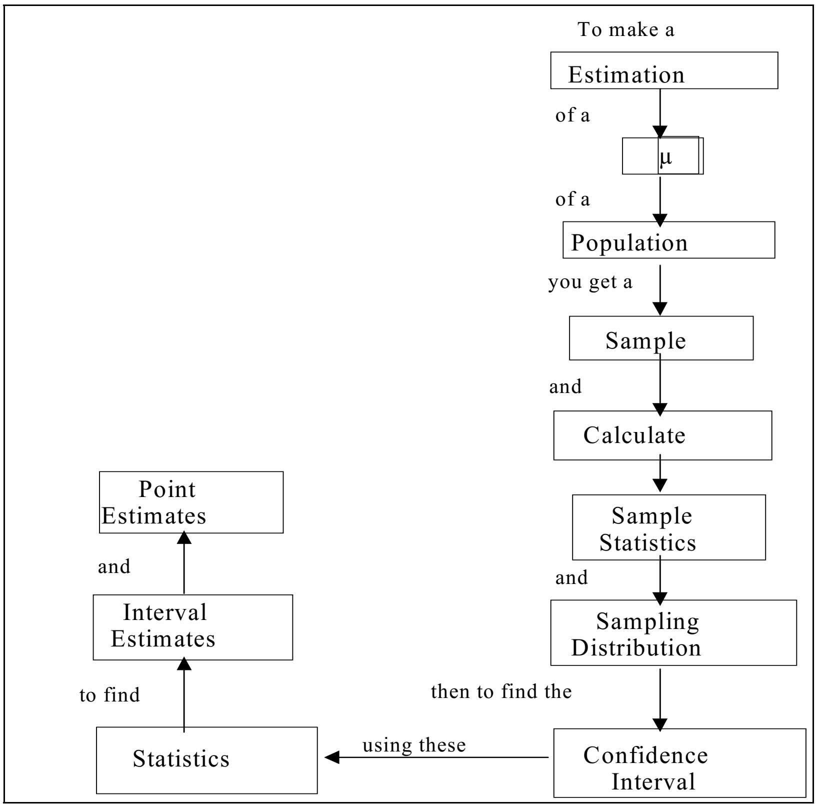


Figure 6.15 Map of estimation for student G2S1 (Group 2)

Map 6: Statistical inference

Suggested terms

confidence interval, decision & conclusion, estimation, hypotheses, hypothesis testing, inferential statistics, interval estimates, P-value, point estimates, population, sample, sample statistics, sampling distribution, significance level, statistical significance, test statistic.

Description

From this map we can see that the student was quite clear as to the purpose of and the methods used for carrying out hypothesis tests and determining confidence intervals. G2S1 indicated clearly that inference was concerned with population parameters, which are unknown. G2S1 was also clear that the decision that was made in statistical testing was whether or not to reject or retain the null hypothesis. The procedure for hypothesis testing was also logically correct. What was again evident was the linear, flow chart like structure of the map, with only one cross-link. And, clearly for this student statistical inference was not concerned with the sampling distribution. The sampling distribution was not integrated into the schema for inference for this student, since, even though the term *sampling distribution* was given on the suggested list of terms, it was omitted altogether from this map. This implied that student G2S1 had constructed a schema for inference that consisted almost entirely of procedural links.

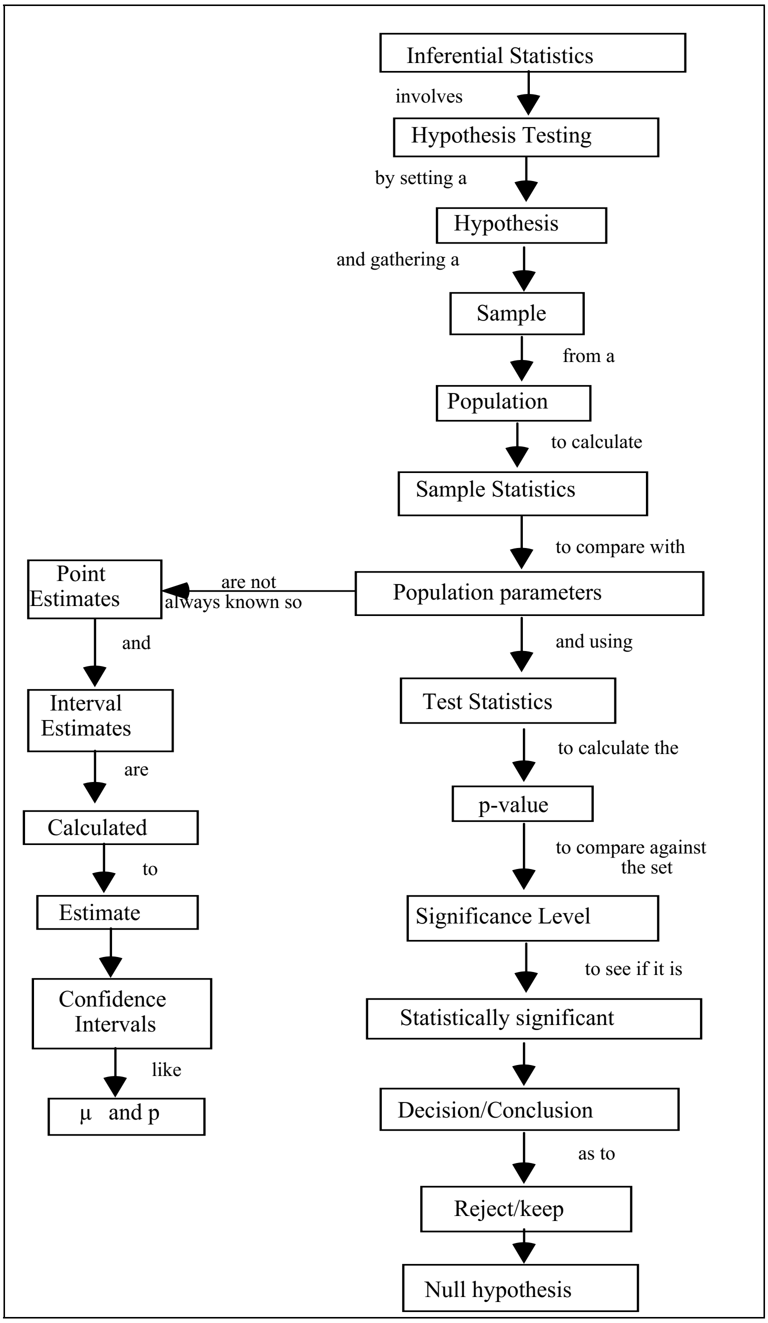


Figure 6.16 Map of Statistical Inference for student G2S1 (Group 2)

Summary of schema development for G2S1

Student G2S1 demonstrated a schema for sampling distribution that showed knowledge of the sampling distribution as describing the sample statistic. Both the empirical and formal/theoretical representations of sampling distribution were identified. The effect of sample size on the spread of the sampling distribution was also recognised. However, this student was not able to integrate these properties of sampling distribution into the schema constructed for statistical inference, which indicated only a procedural understanding of hypothesis testing and estimation. For this student, sampling distribution had no role in statistical inference.

6.3.1.2 Relating conceptual structure to understanding for Student G2S1

The following table gives the scores allocated to each of the tasks for student G2S1. The maximum score allocation for the task is also given in brackets.

Table 6.3 Tasks scores for student G2S1

Task	Score
Chi square	7 (9)
Correlation (a)	5 (5)
Modelling (b)	7 (7)
t-test	10 (10)
Modelling (a)	2 (7)
Sample size	2 (3)
Two group *t*-test	7 (8)
Explanation	6.5 (10)
Radio	2 (5)
Hospital	2 (2)
Confidence interval	0 (2)

Student G1S2 received a factor score of 0.3 for the procedural understanding measure, ranking her as 19th of the 23 students in the group. The student's result in the conceptual understanding measure was 0.2, slightly above average (ranking 8th of the 23 students in the group).

In order to investigate the nature of the relationship between the student's schema and conceptual understanding, student G1S2's responses to the tasks designed to measure conceptual understanding were analysed in detail. Her answers to the some of the tasks indicated knowledge of the sampling distribution. For example, G2S1 correctly answered the hospital question.

However, in other instances G2S1 demonstrated clear misconceptions. For example, in the Confidence Interval task (see section 4.8) where students were asked if the given confidence interval gave the same conclusion as the hypothesis test previously undertaken, G2S1 recalculated the confidence interval to see if the answer achieved was the same as the one given. Since they were comparable, she concluded that there was no contradiction. G2S1 did not conceptually connect the hypothesis test to the confidence interval, an action that was consistent with the form of the student's schema for statistical inference which showed no connection in the student's schema.

As also suggested by this student's schema for statistical inference, most of the explanations given in the Explain task conveyed procedural rather than conceptual understanding. For example, when asked to describe significance level in this task, by G2S1 responded as follows:

> When we do the calculation if it is below this set level it is said to be statistically significant so "Yes" it has had an effect of under this level. If above this Figure then it has not been effective, it is just due to sampling variability.

This interpretation of the significance level is essentially procedural, although sampling variability is mentioned, which is indicates some understanding of the relationship between the hypothesis testing procedure and the sampling process.

In the same task the P-value is described by G2S1 as follows:

> The P-value is derived from a Pearson's r table. Which shows the probability of getting this value from a population where there is no relationship. It is × 2 as it is a non-directional test. It has a low probability of 0.5% of just occurring by chance. Therefore it is then compared to the significance level 5%.

This explanation of P-value mixes a procedural description of how it is determined with a correct understanding of the meaning of P-value.

Later in this task, when explaining the Decision and Conclusion section, G2S1 states:

> ..and the decision then when compared to the α 5% is then to reject the null hypothesis as it has shown that there is a relationship between mothers and children in the general population.

Overall G2S1 has largely described the procedure of hypothesis testing. There is no clear link between this process of hypothesis testing and the process of sampling, but an understanding that the choice is between sampling variability or the existence of a relationship when carrying out the hypothesis test is clearly demonstrated.

The answer given by G2S1 to Radio task indicates that G2S1 has connected this scenario with her schema for hypothesis testing, correctly interpreting the situation with a hypothesised value of 0.2 (the radio station claim), and the observed sample statistic of 0.15. However, in order to apply the procedure of hypothesis testing with which this student feels confident, a sample size of 100 is assumed so as the question can be answered exactly. (The value of $n = 100$ is possibly used because the probabilities are stated as percentages.) G2S1 did not connect the scenario with possible sampling problems, again confirming that for this student there is only a tenuous link between sampling distribution and statistical inference.

Both the factor scores achieved, and the analysis of the responses given to the tasks by student G2S1 indicate that the level of conceptual understanding demonstrated is less than that demonstrated by the students allocated to Group 1. They confirm that for this student an essentially procedural schema for hypothesis testing and estimation is evoked by most situations relating to statistical inference. This is consistent with the prediction made on the basis of the form of the conceptual structure created by G2S1 as evidenced by the concept maps.

6.3.2 Student G2S2

Student G2S2 is a mature age female undertaking her first formal course of study in statistics. She also works full-time in the health science area. Prior to the teaching experiment she rated Low on the questionnaire which self-assessed prior knowledge. She achieved a score of 0.5 on the measure of procedural understanding, and –0.3 on the measure of conceptual understanding.

6.3.2.1 Concept map analysis for student G2S2

Map 1: Sampling distribution of the sample proportion

Suggested Terms:

centre, computer generated, constant, distribution, normal model, p, $\hat{p}$, population, population parameter, population proportion, repetitions, sample proportion, sample size, sample statistic, sample(s), sampling distribution, sampling variability, spread, variable.

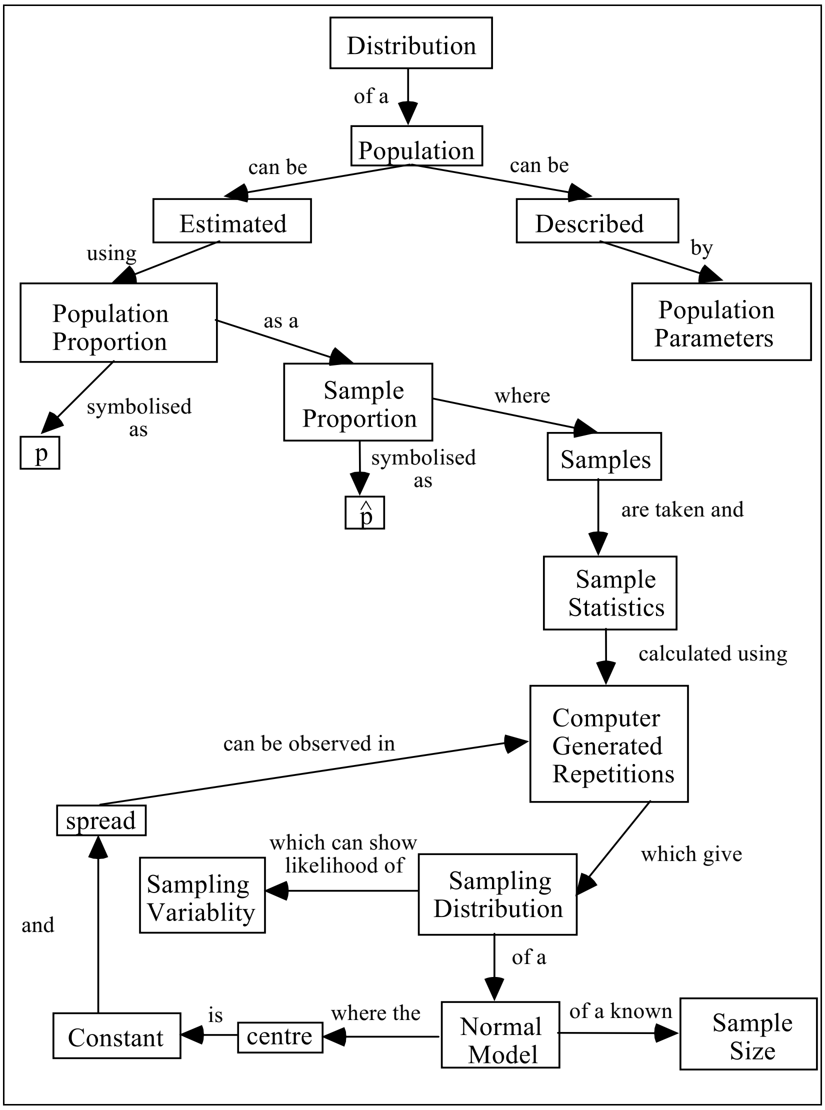


Figure 6.17 Map of sampling distribution of the sample proportion for student G2S2 (Group 2)

Description

In this concept map the sampling distribution was correctly identified as the distribution of the sample proportion, which was generated from repetitive sample using the computer, and a known value of the population proportion. The normal model was also correctly associated with the sampling distribution, and G2S2 recognised that the centre of the sampling distribution was constant (although not related this to the value of the population proportion) whilst the spread varied (although not related this to the sample size).

Comment

In summary, this student created a conceptual structure for sampling distribution of the sample proportion that included both the empirical and the formal/theoretical representations. Whilst there were some omissions, this map included six of the ten propositions identified in the expert concept map for the sampling distribution of the sample proportion.

Map 2: Sampling distribution of the sample mean

Suggested Terms:

centre, computer generated, constant, distribution, normal model, μ, $\bar{x}$, population, population mean, repetitions, sample mean, sample size, sample statistic, sample(s), sampling distribution, sampling variability, spread, variable.

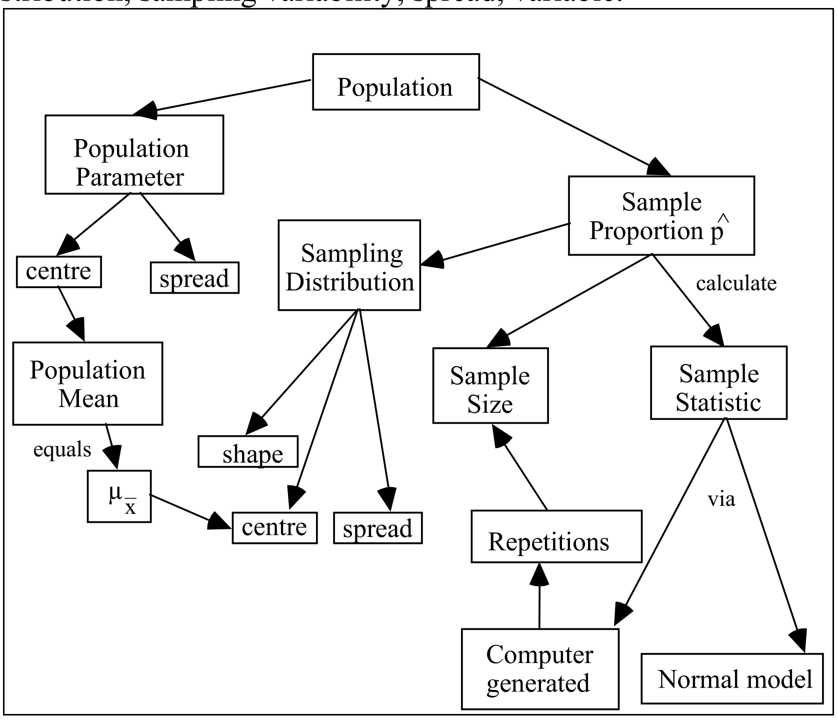


Figure 6.18 Map of sampling distribution of the sample mean for student G2S2 (Group 2)

Description

This map was less clearly labeled than the previous one, but many relationships can still be identified. The sampling distribution was again based on the sample statistic (although incorrectly labeled sample proportion instead of sample mean). Both the computer generated and normal model were given as alternatives, although the link from these to the sampling distribution was unclear. The sample size was again linked to the spread of the distribution, and the centre of the sampling distribution was correctly recognised as occurring at the population mean, a key proposition which was included by very few students.

Comment

This map shows an emerging schema for sampling distribution which shows understanding of the key ideas of sampling distribution in both the empirical and theoretical representations. The map includes six of the 10 key propositions identified in the expert map.

Map 3: Sampling distribution (general)

Suggested Terms: centre, constant, μ, *p*, $\hat{p}$, population, population parameter, *r*, ρ, sample size, sample statistic, sample(s), sampling distribution, spread, variable.

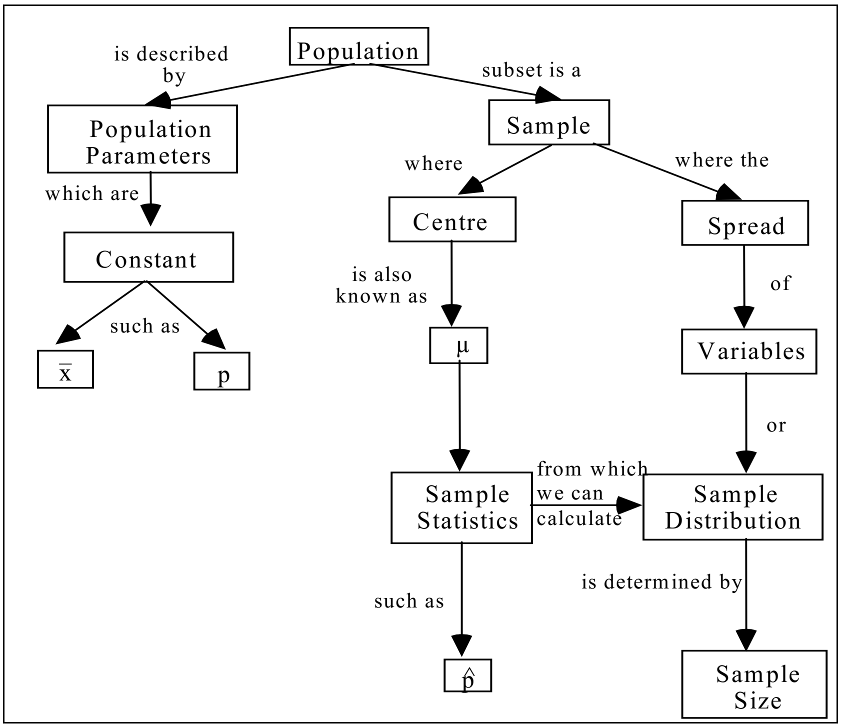


Figure 6.19 Map of sampling distribution (general) for student G2S2 (Group 2)

Description

In this map the symbols for the population mean (μ) and the sample mean ($\bar{x}$) were confused. The *sampling* distribution was also incorrectly named as the *sample* distribution. However, the sampling distribution seems to be what was implied here as it derived from the sample statistic, and was dependent on sample size.

Comment

Whilst retaining many of the key propositions when synthesising the general schema for sampling distribution, student G2S2 has made no reference to the empirical and the theoretical representations of sampling distribution when constructing this concept map. However, the schema for sampling distribution shown is indicative on the formation on many correct links, and includes seven of the ten propositions identified in the expert map.

Map 4: Hypothesis testing

Suggested terms: alternate, decision, hypotheses, null, P-value, population, population parameter, sample statistic, sample(s), sampling distribution, sampling variability, significance level, test statistic.

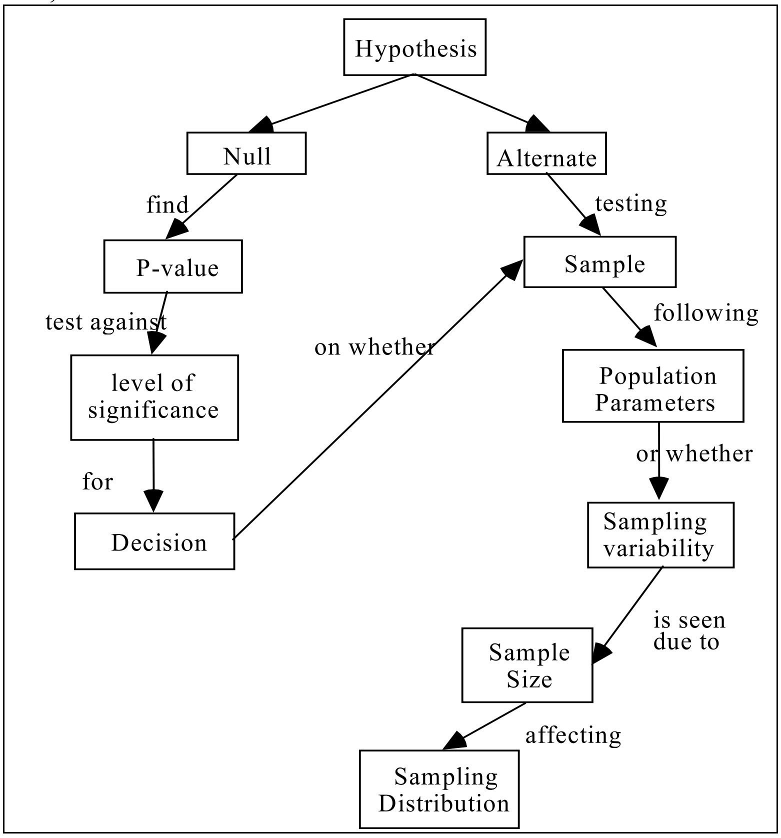


Figure 6.20 Map of hypothesis testing for student G2S2 (Group 2)

Description

This map essentially describes the process of hypothesis testing, with a step by step describes of how one would go about carrying out the hypothesis test. However, the map also indicates clearly that the student G2S2 understands that the sampling distribution played a role in determining whether the difference between the observed sample statistic and the hypothesised population parameter could be reasonably attributed to sampling variability or not. G2S2 also noted that this decision would be affected by the size of the sample used.

Comment

Whilst the sampling distribution has not been identified in this map as an important component in the determination of the P-value, there is a recognition of sampling variability, which is summarised by the sampling distribution, as the basis for the decision in an hypothesis test. However, map is basically a recipe for hypothesis testing, and includes only two of the 14 propositions identified in the expert map.

Map 5: Estimation

Suggested terms

confidence interval, estimation, interval estimates, point estimates, population, sample, sample statistics, sampling distribution,

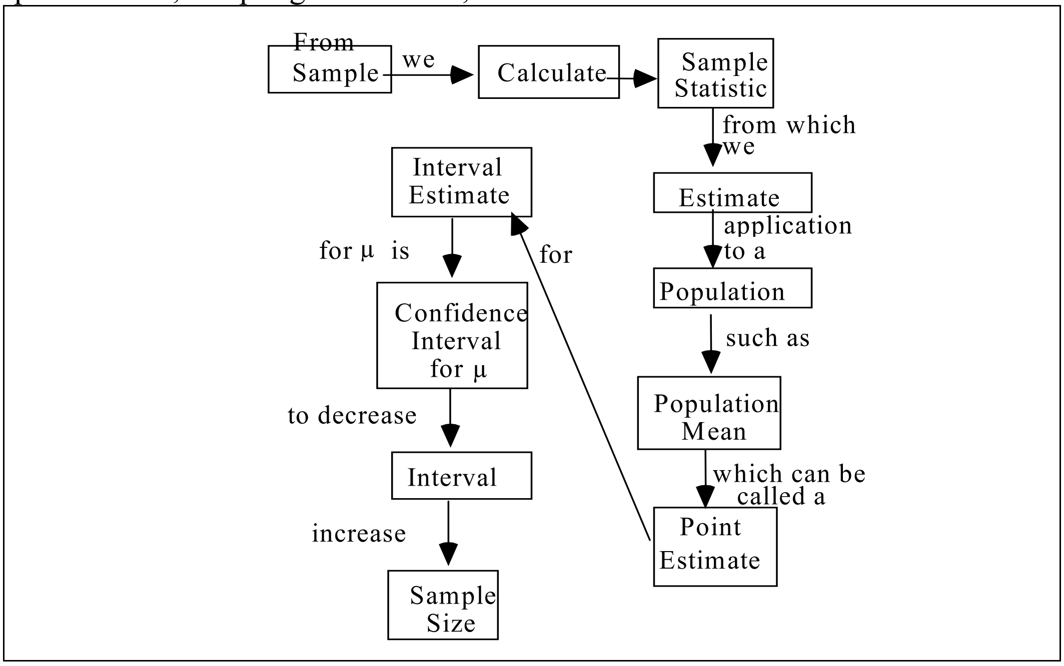


Figure 6.21 Map of estimation for student G2S2 (Group 2)

Description

This map is again quite linear with few cross-links, suggesting knowledge of the process of estimation. Even though the term *sampling distribution* was included on the list of suggested terms, and has been included on the previous maps, it has been excluded here suggesting that G2S2 identified no role for the sampling distribution in estimation. G2S2 has included, however, the effect of the sample size on the width of the confidence interval, but has not related this to the spread of the sampling distribution. The map includes only three of the 10 key propositions identified in the expert map.

Map 6: Statistical inference

Suggested terms

confidence interval, decision & conclusion, estimation, hypotheses, hypothesis testing, inferential statistics, interval estimates, P-value, point estimates, population, sample, sample statistics, sampling distribution, significance level, statistical significance, test statistic.

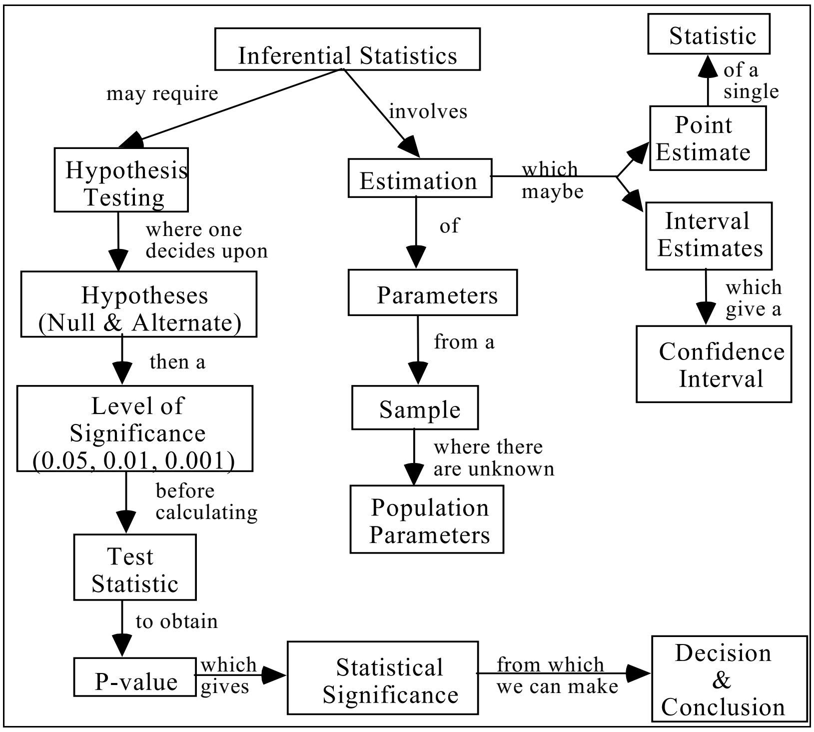


Figure 6.22 Map of Statistical Inference for student G2S2 (Group 2)

Description

In this map the hypothesis testing procedure is shown by a flow chart indicating a sequence of steps to be followed. This student's schema indicates that she knows how to carry out an hypothesis test. Confidence intervals are identified as interval estimates of parameters, but no further information is given concerning how they are determined or interpreted. There is no connection made between hypothesis testing and estimation.

Comment

In this student's schema neither aspect of inference, hypothesis testing nor estimation, has been linked to the sampling distribution, which has again been completely excluded. In fact, the entire process of sampling has been omitted, and there is no indication that the hypothesis testing procedure involves making a decision about a population based on sample information. This map included three of the 14 propositions identified n the expert map.

Summary of schema development for student G2S2

The concept maps for sampling distribution constructed by student G2S2 were complex, included many correct propositions. In particular the sampling distribution was correctly identified as summarising the sample statistic, and both the empirical and theoretical representations of sampling distributions were included in the early maps. However both aspects of statistical inference, hypothesis testing and estimation, were represented quite procedurally in the later maps, with no link made to sampling distribution.

6.3.2.2 Relating conceptual structure to understanding for student G2S2

Table 6.4 gives the scores allocated to each of the tasks for student G2S2. The maximum score allocation for the task is also given in brackets.

Student G2S2 received a factor score of 0.5 for the procedural understanding measure, a result slightly above the average (ranking 9th of the 23 students in the group). Her result in the conceptual understanding measure was -0.3, slightly below average (ranking 14th of the 23 students in the group).

Table 6.4 Tasks scores for student G2S2

Task	Score
Chi square	9 (9)
Correlation (a)	5 (5)
Modelling (b)	7 (7)
t-test	8 (10)
Modelling (a)	5 (7)
Sample size	3 (3)
Two group *t*-test	8 (8)
Explanation	4.5 (10)
Radio	2 (5)
Hospital	0 (2)
Confidence interval	2 (2)

Some insight into the nature of G2S2's understanding of statistical inference can be gained by examination of her answers for the tasks designed to measure conceptual understanding. Although the relationship between hypothesis testing and confidence intervals was not explicitly recognised in the concept map for statistical inference, student G2S2 recognised this relationship in Confidence Interval task, with the following answer:

> The confidence interval does not include the μ of 3.6 & is therefore consistent with the hypothesis that the sample mean has decreased.

When answering the Explain task, Student G2S2 gave the following explanation of the level of significance:

> If a significance level less than the specified (in this case 0.05) is calculated then the difference between the population parameter and the sample statistic is unlikely to be due to sampling variability alone. If a significance level of less than 0.01 is chosen, a P-value less than this indicates the result is strongly unlikely to be due to sampling variability.

Once again, the student G2S2 is clear that the purpose of the test is to determine whether or not sampling variability is a reasonable explanation for the difference between the sample statistic and the hypothesised population parameter. However, the link between this and the significance level is not clear from this explanation.
The P-value step of the Explain task is explained by G2S2 as follows:

> As this is a non-directional or 2-tail test, we multiply the Figure obtained by 2 to get a probability value. This is compared with the level of significance set earlier.

This is a procedural explanation of the P-value, giving no information concerning the meaning of the P-value, or the role of the sampling distribution in its determination.

The Decision and Conclusion explanation given by G2S2 is as follows:

> The P-value was less than the level of significance so the alternate hypothesis is not disproved. The null hypothesis is rejected as the P-value indicates that the appearance of a relationship is not due to sampling variability alone

This logic is a little confused, but an essentially correct conclusion is reached. The link between the hypothesis testing process and the notion of sampling variability does seem to have been recognised by G2S2 at a conceptual level, but exactly how this is determined by the sampling distribution is vague.

The answer given by G2S2 to the Radio task was as follows:

> Sample size may be different (ie too small). Radio station's sample may not have been random: eg those who ring their request line, those at a particular entertainment.
> Not necessarily deliberately misleading, may be a result of poor survey methods.

Interestingly, G2S2 does not associate the difference between the two values with sampling variability at all overtly, although by noting that the sample sizes may have been different G2S2 implies some appreciation of the role of sampling variability. The difference is seen more directly as explainable in terms of data collection problems. The link here has been to sampling, with no clear link to the sampling distribution and hypothesis testing.

Both the factor scores achieved, and the analysis of the responses given to the tasks by student G2S2 indicate that the level of conceptual understanding demonstrated is in general terms less than that demonstrated by the students allocated to Group 1. The analysis shows that this student seems to have connected statistical inference to sampling variability, but is not exactly sure as to the source and meaning of the connection (via the sampling distribution). Hence, questions of interpretation are often resolved by G2S2 by resorting to a procedural description.

6.4 Analysis of Group 3 Students

Students who did not at any stage show evidence of the development the concept of sampling distribution were allocated to Group 3. Two students from this group have been selected at random for further analysis.

6.4.1 Student G3S1

Student G3S1 is a mature-age female undertaking her first formal course of study in statistics. She works full-time in the health science area. Prior to the teaching experiment she rated herself as Medium on the prior knowledge questionnaire. She achieved a factor score of 0.2 on the measure of procedural understanding, and –0.4 on the measure of conceptual understanding.

6.4.1.1 Concept map analysis for student G3S1

Map 1: Sampling distribution of the sample proportion

Suggested Terms: centre, computer generated, constant, distribution, normal model, *p*, $\hat{p}$, population, population parameter, population proportion, repetitions, sample proportion, sample size, sample statistic, sample(s), sampling distribution, sampling variability, spread, variable.

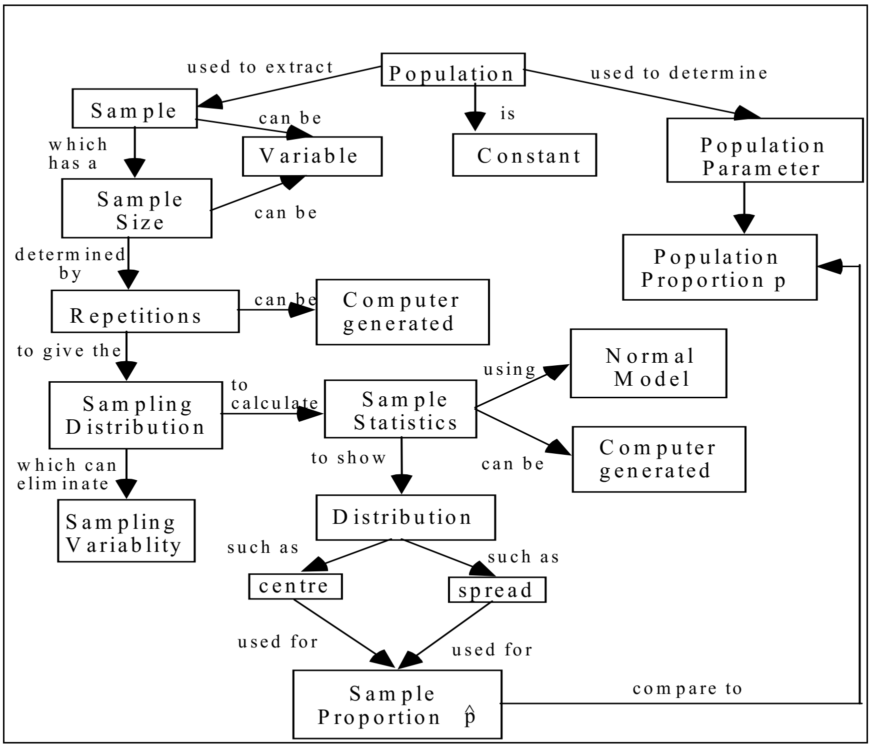


Figure 6.23 Map of sampling distribution of the sample proportion for student G3S1 (Group 3)

The schema represented by the map shown in Figure 6.23 included many important concepts, such as the variability of the sample and the stability of the population. Also shown is the link between the population parameter and the sample statistic. However, it can be seem from study of the map that the sample statistics were seen as descriptive of the sampling distribution, and not the other way around. Whilst both *computer generated* and *normal model* were mentioned, these were related to the sample statistic.

Comment

There is no clear indication from this map that student G3S1 has recognised the variability of the sample statistic, a concept that necessarily forms the foundation of a schema for sampling distribution. The map includes 4 of the ten key propositions identified in the expert map.

Map 2: Sampling distribution of the sample mean

Suggested Terms: centre, computer generated, constant, distribution, normal model, μ, $\bar{x}$, population, population mean, repetitions, sample mean, sample size, sample statistic, sample(s), sampling distribution, sampling variability, spread, variable.

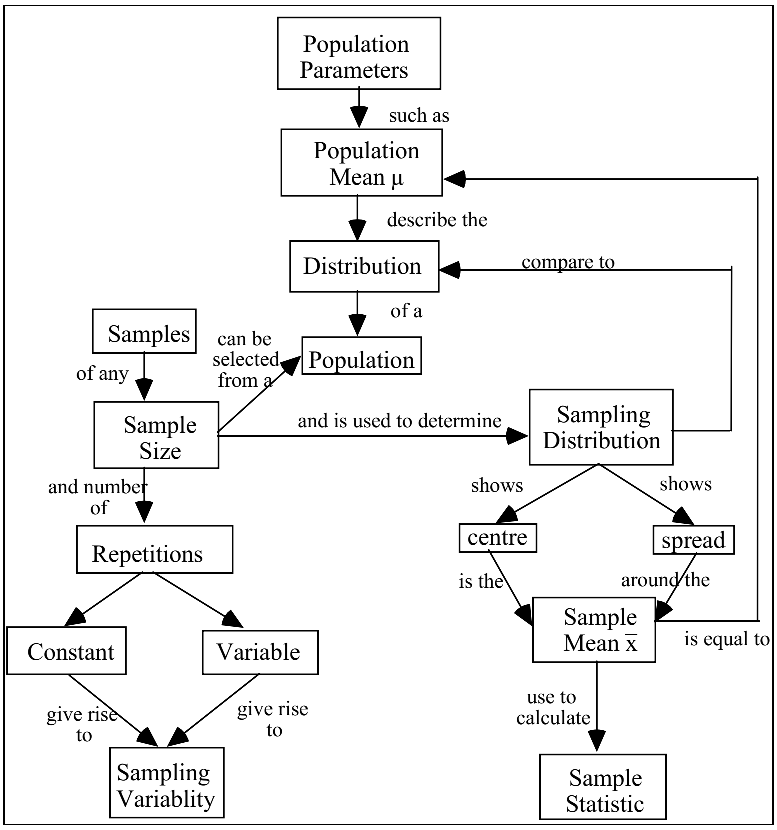


Figure 6.24 Map of sampling distribution of the sample mean for student G3S1 (Group 3)

Description

As in the previous map, the sampling distribution was again described as the source of the sample statistic, with the centre and spread describing the distribution of the sample, and this distribution then compared to the distribution of the population. One would expect, of course, that the distribution of the sample would generally be similar to the distribution of the population, whereas the sampling distribution may be completely different. Neither the normal model nor the computer-generated representations of the sampling distribution were mentioned in this map.

Comment

Again, from this map it appeared that Student G3S1 had not yet recognised the variability of the sample statistic in her schema. The map includes two of the 10 key propositions identified in the expert map.

Map 3: Sampling distribution (general)

Suggested Terms: centre, constant, μ, p, $\hat{p}$, population, population parameter, r, ρ, sample size, sample statistic, sample(s), sampling distribution, spread, variable.

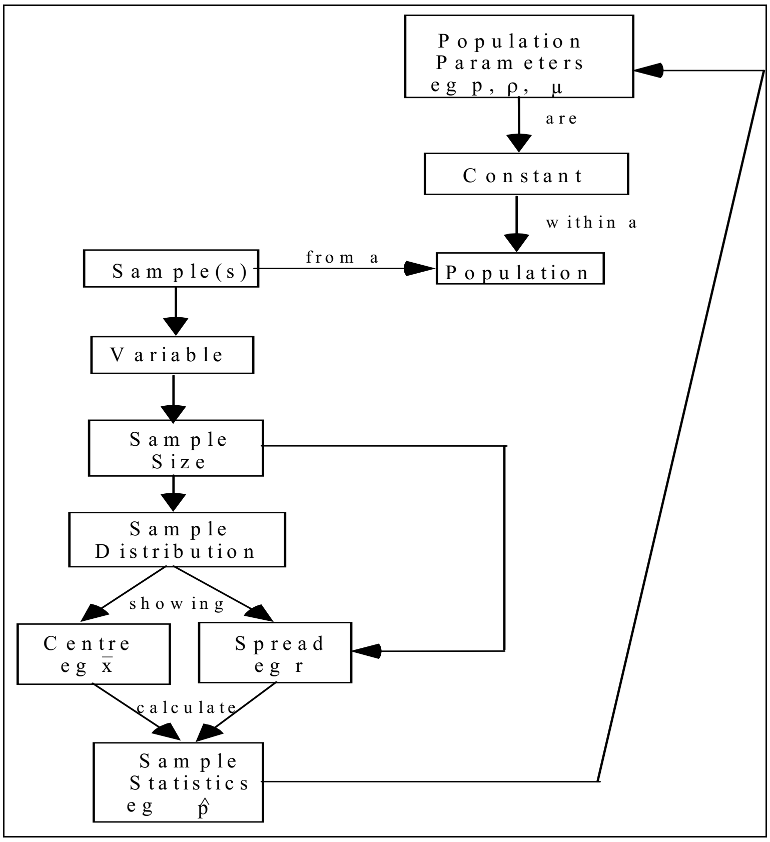


Figure 6.25 Map of sampling distribution (general) for student G3S1 (Group 3)

Description

The map constructed includes several correct propositions relating samples and populations, statistics and parameters. Correct notation has also been used in most instances. The propositions shown in this map again suggested that the schema constructed by student G3S1 had identified sampling distribution as the distribution of the sample. In fact, in this map the term *sampling* distribution was modified to *sample* distribution, which was exactly what G3S1 was describing.

Comment

There is no evidence of the formation of a schema in which sampling distribution is recognised as the distribution of the sample statistic in this map. The map includes three of the 10 propositions identified in the expert map.

Map 4: Hypothesis testing

Suggested terms: alternate, decision, hypotheses, null, P-value, population, population parameter, sample statistic, sample(s), sampling distribution, sampling variability, significance level, test statistic.

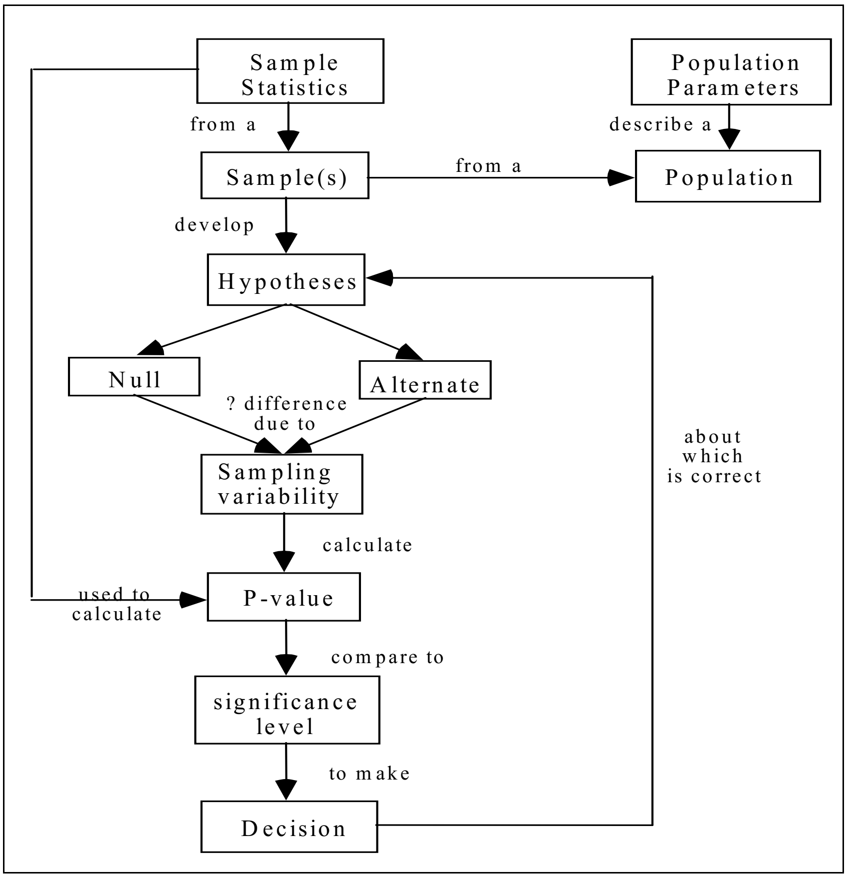


Figure 6.26 Map of hypothesis testing for student G3S1 (Group 3)

Description

Almost all of the valid propositions included in the schema represented by this map are concerned with the procedure of hypothesis testing, and the map resembles a set of steps to be followed when carrying out this procedure. Student G3S1 did correctly pose the question that the choice between the null and alternate hypotheses depended on the notion of *sampling variability*. Although given in the suggested list of terms to be included, the sampling distribution was omitted.

Comment

The lack of a role for sampling distribution in hypothesis testing was predictable since, for this student, the sampling distribution was the distribution describing the sample. As such, it had no clear role in the hypothesis testing procedure. This map included only four of the 14 key propositions identified in the expert map.

Map 5: Estimation

Suggested terms

confidence interval, estimation, interval estimates, point estimates, population, sample, sample statistics, sampling distribution.

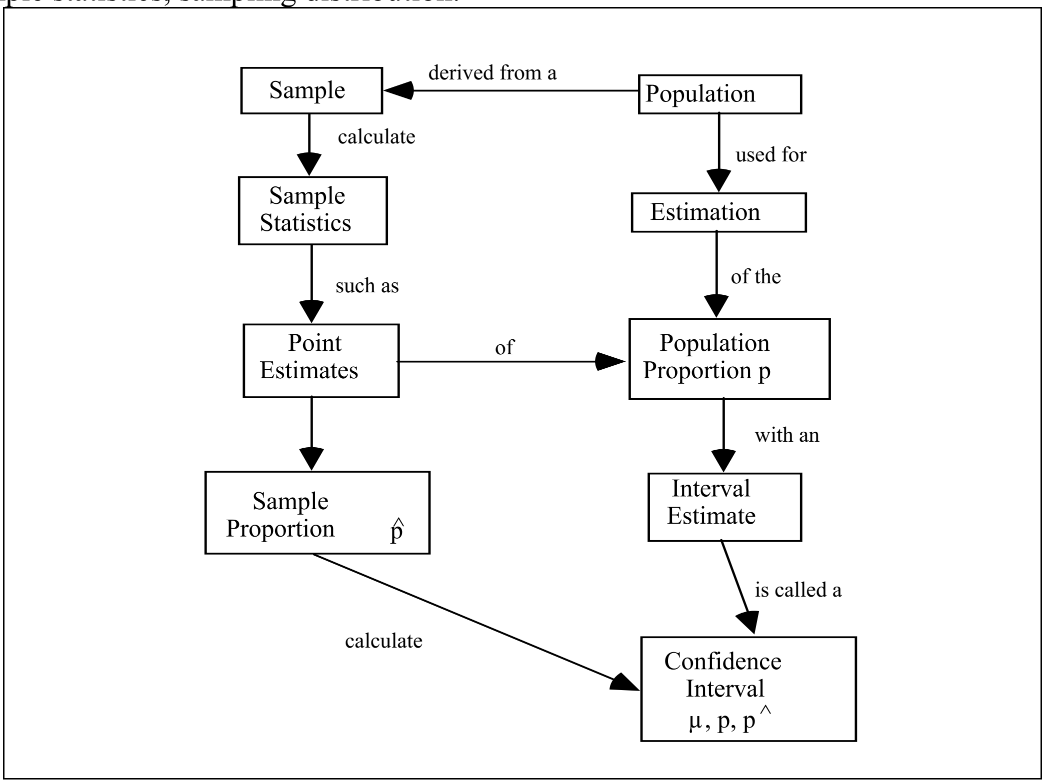


Figure 6.27 Map of estimation for student G3S1 (Group 3)

Description

Again, the map correctly relates samples and populations, statistics and parameters, although there are some problems with notation. And, once again, although on the suggested list of terms, the sampling distribution was omitted.

Comment

The sampling distribution, as this student understands it, had no role in estimation.

Map 6: Statistical inference

Suggested terms

confidence interval, decision & conclusion, estimation, hypotheses, hypothesis testing, inferential statistics, interval estimates, P-value, point estimates, population, sample, sample statistics, sampling distribution, significance level, statistical significance, test statistic.

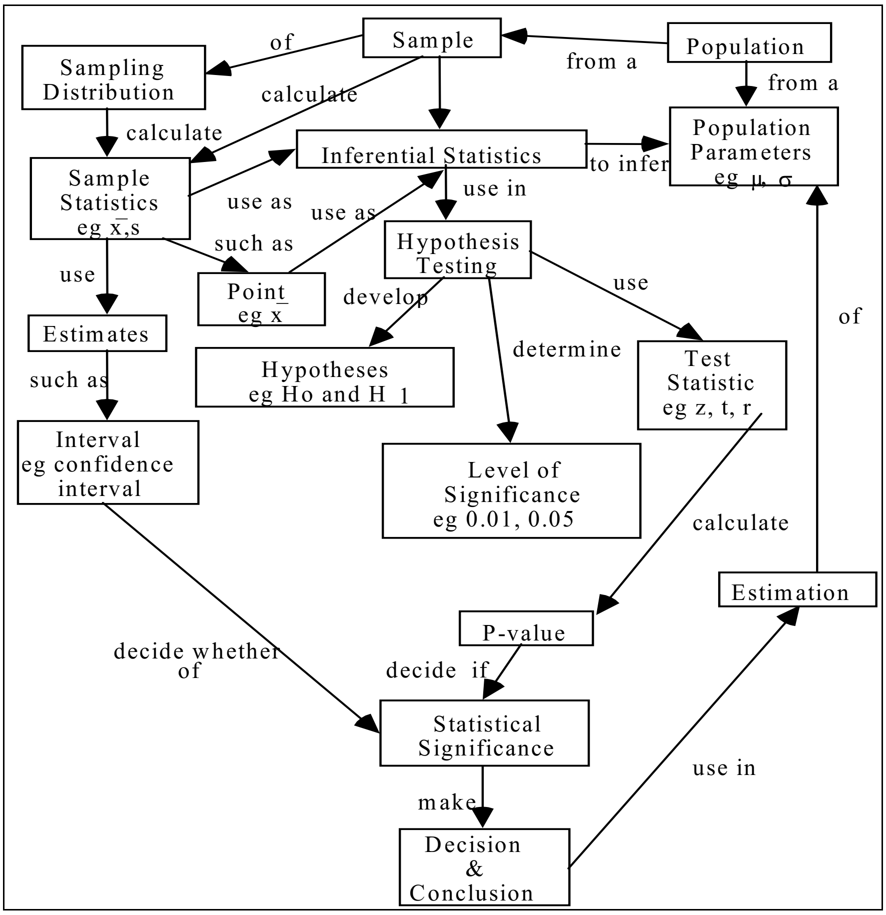


Figure 6.28 Map of Statistical Inference for student G3S1 (Group 3)

Description

There were again many valid propositions included in this map, particularly those relating samples and populations. Correct notation was also used throughout the map. For the first time since the study of inference began, the sampling distribution was included in the map. However, the sampling distribution is still understood by this student as a summary of the sample, leading to the calculation of sample statistics.

Comment

There is still no evidence that this student has connected the sampling distribution to the distribution of the sample statistic. The map includes five of the 14 propositions identified in the expert map.

Summary of schema development of G3S1

The conceptualization of the sampling distribution was as the distribution of the sample in all of the concept maps constructed by this student in which sampling distribution was mentioned. As such, there was no place for the sampling distribution in statistical inference, and consequently there was no link from either hypothesis testing or estimation to the sampling distribution in any of the maps concerned with inference. There was also no link between hypotheses testing and estimation.

6.4.1.2 Relating conceptual structure to understanding for student G3S1

The following table gives the scores allocated to each of the tasks for student G3S1. The maximum score allocation for the task is also given in brackets.

Table 6.5 Tasks scores for student G3S1

Task	Score
Chi square	9 (9)
Correlation (a)	5 (5)
Modelling (b)	6 (7)
t-test	10 (10)
Modelling (a)	5 (7)
Sample size	2 (3)
Two group *t*-test	8 (8)
Explanation	2 (10)
Radio	5 (5)
Hospital	0 (2)
Confidence interval	0 (2)

Student G2S2 received a factor score of 0.2 for the procedural understanding measure, a result close to the average for the group (ranking 16th of the 23 students in the group). The student's result in the conceptual understanding measure was -.4, below average for the group (ranking 17th of the 23 students in the group).

As can be seen from Table 6.5, student G3S1 scored quite poorly on all of the tasks measuring conceptual understanding, except the Radio task.

In Confidence Interval task G3S1 did recognise the relationship between the confidence interval and the hypothesis test, but rather noted that there was no inconsistency between the two results since the value of the sample mean was in the middle of the confidence interval.

In Explain task, where the student is required to interpret the hypothesis test in his or her own words, the significance level was described as follows:

> Not stated in documentation so default significance level of 5% $\alpha = 0.05$ used. If the result found is < 0.05 it is statistically significant.

This is a procedural definition of the significance level, giving no insight at all into what it means.

When describing the P-value G3S1 says:

> Calculate P-value using value of r and number of pairs of data values using Pearson's r table. The P-value must be doubled as this hypothesis test is non-directional.

Once again, a procedural explanation, giving no information about the way in which the P-value should be interpreted, or the relationship between the P-value and the sampling process.

Finally, in the Decision and Conclusion step, G3S1 wrote:

> Compare calculated P-value to stated level of significance. If P-value was > 0.05 we would not conclude that a relationship existed between mothers and their children. However as our calculated P-value is less than 0.05 we reject our null hypothesis and conclude that there is a relationship between the intelligence of children and their mothers.

This explanation of the conclusion of the hypothesis test merely describes in words the process which is being carried out, again showing little conceptual understanding of the procedure.

In contrast to Explain task, G3S1 gave an answer to the Radio task that included both sampling problems and issues of sampling variability as possible explanations. The need to know the sample size to make a decision concerning the claim was also recognised. The completeness of this answer seems to contradict the confusion concerning sampling distribution that was suggested by this student's concept maps. However, in the earlier maps constructed by G3S1 is was clear that this student had established a conceptual structure for the sampling process which included recognition of the variability of samples, and this may be the basis of the answer given here. On further investigation of the student's background it was also revealed that G3S1 worked in an area which was concerned with the collection of survey data, and thus the scenario posed was possibly similar to one with which G3S1 was already familiar.

Consideration of the factor scores achieved, and analysis of the responses given to the tasks by student G3S1 indicate that the level of conceptual understanding demonstrated is in general terms less than that demonstrated by the students allocated to both Groups 1 and 2. The analysis again shows a possible connection between statistical inference and the sampling process. However, at no stage has G3S1 indicated that she knows what is meant by a sampling distribution, and her answers are in the most part descriptions of the procedures involved in statistical inference.

6.4.2 Student G3S2

Student G3S2 is a mature-age female who has studied statistic in previous courses. She rated herself as High on the self-assessed measure of prior knowledge. She works part-time in the market research area. On the procedural measure she achieved a score of 0.5, and on the conceptual measure a score of –1.6.

6.4.2.1 Concept map analysis for student G3S2

Map 1: Sampling distribution of the sample proportion

Suggested Terms:

centre, computer generated, constant, distribution, normal model, p, $\hat{p}$, population, population parameter, population proportion, repetitions, sample proportion, sample size, sample statistic, sample(s), sampling distribution, sampling variability, spread, variable.

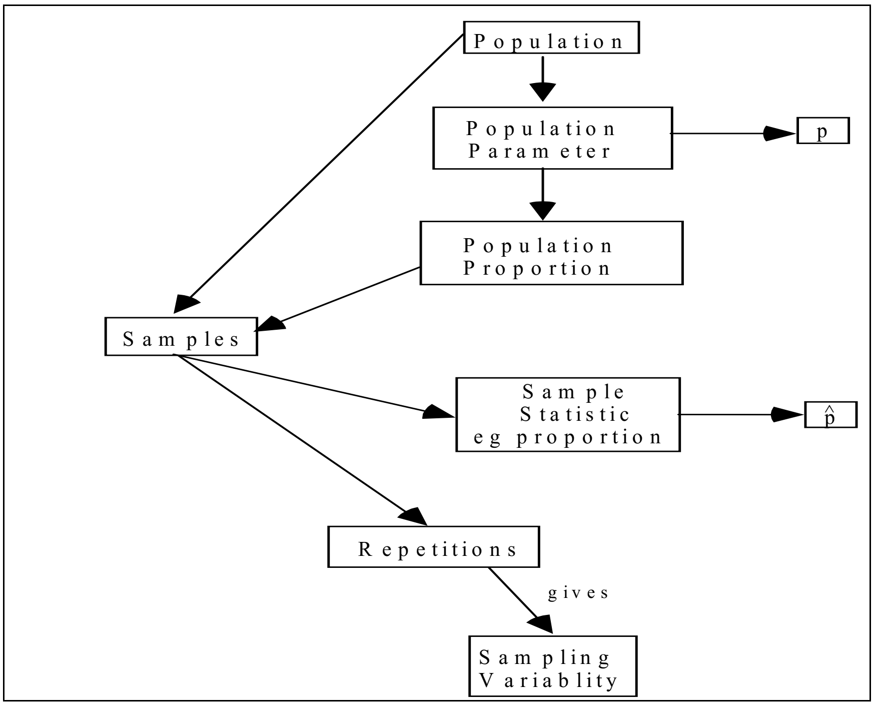


Figure 6.29 Map of the sampling distribution of the sample proportion for G3S2 (Group 3)

Description

The distinction between populations and samples was clearly shown in this concept map, and correct notation was used. The map also indicated that student G3S2 recognised that repetitive sampling led to variable samples. Even though sampling distribution is the subject of the map, it has been left out.

Comment

At this stage the student does not seem to have constructed a schema for the sampling distribution of the sample proportion. The map includes three of the 10 propositions identified in the expert map.

Map 2: Sampling distribution (general)

Suggested Terms:

centre, constant, μ, p, $\hat{p}$, population, population parameter, r, ρ, sample size, sample statistic, sample(s), sampling distribution, spread, variable.

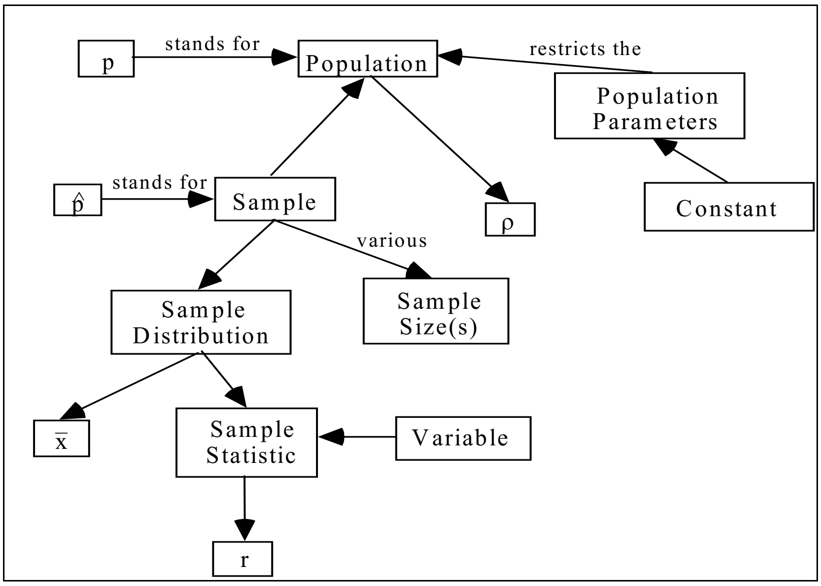


Figure 6.30 Map of the sampling distribution (general) for G3S2 (Group 3)

Description

Once again, student G3S2 included propositions indicating that she realised the distinction between populations and samples, and showed correct use of notation. Population parameters were seen to be constant, and sample statistic as variable. In this map *sample* distribution was mentioned, correctly, as the distribution of the sample.

Comment

Again, there was no mention of a distribution describing the sample statistic (the sampling distribution), indicating that this student had not formed a schema for sampling distribution. The map included 4 of the ten propositions identified on the expert map.

Map 3: Hypothesis testing

Suggested terms

alternate, decision, hypotheses, null, P-value, population, population parameter, sample statistic, sample(s), sampling distribution, sampling variability, significance level, test statistic.

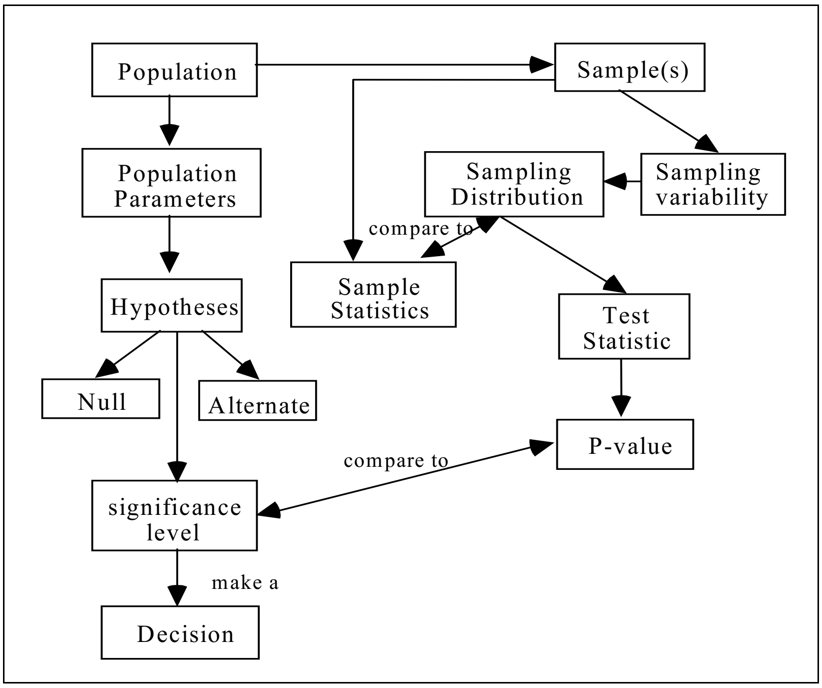


Figure 6.31 Map of hypothesis testing for student G3S2 (Group 3)

Description

This map was again rather like a flow chart for hypothesis testing. This was the first map prepared by this student in which the sampling distribution was mentioned, in reference to the sampling variability of the sample. The sample statistic was compared to the sampling distribution, and the test statistic was also originating from the sampling distribution. Whether this student had conceptualised the sampling distribution as summarising the sample or the sample statistic was not clear from this map, and the role proposed for the sampling distribution in hypothesis testing by this student was also unclear.

Comment

It is possible that student G3S1 had begun to realise at this stage that the sampling distribution referred to the distribution of the sample statistic, and had a role in the process of hypothesis testing. The map included five of the 14 propositions identified from the expert map.

Map 4: Estimation

Suggested terms

confidence interval, estimation, interval estimates, point estimates, population, sample, sample statistics, sampling distribution.

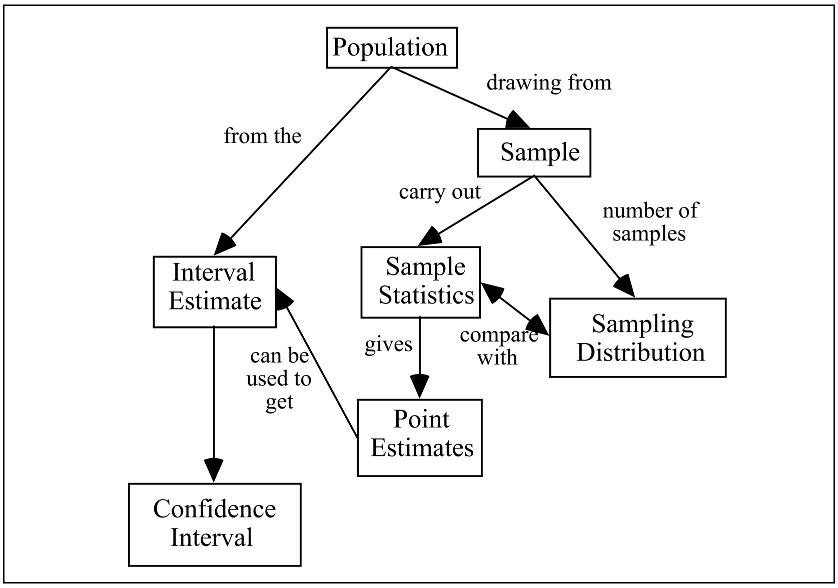


Figure 6.32 Map of estimation for student G3S2 (Group 3)

Description

The sampling distribution was again mentioned in this map for estimation, and was seen as being derived from repetitive sampling (the empirical representation). However, once again it was not clearly identified as the distribution of the sample statistic. Whilst linked to the sample statistic, there was no link between the sampling distribution and the process of estimation.

Comment

Once again, it is possible that the student G3S2 is struggling with the concept of sampling distribution and its role in estimation. However, even if the student has now realised that the sampling distribution is the distribution of the sample statistic, there is no link between the sampling distribution and the determination of the confidence interval. The map included 4 of the 10 propositions identified in the expert map.

Map 5: Statistical inference

Suggested terms

confidence interval, decision & conclusion, estimation, hypotheses, hypothesis testing, inferential statistics, interval estimates, P-value, point estimates, population, sample, sample statistics, sampling distribution, significance level, statistical significance, test statistic.

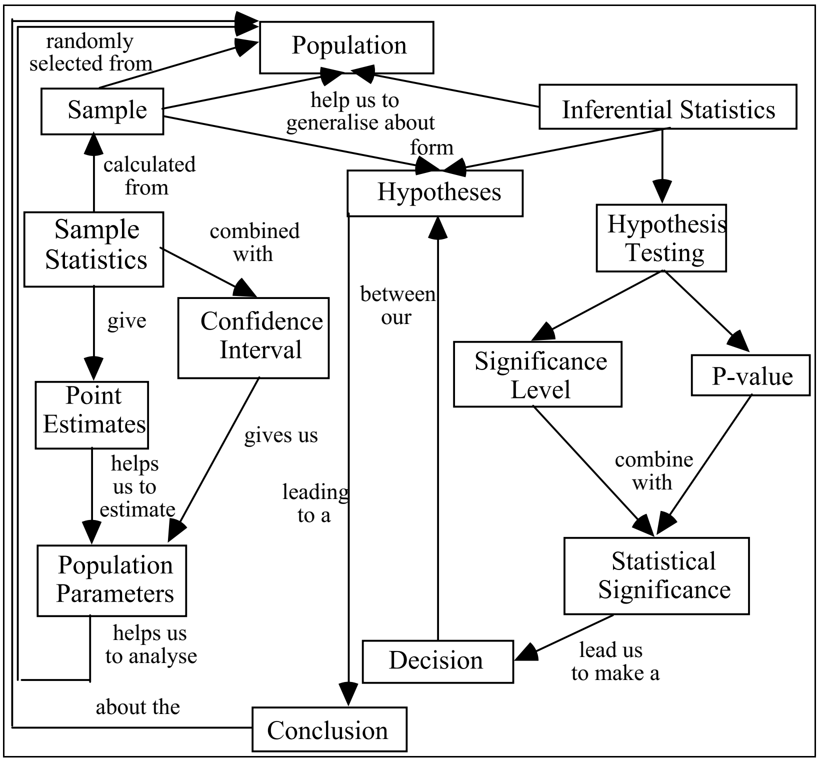


Figure 6.33 Map of Statistical Inference for student G3S2 (Group 3)

Description

The map included many valid propositions that describe the procedure of hypothesis testing, once again reminiscent of a flow chart. There is no mention of the sampling distribution, and in fact neither the sample statistic nor the test statistic is mentioned in relation to hypothesis testing.

Comment

As in the earlier maps, the sampling distribution has been left out from this map. It includes five of the 14 propositions identified in the expert map.

Summary of schema development for student G3S2

In the early maps constructed by this student there was no evidence of a schema in which the sampling distribution is understood to be the distribution of the sample statistic in either the empirical or the formal/theoretical representation. Consequently, it was not possible for the student to form links between sampling distribution and either hypothesis testing or estimation.

6.4.2.2 Relating conceptual structure to understanding for student G3S2

The following table gives the scores allocated to each of the tasks for student G3S1. The maximum score allocation for the task is also given in brackets.

Table 6.6 Tasks scores for student G3S2

Task	Score
Chi square	9 (9)
Correlation (a)	5 (5)
Modelling (b)	7 (7)
t-test	9 (10)
Modelling (a)	5 (7)
Sample size	3 (3)
Two group *t*-test	8 (8)
Explanation	2 (10)
Radio	1 (5)
Hospital	0 (2)
Confidence interval	0 (2)

Student G3S1 received a factor score of 0.5 for the procedural understanding measure, a result above average for the group (ranking 7th of the 23 students in the group). The student's result in the conceptual understanding measure was -1.6, well below average for the group (ranking 22nd of the 23 students in the group).

In the Confidence Interval task the equivalence of the confidence interval and the conclusion previously reached was not recognised, with the G3S2 stating:

> Yes, it is consistent - our value 2.54 lies between the confidence interval (1.626,3.465)

That is, student G3S2 found no inconsistency because the sample mean lies within the confidence interval.

The significance level is described in the Explain task as follows:

> The significance level indicates at what point our results are statistically significant.

This is merely a restatement of the term, and gives no insight whatsoever into the meaning of the concept.

In the same task student G3S2 explained the P-value as follows:

> The P-value is compared with the significance level to determine which hypothesis we use
> ie if $P<\alpha$, we reject H_0
> if $P>\alpha$, we accept H_0

Once again this is a procedural description, indicating how the P-value is used instead of what it means.

Finally, under Decision and Conclusion G3S2 wrote:

> In this case our P-value was smaller than our significance level, so we reject our null hypothesis and accept our alternate hypothesis.

The outcome of the hypothesis test has not been interpreted here, merely restated in terms of the action that would be taken.

The Radio task is recognised by G3S2 as a hypothesis test for a proportion. No hypotheses are explicitly stated, and a sample size is taken as 100 so that the problem can be completed. The conclusion to the test is as follows:

> The z-score value of 0.1056 at a significance level of 0.05, shows that the radio station is misleading its advertisers.

The P-value is mistakenly labeled as a z-score, and the conclusion is incorrect. This answer indicates that the student recognised that the scenario given could be interpreted as a hypothesis test, but there is no link between this hypothesis testing procedure and the sampling process for this student.

Consideration of the factor score achieved and analysis of the tasks undertaken by student G3S2 indicate that the level of conceptual understanding demonstrated by her is less than that demonstrated by the students allocated to Groups 1 and 2. Her interpretations have on every occasion been limited to the procedures of statistical inference. These results are consistent with the prediction made on the basis of the form of the schema for statistical inference created by G3S2 and evidenced by the concept maps.

6.5 Conclusion

In this chapter the series of concept maps constructed by six students, two from each of the categories described in Chapter 5, were analysed in detail. The purpose of the analysis was to document the form of the schema for sampling distribution constructed by each student, and further to investigate the nature of any connections made by the students between their schema for sampling distribution and that for statistical inference.

The theoretical analyses undertaken prior to the conduct of the teaching experiment which was the basis of this study suggested that, in order to develop a high level of conceptual understanding in statistical inference, the student's schema for statistical inference should include links to the sampling distribution, in particular in its empirical representation. Since the empirical representation for the sampling distribution is based on the sampling process, then the existence of such connections in the student's schema provide a means by which the student is able to interpret hypothesis testing and estimation in terms of samples and the sampling process. Thus, examination of the students' schemas should provide evidence that form of the conceptual structure is related to student performance on the measures of conceptual understanding. Since procedural understanding is not by definition associated with connected networks in the students' mental structures (Hiebert & Carpenter, 1992), then it was considered possible for the students to exhibit high levels of procedural understanding whatever the form of their conceptual structure.

The case study analysis carried out in this chapter has shown that on the basis of the students' schemas for sampling distribution and statistical inference, it is possible to predict which students will demonstrate higher levels of conceptual understanding and which students will exhibit lower levels. Those students who have connected statistical inference to the sampling distribution have demonstrated the ability to explain and interpret statistical inference. Those students who have not tend to understand statistical inference as a procedure, and in most instances are unable to related this procedure to the sampling process. However, no relationship was observed between the form of the students' schema for statistical inference and their demonstrated level of procedural understanding.

The factor scores on each of the factors procedural understanding and conceptual understanding calculated using the coefficients arising from the factor analysis described in chapter 5 are given in Table 6.7. Note once again that factor scores are standardised with a mean of zero and a standard deviation of 1.

Table 6.7 Procedural and Conceptual Factor Scores for Case Study Students

Student	Factor Procedural	Factor Conceptual
Group 1		
G1S1	.3798	2.3053
G1S2	.4735	.8143
Group 2		
G2S1	-.3091	.1643
G2S2	.4725	-.2525
Group 3		
G3S1	.2437	-.3672
G3S2	.4824	-1.6015

From this table it can be seen that the factor scores for procedural understanding are similar for all three groups, with the lowest score occurring in Group 2, and the highest score in Group 3, indicating that there is no obvious relationship between the procedural understanding score and the group allocation. There is, however, a clear pattern to factor scores for the measure of conceptual understanding, with the highest scores achieved by the Group 1 members, and the lowest scores achieved by the Group 3 members. The results in Table 6.7 again affirm the relationship between the form of the students' schemas and conceptual understanding, and lack of relationship between the form of the students' schemas and procedural understanding.

Thus, the analyses undertaken in this chapter have confirmed and informed the group level findings reported in Chapter 5.

Chapter 7

Implications for Teaching Practice

7.1 Introduction

The motivation for the conduct of this research came from a concern shared by the researcher and other statistics educators (for example: Garfield & Ahlgren, 1988; Moore, 1992a) that the teaching and learning of statistics was at a crossroads. There had been a sudden expansion of the teaching of statistics at all educational levels, no doubt due in part to the availability of cheap technology with which one could carry out sophisticated statistical analyses. However, many felt that, while increasing numbers of students studied statistics, the number of students who gained a real appreciation of the power and purpose of statistics was extremely small (for example Konald, 1991; Williams, 1996). And, there were few signs that increasingly refined technological support was doing much to change this (for example Weldon, 1986; 1983Hawkins, 1988; Cohen & Chechile, 1997). More research was needed in order to determine how to better structure and teach introductory statistical inference.

The research carried out in this study has confirmed the following hypothesis:

> Students whose schema for sampling distribution demonstrates links to the sampling process and whose schema for statistical inference includes links to the sampling distribution, will show evidence of both conceptual and procedural understanding, whilst those who have not integrated these concepts into their cognitive structure may still demonstrate high levels of procedural understanding, but not of conceptual understanding.

What are the practical implications for the results of these findings for the teaching of statistics? This question forms the basis of Chapter 7, and three aspects of this issue are discussed. Section 7.2 focuses on the implications of the study for the teaching/learning strategy. In section 7.3, recommendations are made concerning the development and use of technology in the introductory statistics course. Issues of assessment are the concern of section 7.4.

7.2 Implications for the course structure and content

In this study an analysis of the schemas constructed by students in relation to statistical inference showed an association between the concept of sampling distribution and the development of conceptual understanding of statistical inference. It was clear from the results of the study that, even though the instructional treatment was designed to facilitate the development of links between previously identified key concepts for students, many students formed schemas which were fragmented and which excluded key relationships. Whilst evaluating the effectiveness of the instructional treatment was not an explicit aim of the study, it is valuable to explore the possible relationships between the teaching/learning strategy used and the observed outcomes, and offer some suggestions concerning the teaching strategy in order to improve classroom practice.

7.1.2 Developing understanding of the sampling distribution

The theoretical analysis undertaken in this research suggested that the sampling distribution, in particular in its empirical representation, would have a central role in the development of conceptual understanding of statistical inference. This proposition was supported by the findings of the research. It has also been shown that the links between the empirical and theoretical representations of the sampling distribution are difficult to establish and maintain. What are the implications of this finding for the content of the introductory statistics course?

Typically, introductory courses in statistics include a study of descriptive statistics, introducing methods of describing, displaying and summarising realisations of a variable based on a sample (see, for example, Moore & McCabe, 1999). The distribution of the variable in the population is also usually discussed, generally in relation to a probability distribution. For many students, and their teachers, the distribution of the population is seen as the ‘true’ distribution while the sample distribution is seen as a particular case of the theoretical population distribution. Often in texts when dealing with empirical distributions, the implication is made that if there were enough data, plotted on a histogram with small enough intervals, then the probability distribution function would result. For example:

> Just as probability is conceived as the long-run relative frequency, the idea of a continuous distribution draws from the relative frequency distribution for a large number of measurements...Proceeding in this manner, even further refinements of relative frequency histograms can be imagined with larger number of observations

> and smaller class intervals...the curve obtained as the limiting form of the relative frequency histogram represents the manner in which the total probability 1 is distributed over the interval of possible values of the random variable X. (Johnson & Bhattacharyya, 1992, pp.229-230)

Others express similar sentiments concerning the relationship between the empirical sampling distribution and the theoretical sampling distribution. For example:

> All of the sampling distributions that we will be referring to in the subsequent chapters of this book will be theoretical sampling distributions. After the present chapter, we will have no further occasion to tabulate or plot an empirical sampling distribution, since the exact theoretical sampling distributions are known for the inferential statistics we will be studying. (Olson, 1987, pp.290-291)

and

> Two theoretical distributions provide a bridge between descriptive and inferential statistics: a probability distribution and its close relative, a sampling distribution. A probability distribution associates a probability with each value of a random variable. If the random variable is some function of two or more population elements, say a mean or a standard deviation, a probability distribution is called a sampling distribution. (Kirk, 1990, p.306)

The instructional strategy used in this study deliberately took the position that the empirical sampling distribution was an equally valid and necessary representation of the behavior of the sample statistics as the theoretical representation, useful in the empirical form for subsequent decision making in statistical inference. To make the link between the empirical and theoretical representations, it was further shown that, under certain circumstances, the empirical sampling distribution could be *modelled* by a known probability distribution, and that this known distribution is used in the determination of P-values. To reinforce and maintain this link, the relationship between the empirical and theoretical representations was established and revisited often throughout the study. The theoretical content analysis underpinning the instructional strategy used in this study had identified the establishment of this link as an important feature of the teaching/learning strategy, a necessary pedagogic action to facilitate the development of this link in the student mental structure.

While the study was not specifically designed to evaluate the effectiveness of the instructional strategy on the maintenance of both representations of the sampling distribution, it did provide some information on the success of this strategy. For instance, from the concept maps it could be determined that 9 of the 21 students (43%) who constructed the concept map for sample proportion explicitly noted that the empirical sampling distribution could be modelled by the normal distribution. Whether or not this is more than could be expected by a more traditional teaching strategy it is not possible to say. However, the study does show that those students who did form this link score higher on the measure of conceptual understanding than those who did not. Thus, the findings of the study are consistent with the theoretical content analysis which established the importance of this link in the understanding of statistical inference, and suggests that further research to identify educational strategies that establish and maintain the multiple representations of sampling distribution would be worthwhile.

7.2.2 Distinguishing the sampling distribution from the sample distribution

An unexpected finding from the analysis of the students' schemas was the confusion between the sample distribution and the sampling distribution. This lack of recognition by students of the sample statistic as the source of the sampling distribution has the potential to render all subsequent development of statistical inference as meaningless.

The similarity between the terms *sample* distribution and *sampling* distribution is obvious. Whilst the distinction between the concepts they describe is fundamental to a study of inference, this study has shown that it is clearly not so obvious to the student of statistics. On reflection, there is a basis for this confusion, and yet it appears to be unrecognised in the current literature, and certainly by textbook writers. Inasmuch as almost all courses in statistics define what a sampling distribution is, almost none explicitly differentiate it from the distribution of a sample. Exceptions to this are the computer packages *Sampling Laboratory* (Rubin, 1990), and *Models'n'Data* (Stirling, 1991) where dynamically linked windows simultaneously display the sampling distribution and the contents of the sample last selected, making obvious the distinction between the two distributions.

This study found that the confusion between the sample distribution and the sampling distribution arose for some students from the first mention of sampling distribution, and stayed with them throughout the whole course. For others, the term sampling distribution was initially used to correctly describe the distribution of a sample statistic, but somewhere along the line it merged with the sample distribution. Some students used an incorrect name (sample distribution instead of sampling distribution) to describe the distribution of the sample statistic. For others the reverse occurred, with the term sampling distribution applied incorrectly to the distribution of the sample.

Clearly there is a problem with the language being used to name the distributions under discussion. It is an extremely challenging task for a student to recognise the subtle differences in the three distributions being referred to, even if they are aware that there exist three distinct concepts. Given the similarity in the names of the distributions, it is highly likely that many students were unaware that more than one distribution actually existed. For purposes of discussion, the three distributions were termed the sample distribution, the empirical sampling distribution, and the theoretical sampling distribution. These distributions need to be given names which are more clearly discernible, such as perhaps *data distribution* (for sample distribution), *sampling pattern* (for the empirical sampling distribution) and *sampling distribution* (for the theoretical sampling distribution).

In the light of this unexpected confusion the instructional strategy needs to be modified to specifically include representations of all three distributions; the population distribution, the sample distribution and the sampling distribution, and to ensure that the similarities and differences in the language and notation are made explicit. Exercises should be included in the teaching and learning materials which draw the students' attention to the specificity of the language, such as asking students to name a distribution as *population*, *sample* or *sampling*, from either a graphical representation, a verbal description or both. A concept mapping activity which explicitly asks students to identify the relationships between the population distribution, the sample distribution and the sampling distribution would encourage students to undertake a metacognitative analysis of these concepts could also be of potential benefit.

The desirability of comparing and contrasting explicitly the population distribution, the sample distribution and the sampling distribution also has implications for the design of computer software for use in the teaching and learning of sampling distributions. This was previously recognised by software developers Rubin (1990) and Stirling (1991). More recently, an instructional technology was developed by Cohen et al (1994) called ConStatS, software designed to teach probability distributions. According to Cohen & Chechile (1997):

> Understanding the *differences* between a data distribution and a probability distribution is one of the most profound insights a student can have. (p. 254)

The results of this research study add weight to this contention, and extend it to include the desirability of students also understanding the difference between the sample distribution and the sampling distribution.

7.2.3 The role of the concept map in instruction

As a result of this study it is the researcher's opinion that the concept map does have a role in the teaching of statistical inference, but not the role for which they were used in here. In this study concept mapping was used as a research tool for externalising the students' internal mental structures. Almost always, the concept maps were constructed in class time. On the one occasion that the concept map was set for homework only 9 of the 23 students completed the task.

At the end of the study, students were asked their opinion of concept mapping as an activity. The overall response was not wildly enthusiastic, with 24% of the student in the main study rating concept mapping as not helpful, 38% as average and 38% as helpful. This was a much more negative response to that of the group that undertook the preliminary study, who had been extremely positive about the potential of mapping exercises to support conceptual growth (and had in fact requested that lecturers in other courses include concept mapping activities).

The difference in attitudes of the two groups suggests that a difference in the effectiveness of the way in which the concept mapping activity was conducted with the two groups. The essential difference in the way in which the concept maps which were used with the two groups was the collaborative discussion and evaluation of maps that took place with the preliminary study group. Since there was no need to attribute each map to a particular student, such discussions were encouraged. The value of doing this has been noted recently by Wilcox (1998):

> But what seemed even more valuable were the discussions we had as we constructed the maps and them shared them with out team members. We discovered that we had intense conversations about what to include and what not to include and why. (p. 465)

The students in the preliminary study were also shown the maps which had been prepared by the researcher and colleagues (the expert maps), and were able to compare these to their own. In this way, these students were able to identify and explicitly discuss the relationships that were important to the expert, but were missing from their own maps. There is no doubt that the student maps in the preliminary study were "better" than the maps of the student in the main study, in terms of the number and complexity of the relationships that they included. What was not able to be determined from the preliminary study was whether the student maps were becoming more like the instructor maps because the students were in a sense copying the instructor maps, or whether the maps themselves were evidencing real cognitive growth in the students.

The use of concept maps with the students in the study was primarily as a research instrument to identify qualitative differences in schema between students, and to document cognitive changes in the students over the course of the instructional treatment. Since the maps prepared by the students where necessarily highly individual, group discussions and mapping activities were not undertaken during the study. This left students feeling frustrated at times, with several requesting feedback on the maps that they had prepared. Whilst this lack of discussion was a necessary component of the research process, in the opinion of the researcher concept mapping is a far more profitable aid to student learning when used as a small group or class activity, encouraging students to explicitly identify relationships between concepts. That is, as a learning strategy for the student rather than as an evaluative tool for the researcher. As stated by Novak (1990b, p.37), "the primary benefit of concept maps accrues to the person who *constructs* the maps".

The recommended role for concept mapping activities is thus to encourage the students to participate in reflective thinking. However, there is a natural resistance by some students for engaging in meaningful learning. This phenomenon is well documented. According to Novak (1990b):

> We were not alone in our failure to recognise the pervasive, pernicious character of rote learning and associated pedagogical strategies; however, most reports on the use of metacognitive strategies make no mention of this as a potential or real problem. We now see this as a major (if not the major) concern to be addressed when moving to include metacognitive strategies into real school settings. (p. 43)

Some recent studies concerning the development of conceptual understanding in statistics at post-secondary level have recommended the use of small, co-operative learning groups (Garfield, 1993b; Giraud, 1997; Keeler & Steinorst, 1995; Magel, 1998). According to Garfield:

> …using cooperative groups relates to the constructivist theory of learning...Small-group learning activities may be designed to encourage students to construct new knowledge as they learn new material, transforming the classroom into a community

> of learners, actively working together to understand statistics. (1993b, paragraphs 10-11)

Working in small groups requires the students to synthesise and organise their knowledge so that they may verbalise. This is supported by Giraud (1997, para 4) who states:

> Discussions among students create external connections with the material by reference to a variety of experiences and perspectives.

Many suggestions are given in the literature concerning the types of activities that students might undertake in small groups. Most of these are based on problem solving or data analysis scenarios. However, in the opinion of the researcher, a suitable small-group cooperative activity would be concept mapping. When used in this way the concept mapping activities can contribute positively to the growth of students understanding by:

1. engaging students who actively avoid metacognitive activities;
2. facilitating discussion concerning concepts and the links between concepts, which is based on student reflection on their learning;
3. encouraging the formation and explicit recognition of links between concepts in the mental structures created by students;
4. providing students with feedback concerning the validity or otherwise of relationships between concepts;
5. giving students insight into the conceptual structures of others, some of which may be considered to be more expert than their own.

The way in which concept maps were used in this study addressed only steps 1 and 3 of this list of recommendations. After the analysis of the outcomes of this study, and comparison with the outcomes of the preliminary study, concept mapping activities designed to facilitate cognitive growth rather that the research process would encompass all five steps noted here.

7.3 Implications for technology use in teaching statistics

As previously mentioned, there has been a recent proliferation of technology use in statistics courses, but sometimes a confusion concerning the intended impact of the technology on student learning. In order to discuss the role of technology in teaching statistics, it is useful to differentiate between the two roles for which technology was used in this research.

7.3.1 Effects *with* and effects *of* technology

As previously mentioned, Salomon, Perkins and Globerson (1991) describe two possible purposes for the use of technology in teaching and learning; the first of which they term the effects *with* technology relating to the capability of the combined system of technology and user. The second is termed effects *of* technology, related to the longer-term effect of working with the technology on the cognitive structure of the individual. What can this study contribute to our understanding of the role of technology in the teaching and learning of statistics?

7.3.1.1 Effects *of* technology

How effective were the computer activities used in this study in encouraging the development of understanding in these students? Computer based sampling software was used in the study to provide representations and simulations which might facilitate the formation of links between concepts in students' mental structures. That is, to support the development of conceptual understanding in statistical inference.

The effect of using the technology to help developing understanding is unclear from this study. The dynamic sampling packages *Sampling Laboratory* (Rubin, 1990) and *Models'n'Data* (Stirling, 1991) were used to help introduce the idea of the sampling distribution for the sample proportion and the sample mean. While reasonable proportions of students in the study recognised that the sampling distribution was the distribution of a sample statistic (62% and 80% for the proportion and the mean respectively), only 26% of students were able to carry out the integrative reconciliation necessary to recognise each of these as specific examples of the general concept of sampling distribution.

Could these software packages have been used more effectively? The computer sessions, which formed a part of the teaching strategy in this study, were instructor-centred expositions. However, it seems that a desirable student experience with which to complement the group sessions would be laboratory sessions in which the students gained first hand experience with the software, in a small-group cooperative situation. This would enable the students to experiment for themselves with the sampling process, and the effects of varying certain parameters on the sampling distribution. Such sessions were not held during the study, due to the limited availability of the software at the time. However, with such software becoming more available on all computer platforms, particularly via the web, then this should not be such a limitation in the future (see for example the collection of demonstrations at http://it.stlawu.edu/~rlock/tise98/java.html).

From this study it seems possible that the instructional treatment, including the use of the sampling software, was helpful in elucidating some important concepts of sampling distribution in each of the specific contexts in which it was applied for some students. However, the nature of these software packages is such that they are always context specific, and as such they have no specific role in illustrating the concepts and links which together form a schema for the generalised sampling distribution. That is, the recognition of each of the contexts as a particular example of a general concept is an act of integrative reconciliation cannot be assumed to follow from experience with the computer software under discussion. Is it possible to design computer software which, when appropriately integrated in a teaching/learning strategy, could encourage the formation of the generalised sampling distribution concept? This would seem to be the next challenge for the educational software developer. Whilst many packages exist and continue to be developed which are designed to support student learning in the area of sampling distribution, research based studies to establish the effects of most of these computer/student interactions still need to be carried out. Cohen and Chechile (1997) suggest that in order to evaluate an instructional technology the research must ask questions such as:

> When students are shown a probability distribution in an instructional exercise, what exactly do they see? What concepts does it evoke? And what concepts must it evoke for students to use it in a learning exercise? (pp. 261-262)

Only after such evaluations are undertaken will a clear picture emerge concerning the potential of the software to support the development of conceptual understanding in statistical inference.

7.3.1.2 Effects *with* technology

The instructional strategy implemented in the study was designed to incorporate current technology. A statistical calculator and a specialised statistical package (*Minitab*, 1992) were used as tools throughout the course of instruction, freeing students from labour intensive hand calculations. The students were only expected to carry out first principles hand calculations in the most simple of cases. Once the more complex statistical processes were under discussion, a statistical package was always used.

The impact of the use of technology as a tool for calculations alone should not be underestimated, given the volume and difficulty of the calculations necessary in a statistics course. However, there are those who resist abandoning ‘first-principles’ calculations, arguing that conceptual understanding must be preceded by familiarity with formulae and how they work. For example, Khamis (1991, p295) suggests:

> Hand computations involving data lead to an appreciation of, and insight into, the nature of chance phenomena and give the student that empirical basis without which no real understanding of a science can be achieved.

The belief of such proponents seems to be that an understanding of the elements of the formula and the procedure required for the students to perform the calculations correctly is necessary to the development of understanding of the concept represented by the formula. However, what students develop by performing a hand calculation is the understanding of the process involved in making such a calculation, which is often totally unrelated to the underlying process at work, because the formulae used are computational formulae rather than definitional formulae.

Does working with the formulae from first principles necessarily precede conceptual understanding? According to the measures used in this study, many students exhibited high levels of conceptual understanding, without having had the experience of working through all of the calculations by hand. Since this is a heterogeneous group in terms of previous statistical experience, it could be argued that those students who developed the higher levels of conceptual knowledge had previous educational experiences that had given them 'hands-on' experience with the formulae. However, since the correlation between the score on the prior statistical experience questionnaire and the score on the conceptual measure was negligible ($r = 0.042$, $n = 21$), then this would appear to be an unlikely explanation.

A counter argument to theory that first principles calculations are necessary for understanding is concerned with the amount of time saved by using technology, time which can be devoted to teaching for understanding. A second argument against the need for first-principle calculation is concerned with the cognitive load argument (Sweller, 1993). That is, that by reducing the intellectual demand on students associated with the process of calculation, the student can focus on the wider issues, such as appropriateness of the analysis and interpretation of the results.

However, the results of this study confirm that conceptual understanding cannot be *assumed* to follow from demonstrated expertise with the procedures of statistics. That means that the reduction of cognitive load that accompanies the introduction of an appropriate technology does not uniformly lead to the development of conceptual understanding. Thorough theoretical analyses of both content and student learning processes are required in order to develop materials and strategies which may encourage development of conceptual understanding, the other role for which technology has been used in this study.

7.3.2 Computer Intensive Methods

The results of this study have implications for the use of Computer Intensive Methods in introductory statistics courses. As previously discussed in Section 3.5, Computer Intensive Methods is a terms describing a set of techniques in statistics which essentially rely only on empirical distributions to carry out analyses. This set of techniques requires statisticians to make fewer assumptions concerning the distribution of their data and was made possible by the development of fast and efficient computer technology.

Early research carried out on the educational potential of empirical methods (termed Monte Carlo method) by Simon, Atkinson & Shevokas (1976) indicated that both achievement in statistics and attitude towards the subject were improved through teaching such a course. However, the reasons given for this are supposition at best. For example:

> The advantage of the Monte Carlo method seems to stem from its greater simplicity in a fundamental intuitive sense due to having fewer "working parts", and because the student never needs to take anything on faith, especially the sort of faith that is necessary with analytic methods that work by way of the central limit theorem. (Simon, Atkinson, & Shevokas, 1976, p. 739)

Recent studies which support the use of a computer intensive method instead of or in conjunction with traditional methods, are supportive of this as a strategy but are not based on theoretical analysis nor empirical evidence (Alper & Raymond, 1998).

This study adopted some of the principles of Computer Intensive Methods, by using the empirical sampling distribution to determine the P-value in hypothesis tests for means and proportions. These computer-generated empirical sampling distributions were used in this study in two ways. By explicitly forming the sampling distribution from the repeated sampling, the empirical sampling distribution served as a conceptual bridge between the sampling process and the determination of the P-value. They were also used to show that, in most circumstances, the theoretical distribution was an appropriate model for the empirical distribution.

The results of the study show that most students (87%) were able to use the empirical sampling distributions to carry out an hypothesis tests. This result gives some empirical validation of the potential for empirically based, or computer intensive methods. That is, that the logic and methods of hypothesis testing were clear to most students on the basis of the empirical sampling distribution. Whether or not the level of conceptual understanding developed by the students was greater because of their experiences with the empirical sampling distribution was not a question addressed by this study. However, the theoretical analyses carried out earlier in the study, based on a view of knowledge organised in semantic networks comprised of various interrelated schema and an analysis of the content domain base of statistical inference, suggest that the use of computer intensive methods for statistical inference would support the development of conceptual understanding in this area. Further research is needed which specifically addresses this question.

7.4 Implications for assessment in statistics

The need to include a variety of assessment tasks when evaluating student learning is clear from the results of this study. Had only the traditional skills-based measurement tasks such t-tests or chi-square tests been used to evaluate student performance, only two students would have been deemed to have performed at an ‘unsatisfactory’ level (that is, failed the examination). And yet, an analysis of tasks designed to elicit deeper understanding revealed substantial variation in the level of conceptual understanding demonstrated. Thus, further thought needs to be given to the nature of tasks which could be used to measure understanding in statistics.

At the time at which this study commenced, the statistics education profession had not really addressed the issue of assessment of understanding in statistics. Some suggestions had been made concerning assessment, but the tasks suggested were developed in isolation, in an ad hoc way, which made it difficult to relate these tasks to a measure of overall student performance. An important consideration when developing assessment tasks for the classroom is that the assessment be applicable with a group of students with whom contact is only available for limited periods of time, and almost entirely in a group situation. Thus, activities, which take a lot of time, or involve intense on-to-one interaction, such as interviews, are really not feasible.

The current study provides a framework which ensures that a full range of aspects of understanding are addressed by a variety of assessment tasks. This framework can be used by other statistics educators to address issues of assessment, in that tasks currently used can be classified within the framework, and gaps or omissions in assessment identified and appropriate new tasks developed.

The framework can be applied to recent developments in assessment in statistics, beginning with publications by Garfield (1994), Konold (1995) and Gal & Ginsberg (1994), and culminating in the publication of *The Assessment Challenge in Statistics Education* (Gal & Garfield, 1997). This is a collection of nineteen papers focussing on assessment in statistics at all age levels. There is now available to statistics educators a body of works which contain suggestions for the measurement of statistical understanding. Garfield and Gal state (1997):

> Now that attention to the teaching of statistics has become visible at all educational levels, it has become apparent that assessment of student learning and understanding of statistics is not being adequately addressed in current projects and instructional efforts. (p. 1)

Begg (1997) suggests that assessment instruments should be based on a constructivist rather behaviorist theory of learning, with tasks designed to illicit what the student understands as opposed to what the student can do as has been the case in the past. This view is consistent with the approach adopted in this study, where tasks have been developed to assess both conceptual and procedural understanding.

Colvin and Vos (1997) suggest that assessment tasks should be authentic, in that they assess student performance on tasks that are relevant outside the classroom or the lecture theatre, and give some measure of the student's ability to transfer learning to new situations. Authentic assessment strategies are also strongly recommended by Chance, (1997) and MacGillivray (1998). Colvin and Vos suggest that if a situation is judged by the student to be one which could actually occur in real-life, then the student will be more likely to attempt the task. To be judged an authentic assessment, the task needs to:

> …capture the interest of the student, make the situation believable to the student, incorporate sufficient real data to lend credibility and, most importantly, focus the process towards important statistical concepts. (Colvin & Vos, 1997, p.29)

Authentic assessment tasks would seem to fall into the categories previously identified and discussed as *understanding as connections between types of knowledge*, *and understanding as situated cognition*. Consideration of the assessment tasks used in this study shows that several fall into the category of authentic assessment tasks, as recommended.

Kelly, Sloane and Whittaker (1997) were concerned with assessing conceptual understanding of statistical ideas, and suggest some tasks which uncover students misconceptions concerning the concept of the mean, and analysis of variance. Both of these tasks confirmed the sentiments expressed earlier in this thesis, that many students are concerned with rule-based or instrumental learning (Skemp, 1979). Kelly et al (1997, pp. 85-86) suggest that one should

> …adopt assessment techniques that allow the teacher some insight into the students' thinking. These techniques should be relatively easy to use, and economical on the teacher's time.

That students still lack conceptual understanding when interpreting hypothesis testing was confirmed by Kelly et al (1997) and also by Schau and Mattern (1997), who attribute the difficulties many students face in statistical reasoning as a lack of connected understanding. That is, that many students lack connections between importance concepts in their schemas. This argument is similar to that proposed by the researcher in this study, and Schau and Mattern, as well as others (Schau, 1997; Williams, 1995) recommend the use of concept maps to assess the degree of connection of student schemas, as well as the evaluation of student concept maps by comparison with expert maps.

Joliffe (1997) confirms the researcher's analyses by recommending open-ended questions, such as those used in the Explain task and the Radio task, for measuring conceptual understanding. She states:

> …open-ended questions can go a long way towards meeting the requirements of good assessment questions. They are able to pose 'real' questions with 'real' data and assess choice of appropriate techniques and the doing and (written) communication of statistics. They also ask students to explain concepts – a sure test of understanding. (p. 199)

An analysis of the more recent literature concerned with assessment in statistics showed that, whilst assessment in statistics has become of greater interest to researchers in recent years, there is still a lack of instruments with which to assess conceptual understanding in statistical inference. However, there many valuable suggestions have been made concerning the possible style and form of assessments that could be used.

This analysis of the suggestions has shown that the tasks developed for measuring conceptual understanding in this study are consistent with current recommendations from the profession, and that in fact these tasks could be valuable to other researchers and educators for inclusion with or without modification into their assessment programs,

In conclusion, it is worth noting that whatever assessments are used, it is necessary for the educator to remain aware of the students' active preference for procedural learning, and their consequent tendency to "practice" even novel questions until a procedure is created. For example, consider the questions requiring students to carry out an hypothesis test for the same scenario using both an empirical sampling distribution and a theoretical sampling distribution (the modelling questions). The instructional purpose was to develop and establish externally the link between the theoretical and empirical representations of the sampling distribution, in order to facilitate the establishment of the link in the student schema. However, analysis of the data indicated that many students had proceduralised both tasks as both were found to contribute to the students measure of procedural understanding. As stated by Hubbard (1997):

> If an instructor produces a non-standard question and keeps repeating it, then it becomes a standard question and students will learn a standard response. (para 13)

Thus, in order to reveal a student's level of conceptual understanding, families of tasks need not only to be developed but also continually modified so not to become proceduralised.

The framework developed for the assessment of statistics in this study, and the tasks developed within this framework, provide a starting point for the development of suitable assessment instruments in the area of statistical inference, able to be used by others in the field of statistics educator to suit the content and objectives of their particular courses of study.

7.5 Conclusion

The results of the study have implications for teaching practice in statistics. Firstly, they provide some evidence in support of the theoretical notion of understanding as the ability to maintain and move between representations of a concept, in this case the theoretical and empirical representations of sampling distributions. This suggests that instructional strategies, which endeavor to address and maintain both representations of the sampling distribution, are necessary. The study has also identified a misconception that students often develop concerning the distinction between the sample distribution and the sampling distribution. As a result an instructional strategy which explicitly identifies and discusses the distributions of the population, the sample and the sample statistic is suggested.

The students' responses to the value of concept maps suggested that, whilst concept mapping proved to be a valid and useful method for determining an external representation of the students' schemas in this study, they could perhaps be more profitably used as a tool to encourage reflection and hence conceptual growth in the students. Specific recommendations concerning the use of concept mapping for this purpose have been given in this chapter.

Some issues concerning the role of technology in the teaching and learning of statistics have also been addressed. It was found that many students were able to exhibit high levels of conceptual understanding in content areas where they had not previously undertaken calculations from first principles, suggesting that familiarity with calculations is not an important step in the process of developing understanding. This is contrary to popular belief. It was also found that conceptual understanding does not automatically result from the student's interaction with a software package, even those specifically designed to support the development of such understanding. However, the findings also show that some students can undergo cognitive reorganisation on the basis of such interactions, and more research is needed in order to explore the nature of such changes, and how they relate to the students' instructional experiences.

The study has confirmed the belief held by many statistics educators, that conceptual understanding does not necessarily accompany procedural understanding in statistical inference, and that there is a need for new, but adaptable, assessment instruments. Those instruments developed for this study provide a basis for the further development of assessment tasks specifically designed to measure both conceptual and procedural understanding in statistical inference.

Chapter 8
Conclusion

8.1 Introduction

In the past many statistics educators have suggested that few students develop the level of conceptual understanding which is essential for these students to apply the statistical techniques at their disposal and to appropriately interpret their outcomes (for example Gordon & Gordon, 1992a; Hogg, 1992, Moore 1992a, 1992b, Garfield & Ahlgren, 1988). More recent publications suggest that this is still the situation (for example Gal & Garfield, 1997; Schau & Mattern, 1997; Burrill, 1998; Thomason & Les; 1998; Williams, 1998).

The goal of this study was to understand better the process of learning in the area of statistical inference in order to clarify the role of the sampling distribution in student understanding. Based on this knowledge, it was hoped that recommendations could be made which would assist statistics educators, concerning the content and conduct of teaching and learning strategies in this area.

8.2 Summary of the study

This study has viewed learning and the nature of knowledge from a cognitive science point of view, synthesised from previous research into learning in general and learning in mathematics and statistics in particular. With this as a basis, an analysis of the knowledge domain known as introductory statistical inference was undertaken then, using the content, structure and organisation of the schema created by experts in this area, critical concepts, and relationships between concepts were identified. With knowledge of these key concepts and relationships between concepts an instructional strategy was developed, the application of which was situated within the theoretical framework of a constructivist teaching experiment modified for the post-secondary teaching environment.

During the course of the teaching experiment, the process of knowledge construction by each of the subjects was monitored regularly using concept maps, and evaluated by comparison with those prepared by experts. At the end of the teaching experiment, student achievement was measured in terms of both procedural and conceptual understanding.

8.3 Summary of results

This study is concerned with the content and form of the mental structures that are likely to be associated with the development of conceptual and procedural understanding in students of introductory statistical inference.

The specific hypothesis addressed by the research can be stated as follows:

> Students whose schema for sampling distribution demonstrates links to the sampling process and whose schema for statistical inference includes links to the sampling distribution, will show evidence of both conceptual and procedural understanding of statistical inference, whilst those who have not integrated these concepts into their cognitive structure may still demonstrate high levels of procedural understanding, but not of conceptual understanding.

Through an analysis of the concept maps prepared by the students it was evident that individuals did differ in the number of concepts and relationships identified, and the ways in which their knowledge structures were organised. The extent to which subjects had undertaken the integrative reconciliation of the specific sampling distributions discussed to construct a generalised schema for sampling distribution varied between students. As expected, the role of the sampling distribution in statistical inference, as evidenced by the links between the sampling distribution and the processes of estimation and hypothesis testing, was clearly recognised by some students, but not at all by others.

Based on a qualitative analysis of the schema constructed, students were divided into three groups:

Group 1 Students who showed evidence of the development of the concept of sampling distribution and subsequently integrated sampling distribution appropriately into their schema for statistical inference.

Group 2 Students who showed evidence of the development of the concept of sampling distribution but did not relate this to their schema for statistical inference.

Group 3 Students who did not at any stage show evidence of the development the concept of sampling distribution.

The research hypothesis was addressed by an analysis of the relationship between conceptual structure and understanding of statistical inference. When analysing student understanding at the end of the teaching experiment period no difference was found between the three groups on measures of procedural understanding. However, as hypothesised, students who showed evidence of the development of the concept of sampling distribution, and identified links between the sampling distribution and statistical inference (Group 1), exhibited higher levels of conceptual understanding than those whose concept of sampling distribution was not linked to inference (Group 2). These in turn performed better on tasks which measure conceptual understanding than those whose schema for sampling distribution included very few valid relationships (Group 3).

Thus, the research hypotheses suggested by previous research and further developed from the theoretical analyses was confirmed by the empirical studies undertaken.

8.4 Reflections on the study

The results of this study suggest that the language, content, teaching methodology and assessment of many traditional statistics courses currently offered should be rethought. There is no doubt that after a course on statistical inference, many students are still not able to link their knowledge of statistical inference through the sampling distribution to the sampling process itself. The lack of such a path in the student constructed schemas signals limitations in the students' ability to apply their knowledge in any but the most limited of applications, that is those familiar from the teaching context. And the ability to transfer their knowledge in real world situations is certainly a goal for most statistics educators.

An outcome of this study is a plan of research which, together with the evaluation instruments, enables others, both researchers and classroom practitioners, to study the mental process of their students as they undertake learning experiences. There is a growing demand for formal evaluation of new and different learning strategies in statistics education, particularly with the proliferation of exciting new computer packages, such as StatPlay (Thomason, Cumming & Zangari, 1994) and Dataspace (Finzer & Erickson, 1998), and the equally attractive concept of the interactive and web-based texts currently being developed such as Computer Assisted Statistics Teaching, currently being developed and trialled by Stirling (2000).

However, in spite of their attractiveness to both students and teachers, the educational value of these exciting new teaching resources cannot be assumed. Equally important is the need to evaluate these resources within in a theory of learning and knowledge construction. As stated by Hiebert (1997):

> …careful analyses of students' developmental capacities and their likely learning strategies, and of the key subject matter ideas, can produce intervention programs that yield impressive gains in mathematics. (p. 93)

Whilst a large body of research exists which validates the use of concept maps to document a student's learning journey, this is an area which invites further investigation in the future in two particular areas. Firstly, the variation which occurs between the concept maps constructed by a student is assumed to reflect changes in that student's conceptual structure, and yet it is entirely possible that there will be a degree of within student variation, as within any real world system. The extent of that variation needs to be ascertained in order to determine whether or not the student's map is reflecting real changes. Secondly, there is a need for more research investigating reliability of the interpretation schemes used by researchers when analysing the concept maps prepared by students, in order to confirm that consistent interpretations are made.

Another outcome of the study has been the development of an instructional strategy supported by a set of student materials, which is neither the recipe book nor the mathematical treatise lamented by Shaughnessy (1992). This approach uses empirical strategies rather than mathematical rules for the ideas of probability necessary to understand statistical inference, as suggested by many statistics educators (Garfield & Ahlgren, 1988; Moore, 1992a). This approach was supported by the use of the dynamic and interactive representations offered by the new technologies (Kaput, 1992; Rubin, Rosebery, & Bruce, 1988), while the capacity of technology to redirect the emphasis from calculation to concept (Jones, 1996) was extensively utilised. Finally, and perhaps more importantly, the learning materials and activities developed for this study were designed to support the development of both procedural and conceptual understanding in students. However, further developmental work on the teaching and learning strategy needs to be undertaken, in particular with larger and more diverse groups of students, before any conclusions can be reached concerning the success of the teaching model.

8.5 Conclusion and Recommendations

The findings of this study have emanated from a small, and in some sense specialised group of students. Whilst research hypothesis have been supported within this group, the generalisability of the results to other students, particularly undergraduates, is not theoretically justifiable without replication of the research with other student groups of more diverse age and academic backgrounds. This would seem to be the next step in extending our understanding of the knowledge construction process of students learning introductory inference.

However, the data collected and analysed here do support the conclusion that for some students both procedural and conceptual understanding of statistical inference did develop over the period of this study. For other students, however, there was little evidence of conceptual understanding and, in fact, signs that some students made promising beginnings which were then curtailed. Was this the result the students' preconceptions before the course, of the instruction, or the students' attitude to their learning, or some other factors which have not been considered? These issues were not directly addressed, and they need to be included in order to understand more fully the nature of the process by which students acquire knowledge in statistical inference.

The role of the technology in facilitating the construction of links between representations also needs to be more deeply researched. We need to develop a clearer idea of what the student perceptions are when engaging with pedagogic computer based technologies, if we accept the constructivist argument that the type of information communicated, and the learning occurring, is clearly not the same for all students.

This dissertation is based on a cognitive science model of knowledge and knowing which was current when the research began. However, the theories of learning continue to develop and change as the body of research increases (Davis, 1996). It may be that some of that within an expanding theoretical framework which considers these recent developments, some insights may be gained into the questions posed here but which remain unanswered.

References

Adams, J. L. and Stephens, L. J. (1991). A comparison of computer-assisted instructional methods. *International Journal of Mathematics Education in Science and Technology,* 22(6), 889-893.

Allan, L. and Lord, S. (1991). *For a number of reasons*. Redfern, NSW: Adult Literacy Information Office.

Alper, P, and Raymond, R.L. (1998). Some experiences comparing simulations and conventional methods in an elementary statistics course. In L.Pereira-Mendoza, L. S. Kea, T. W. Kee and W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching of Statistics* (pp.885-890). Voorburg, The Netherlands: International Statistics Institute.

Astruc, J., Cracknell, N., Diver, R., Evans, M., Gauld, E., Taylor, N. and Woolacott, B. (1993). *Mathematics with the Macintosh*. Melbourne: Scotch College.

Ausubel, D. P., Novak, J. D., and Hanesian, H. (1978). *Educational psychology: A cognitive view*. (2nd Edn.). New York: Holt, Rinehart and Winston.

Begg, A. (1997). Some emerging issues underpinning assessment in statistics. In I. Gal and J. Garfield (Eds.), *The Assessment Challenge in Statistics Education* (pp. 17-25), Amsterdam: IOS Press.

Betteley, G. (1990). Something approaching normality – the central limit theorem, *'A' level Ahead*. Birmingham: Centre for Extension Education.

Biehler, R. (1993). Software tools and mathematics education: the case of statistics. In W. Dorfler, C. Keitel and K. Ruthven (Eds.), *Learning from Computers: Mathematical Education and Technology* (pp. 1-37), Berlin, Springer.

Biehler, R. (1991). Computers in probability education. In R. Kapadia and M. Borovcnik (Eds.), *Chance Encounters: Probability in Education* (pp. 169-211). Netherlands: Kluwer Academic Publishers.

Bluman, A. G. (1997). *Elementary statistics: a step-by-step approach*. Boston: WCB/McGraw-Hill.

Brown, A. L., Ash, D., Rutherford, M., Nakagana, K., Gordon, A., and Campione, J. C. (1993). Distributed expertise in the classroom. In G. Salomon (Ed.), *Distributed cognitions* (pp. 188-228). Cambridge: Cambridge University Press.

Brown, C., and Schrage, G. (1989). Elementarization of statistical reasoning using simulation and computer graphics. *International Journal of Mathematics Education in Science and Technology*, 20(3), 375-381.

Burrill, G., Burrill, J. C., Coffield, P., Davis, G., de Lange, J., Resnick, D., and Siegel, M. (1992). *Data Analysis and statistics across the curriculum*. Reston, Virginia: NCTM.

Burrill, G. (1998). Beyond data analysis: Statistical inference. In L.Pereira-Mendoza, L. S. Kea, T. W. Kee and W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching of Statistics* (pp.663-669), Voorburg, The Netherlands: International Statistics Institute.

Chance, B. L. (1997). Experiences with authentic assessment techniques in an introductory statistics course. *Journal of Statistical Education*, [Online], 5(3),

Chervany, N. L., Collier, R. O., Fienberg, S. E., Johnson, P. E. and Nieter, J. (1977). A Framework for the development of measurement instruments for evaluating the introductory statistics course. *The American Statistician*, 31(1), 17-23.

Cobb, P., Yackel, E. and Wood, T. (1992). A constructivist alternative to the representational view of mind in mathematics education. *Journal for Research in Mathematics Education*, *23*(1), 2-33.

Cohen, S. and Chechile, R. A. (1997). Probability distributions, In I. Gal and J. Garfield (Eds.), *The Assessment Challenge in Statistics Education* (pp. 253-262), Amsterdam: IOS Press.

Cohen, S., Smith, G. E., Chechile, R. A. and Cook, R. (1994). Designing software for conceptualizing statistics, In L. Brunelli and G. Cicchitelli (Eds.), *IASE: Proceedings of the First Scientific Meeting* (pp.237-245). Perugia, Italy: University of Perugia.

Colvin, S. and Vos, E. V. (1997). Authentic assessment models for statistics education, In I. Gal and J. Garfield (Eds.), *The Assessment Challenge in Statistics Education* (pp. 27-36), Amsterdam: IOS Press.

Cox, C. and Mouw, J. T. (1992). Disruption of the representativeness heuristic: Can we be perturbed into using correct probabilistic reasoning. *Educational Studies in Mathematics*, 23, 163-178.

Davis, R. B. (1986). Conceptual and procedural knowledge in mathematics: A summary analysis. In J. Hiebert (Ed.), *Conceptual and Procedural Knowledge: The Case of Mathematics* (pp. 265-300). Hillsdale, NJ: Lawrence Erlbaum Associates.

Davis, R. B. (1996). Classrooms and cognition. *Journal of Education*, *178*(1), 3-12.

Demitrulias, D. M. (1988). (Creatively) Teaching the meanings of statistics. *Statistics*, *62*, 168-170.

Devore, J., and Peck, R. (1986). *Statistics*. St Paul: West Publishing Company.

Diaconis, P., and Efron, B. (1983). Computer-intensive methods in statistics. *Scientific American* (May), 96-108.

Dubin, R. and Taveggia, T. (1968). *The teaching-learning paradox*, Eugene, Oregon: Center for the Advanced Study of Educational Administration, University of Oregon.

Edwards, J., and Fraser, K. (1983). Concept maps as reflectors of conceptual understanding. *Research in Science Education*, 13, 19-26.

Falk, R. (1993). *Understanding probability and statistics: A book of problems*. Wellesley, MA: A.K.Peters.

Finzer, W. F. and Erickson, T. E. (1998). Dataspace - A computer learning environment for data analysis and statistics based on dynamic dragging, visualisation, simulation and networked collaboration. In L.Pereira-Mendoza, L. S. Kea, T. W. Kee and W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching of Statistics* (pp.825-829), Voorburg, The Netherlands: International Statistics Institute.

Fuller, C. (1992). *Doubts on letters promise*. Melbourne Herald-Sun, November 25.

Gal, I. and Garfield, J. (Eds.). (1997). *The assessment challenge in statistics education*. Amsterdam: International Statistics Institute.

Gal, I. and Ginsberg, L. (1994). The role of beliefs and attitudes in learning statistics: Towards an assessment framework. *Journal of Statistics Education*, [Online] *2*(2).

Garfield, J. (1993a). Notes: Assessment and teaching statistics. *Journal of Statistics Education*, [Online] 1(1).

Garfield, J. (1993b). Teaching statistics using small-group cooperative learning. *Journal of Statistics Education*, [Online], 1(1).

Garfield, J. and Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal for Research in Mathematics Education*, 19, 44-63.

Garfield, J. B. (1994). Beyond testing and grading: Using assessment to improve student learning. *Journal of Statistics Education*, [Online], 1(2).

Giraud, G. (1997). Cooperative learning and statistics instruction. *Journal of Statistics Education*, [Online], 5(3).

Glencross, M. (1986). *A practical approach to the central limit theorem*. In R. Davidson and J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics* (pp. 91-95), Canada: The University of Victoria.

Goldman, R.N. and Weinberg, J.S. (1985). *Statistics: An introduction*. New Jersey: Prentice-Hall.

Gordon, F. and Gordon, S. (Eds.). (1992a). *Statistics for the twenty-first century, MAA Notes, Number 26*, The Mathematical Association of America

Gordon, F.S., and Gordon, S.P. (1992b). Sampling + simulation = statistical understanding. In F. and S. Gordon (Eds.), *Statistics for the Twenty-First*

Century, MAA Notes, Number 26 (pp. 207-216), The Mathematical Association of America.

Gordon, T. and Hunt, D. (1986). Teaching statistics with the aid of a microcomputer. *Teaching Statistics*, 8(3), 66-72.

Green, D.G. (1982). A survey of probability concepts in 3000 pupils aged 11-16 years. In D. V. Grey (ed.), *Proceedings of the First International Conference on Teaching Statistics* (pp.766-783), London: Statistics Teaching Trust.

Hawkins, A. (1990). Success and failure in statistical education - A UK perspective. In D.Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* (pp 24-32), Voorburg, The Netherlands: International Statistics Institute.

Hawkins, A.S. (1986). Statistics or how to know your onions. *Mathematics in School*, 15(4), 14-15.

Hawkins, A., Jolliffe, F., and Glickman, L. (1992). *Teaching statistical concepts*. Essex, UK: Longman.

Hiebert, J. (1997). Re-thinking what cognitive science can contribute to improving students' learning. *Issues in Education*, *3*(1), 93-101.

Hiebert, J. and Carpenter, T.P. (1992). Learning and teaching with understanding. In D. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 65-97). New York: MacMillan.

Hiebert, J. and Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and Procedural Knowledge: The Case of Mathematics* (pp. 1-27). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hiebert, J. and Wearne, D. (1986). Procedures over concepts: The acquisition of decimal number knowledge. In J. Hiebert (Ed.), *Conceptual and Procedural Knowledge: The Case of Mathematics* (pp. 199-223). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hogg, R.V. (1992). Towards lean and lively courses in statistics. In F. and S. Gordon (Eds.), *Statistics for the Twenty-First Century* (pp. 3-13), MAA Notes No 26, Mathematical Association of America.

Holmes, P. (1985). Using microcomputers to extend and supplement existing material for teaching statistics. In L.Rade and T.Speed (Eds.), *Teaching statistics in the computer age*, Canberra: Chartwell-Bratt.

Howard, D.V. (1983). *Cognitive psychology*. New York: Macmillan.

Hubbard, R. (1997). Assessment and the process of learning statistics. *Journal of Statistics Education*, [Online], 5(1).

Iversen, G.R. (1992). Mathematics and statistics: An uneasy marriage. In F. and S. Gordon (Eds.), *Statistics for the Twenty-First Century* (pp. 37-44), MAA Notes No 26, Mathematical Association of America.

Janvier, C. (1987). Representation and understanding: The notion of function as an example. In C. Janvier (Ed.), *Problems of Representation in the Teaching and Learning of Mathematics* (pp. 67-71). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Jolliffe, F. (1990). Assessment of understanding of statistical concepts. In D.Vere-Jones (Ed.)*, Proceedings of the Third International Conference on Teaching Statistics* (pp. 461-466), Voorburg, The Netherlands: International Statistics Institute.

Jolliffe, F. (1997). Issues in constructing assessment instruments for the classroom. In I. Gal and J. B. Garfield (Eds.), *The Assessment Challenge in Statistics Education* (pp. 191-204), Amsterdam, IOS Press.

Johnson, R. and Bhattacharyya, G. (1992). *Statistics principles and methods (2nd Edn.)*, New York: John Wiley and Sons.

Jonassen, D. H., Beissner, K., and Yacci, M. (1993). *Structural knowledge: Techniques for representing, conveying and acquiring structural knowledge*. Hillsdale, NJ: Lawrence Erlbaum.

Jones, P. (1996). Examining the educational potential of computer based technology in statistics. In C. Batanero (Ed.), *Proceedings of the International Association for Statistical Education Round Table on Research on the Role of Technology in Teaching and Learning Statistics* (pp. 229-239). Spain: University of Granada.

Jones, P. L. and Lipson, K. L. (1993). Determining the educational potential of computer based strategies for developing an understanding of sampling distributions. In W. Atweh (Ed.) *Proceedings of the Mathematical Education Research Group of Australia 16th Annual Conference* (pp.355-360), Brisbane: MERGA.

Kader, G. (1990). Simulations in mathematics – probability and computing. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* (pp. 178-186), Voorburg, The Netherlands: International Statistics Institute.

Khamis, H. J. (1991). Manual computations - a tool for reinforcing concepts and techniques, *The American Statistician,* 45(4), 294-299.

Kahneman, D. and Tversky, A. (1972). Subjective probability. A judgement of representativeness. *Cognitive Psychology*, 3(3), 430-453.

Kapadia, R. and Borovnik, M. (1992). *Chance encounters: Probability in education*. Dordrecht: Kluwer.

Kaput, J.J. (1987). Representation Systems and Mathematics. In C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics* (pp. 19-26). Hillsdale, New Jersey: Lawrence Erlbaum Associates.

Kaput, J.J. (1992). Technology and mathematics education. In D.A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 515-556). New York: MacMillan.

Keeler, C. M. and Steinorst, R. K. (1995). Using small groups to promote active learning in the introductory statistics course: A report from the field. *Journal of Statistical Education*, [Online], 3(2).

Kelly, A. E., Sloane, S. and Whittaker, A. (1997). Simple approaches to assessing underlying understanding of statistical concepts. In I. Gal and J. B. Garfield (Eds.), *The Assessment Challenge in Statistics Education* (pp. 85-90), Amsterdam: IOS Press.

Kinnear, J., Glesson, D. and Commerford, C. (1985). Use of concept maps in assessing the value of a computer-based activity in biology. *Research in Science Education*, *15*, 103-111.

Kirk, R. E. (1990). *Statistics*. (3rd Edn.). Fort Worth, Texas: Holt, Rinehart and Winston Inc.

Kirkwood, V., Symington, D., Taylor, P. and Weiskopf, J. (1994, November). *An alternative way to analyse concept maps*. Paper presented at the Contemporary Approaches to Research in Mathematics, Science and Environmental Education Conference, Deakin University.

Klein, P. (1994). *An easy guide to factor analysis*, London: Routledge.

Konald, C. (1989). Informal conceptions of probability. *Cognition and Instruction*, 6, 59-98.

Konald, C. (1991). Understanding student's beliefs about probability. In E.von Glaserfield (Ed.), *Radical Constructivism in Mathematics Education* (pp. 139-156). Netherlands: Kluwer Academic Publishers.

Konold, C. and Garfield, J. (1993). Statistical reasoning assessment. Part 1: Intuitive thinking, unpublished, SRRI, University of Massachusetts.

Konold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education*, [Online], 3(1).

Kreiger, H. and Pinter-Lucke, J. (1992). Computer graphics and simulations in teaching statistics. In F. and S. Gordon (Eds.), *Statistics for the Twenty-First Century* (pp. 198-206), MAA Notes No 26, Mathematical Association of America.

Laturno, J. (1994). The validity of concept maps as a research tool in remedial college mathematics. *Proceedings of the Sixteenth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, 2, 60-66.

Lesh, R., Post, T. and Behr, M. (1987). Representations and translations among representations in mathematics learning and problem solving. In C. Janvier (Ed.), *Problems of Representation in the Teaching and Learning of Mathematics* (pp. 33-40), Hillsdale, NJ: Lawrence Erlbaum Associates.

Lindley, D. V. (1990). Statistical inference. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* (pp. 33-41), Voorburg, The Netherlands: International Statistics Institute.

Lipson, K. (1995). *Assessing understanding in statistics*. Paper presented at the Fourth International Conference on Teaching Statistics, Marrakech, Morocco.

Lipson, K. (1996). What do students gain from computer simulation exercises? Examining the educational potential of computer based technology in statistics. In C. Batanero (Ed.), *Proceedings of the International Association for Statistical Education Round Table on Research on the Role of Technology in Teaching and Learning Statistics* (pp. 131-144). Spain: University of Granada.

Lipson, K., and Jones, P. (1995). *An introduction to statistics*. Melbourne: Swinburne University of Technology.

MacGillivray, H. (1998). Developing and synthesizing statistical skills for real situations through student projects. In L.Pereira-Mendoza, L. S. Kea, T. W. Kee and W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching of Statistics* (pp.1149-1156), Voorburg, The Netherlands: International Statistics Institute.

Magel, R. C. (1998). Using cooperative learning in a large introductory statistics class. *Journal of Statistical Education*, [Online], 6(3).

Markham, K. M. (1994). The concept map as research and evaluation tool: Further evidence of validity. *Journal of Research in Science Teaching*, 31(1), 91-101.

Marshall, S. P. (1995). *Schemas in problem solving*. Cambridge: Cambridge University Press.

Martin, P., Roberts, L., and Pierce, R. (1994). *Exploring statistics with Minitab*. Melbourne: Nelson.

McInerney, D. and McInerney, V. (1994). *Educational psychology: Constructing learning*. Sydney, Australia: Prentice Hall.

Mendenhall, W., Wackerly, D. D. and Scheaffer, R. L. (1990). *Mathematical statistics with applications*. (4th Edn.). Boston: PWS–Kent Publishing Company.

Minitab Version 8.1 (1992), Boston: *Minitab* Inc., computer program.

Moore, D. and McCabe, G. (1993). *Introduction to the practice of statistics*. (2nd Edn.). New York: W.H.Freeman and Company.

Moore, D. S., and McCabe, G. P. (1999). *Introduction to the practice of statistics*. (3rd Edn.). New York: W. H. Freeman and Company.

Moore, D.S. (1992a). Teaching statistics as a respectable subject. In F. and S. Gordon (Eds.), *Statistics for the Twenty-First Century* (pp. 14-25), MAA Notes No 26, Mathematical Association of America.

Moore, D.S. (1992b). What is statistics? In D.C. Hoaglin and D.S. Moore (Eds.), *Perspectives on Contemporary Statistics* (pp. 1-15), MAA Notes No 21, Mathematical Association of America.

Nitko, A.J. and Lane, S. (1990). Solving problems is not enough: Assessing and diagnosing the ways in which students organise statistical concepts. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* (pp. 467-474), Voorburg, The Netherlands: International Statistics Institute.

Noreen, E.W. (1989). *Computer intensive methods for testing hypotheses: An introduction*. New York: John Wiley and Sons.

Novak, J.D. (1990a). Concept mapping: A useful tool for science education. *Journal of Research in Science Teaching*, 27(10), 937-949.

Novak, J.D. (1990b). Concept maps and vee diagrams: two metacognitive tools to facilitate meaningful learning. *Instructional Science*, (19), 29-52.

Novak, J.D., and Gowin, D.B. (1984). *Learning how to learn*. Cambridge: Cambridge University Press.

Olson, C.L. (1987). *Statistics: Making sense of data*. Newton, Massachusetts: Allan and Bacon.

Ott, L. and Mendenhall, W. (1990). *Understanding statistics*. (5th Edn.), Boston: PWS–KENT Publishing Company.

Papert, S., and Turkle, S. (1992). Epistemological pluralism and the revaluation of the concrete. *Journal of Mathematical Behaviour*, 11(1), 2-33.

Pea, R. (1985). Integrating human and computer intelligence. In E. L. Klein (Ed.), *New Directions for Child Development: No 28, Children and Computers* (pp. 75-96). San Francisco: Jossey-Bass.

Pea, R.D. (1987). Cognitive technologies for mathematics education. In A.H. Schoenfeld (Ed.), *Cognitive Science and Mathematics Education* (pp. 89-122). Hillsdale, NJ: Lawrence Erlbaum Associates.

Pea, R.D. (1993). Practices of distributed intelligence and designs for education. In G. Salomon (Ed.), *Distributed Cognitions* (pp. 47-87). Cambridge: Cambridge University Press.

Pedler, P. J. (1991). Understanding sampling distributions in inferential statistics. *International Journal of Mathematics Education in Science and Technology,* 22(1), 69-87.

Peterson, R. F. and Treagust, D. F. (1989). Development and application of a diagnostic instrument to evaluate grade 11 and 12 students' concepts of covalent bonding and structure following a course of instruction. *Journal of Research in Science Teaching*, 26(4), 301-314.

Piaget, J. (1970). *The principles of genetic epistemology*. New York NY: Columbia University Press.

Pollatsek, A., Lima, S. and Well, A. D. (1981). Concept or computation: Student's understanding of the mean. *Educational Studies in Mathematics*, 12, 191-204.

Pukkila, T. and Putanen, S. (1986) The role of computers in the teaching of statistics. In R. Davidson and J. Swift (Eds.), *Proceedings of the Second International Conference on Teaching Statistics* (pp. 163-167), Canada: University of Victoria.

Putnam, R.T., Lampert, M. and Peterson, P.L. (1990). Alternative perspectives on knowing mathematics in elementary schools. In C. B. Cazden (Ed.), *Review of Research in Education* (pp.57-150). Washington, DC: American Educational Research Association.

Rennolls, K. and Massay, M. (1991). Discover the laws of sampling. *Teaching Statistics,* 13(3), 71-73.

Romberg, T.A. (1992). Perspectives on scholarship and research methods. In D.A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 49-64). New York: MacMillan.

Roth, W. and Roychoudhury, A. (1993). The concept map as a tool for the collaborative construction of cnowledge: A microanalysis of high school physics students. *Journal of Research in Science Teaching*, 30(5), 503-534.

Rubin, A. (1990) *Sampling Laboratory*, computer program. Unpublished.

Rubin, A., Bruce, B., Roseberry, A. and DuMouchel, W. (1988). *Getting an early start: Using interactive graphics to teach statistical concepts in high school*. Paper presented at the American Statistical Association Section on Statistical Education, New Orleans.

Rubin, A., Bruce, B. and Tenney, Y. (1990). Learning about sampling: Trouble at the core of statistics. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* (pp. 314-319), Voorburg, The Netherlands: International Statistics Institute.

Rubin, A.V., Rosebery, A.S. and Bruce, B. (1988). *ELASTIC and reasoning under uncertainty*(6851), computer program, BBN Systems and Technologies Corporation.

Salomon, G. and Globerson, T. (1987). Skill may not be enough: The role of mindfulness in learning and transfer. *International Journal of Educational Research*, 11(6), 623-638.

Salomon, G., Perkins, D.N. and Globerson, T. (1991). Partners in cognition: Extending human intelligence with intelligent technologies. *Educational Researcher, 20(3),* 2-9.

Schau, C. (1997). *Use of fill-in concept maps to assess middle school students' connected understanding of science*. Paper presented at the Annual Meeting of the American Education Research Association, Chicago, IL.

Schau, C. and Mattern, N. (1997). Assessing students' connected understanding of statistical relationships. In I. Gal and J. B. Garfield (Eds.), *The Assessment Challenge in Statistics Education* (pp. 91-104), Amsterdam: IOS Press.

Schoenfeld, A.H. (1994). Some notes on the enterprise (Research in collegiate mathematics education, that is). *Issues in Mathematics Education*, 4, 1-19.

Schoenfeld, A. H. (1987). What's all the fuss about metacognition. In A. H. Schoenfeld (Ed.), *Cognitive Science and Mathematics Education* (pp. 189-215), Hillsdale, NJ: Lawrence Erlbaum Associates.

Shaughnessy, J.M. (1992). Research in probability and statistics: Reflections and directions. In D.A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp.465-494). New York: MacMillan.

Shaughnessy, J.M. (1994). Cognitive snapshots of the stochastic river. *Journal for Research in Mathematics Education*, 25(1), 70-77.

Shavelson, R.J. (1993). *On concept maps as potential "authentic" assessments in science. Indirect approaches to knowledge representation of high school science*. Los Angeles, CA.: National Centre for Research on Evaluation, Standards, and Student Testing

Simon, J.L., Atkinson, D.T. and Shevokas, C. (1976). Probability and statistics: Experimental results of a radically different teaching method. *Mathematical Education*, November, 733-739.

Skemp, R.R. (1971). *The psychology of learning mathematics*. (1st. Edn.), Penguin.

Skemp, R.R. (1978). Relational understanding and instrumental understanding. *Arithmetics Teacher*, 1978(November), 9-15.

Skemp, R.R. (1987). *The psychology of learning mathematics*. (2nd Edn.), Penguin.

Soon, T. (1990). The computer spreadsheet: A versatile tool for the teaching of basic statistical concepts. In D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics* (pp. 187-192), Voorburg, The Netherlands: International Statistics Institute.

Starr, M.L. and Krajcik, J.S. (1990). Concept maps as a heuristic for science curriculum development: Toward improvement in process and product. *Journal of Research in Science Teaching*, 27(10), 987-1000.

Steffe, L.P. (1991). The constructivist teaching experiment: Illustrations and implications. In E.von Glaserfeld (Ed.), *Radical Constructivism in Mathematics Education* (pp. 177-194). Dordrecht: Kluwer Academic Publishers.

Steffe, L.P., and Thompson, P.W. (1996). Teaching experiment methodology: Underlying principles and essential elements. *Unpublished*.

Steinbring, H. (1992). The theoretical nature of probability in the classroom. In R. Kapadia and M. Borovcnik (Eds.), *Chance Encounters: Probability in Education* (pp. 45-63). Dordrecht: Kluwer Academic Publishers.

Steiner, G. (1994). From Piaget's constructivism to semantic network theory: Applications to mathematics education - a microanalyis. In R. Biehler, R. Scholz, R. Straber and B. Winkelmann (Eds.), *Didactics of Mathematics as a Scientific Discipline* . Dordrecht, The Netherlands: Kluwer Academic Publishers.

Stirling, D. (2000). *Computer assisted statistics teaching (Version 1.0)*, [Online], http://www-ist.massey.ac.nz/CAST/CAST.zip.

Stirling, D. W. (1995). *Statistical exercises using Models'n'Data (Version 2.3)*. Milton, Queensland: John Wiley and Sons.

Stirling, D. (1991) *Models'n'Data (Version 1)*. computer program. Santa Barbara: Intellimation.

Sweller, J. (1993). Our cognitive architecture and its consequences for teaching and learning mathematics. *Reflections*, 18(4), 1-16.

Tall, D. and Bakar, M. (1992). Students' mental prototypes for functions and graphs. *International Journal of Mathematics Education in Science and Technology*, 23(1), 39-50.

Tall, D.O. and Vinner, S. (1981). Concept image and concept definition in mathematics, with special reference to limits and continuity. *Educational Studies in Mathematics*, 12, 151-169.

Tanis, E.A. (1992). Computer simulations to motivate understanding. In F. and S. Gordon (Eds.), *Statistics for the Twenty-First Century* (pp. 217-225), MAA Notes No 26, Mathematical Association of America.

Thisted, R.A., and Velleman, P.F. (1992). Computers and modern statistics. In D.C. Hoaglin and D.S. Moore (Eds.), *Perspectives on Contemporary Statistics* (pp. 41-53), MAA Notes No 21, Mathematical Association of America.

Thomas, D.A. (1984). Understanding the central limit theorem. *Mathematics Teacher*(October), 542 - 543.

Thomason, N., Cumming G. and Zangari, M. (1994) Understanding central concepts of statistics and experimental design in the social sciences, In K. Beattie, C. McNaught and S. Wills (Eds.) *Interactive Multimedia in University Education: Designing for Change in Teaching and Learning*. Amsterdam: North-Holland.

Thomason, N. and Les, J. (1998), Towards 2000: Reform in research practice and statistical education, In L.Pereira-Mendoza, L. S. Kea, T. W. Kee and W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching of Statistics* (pp.695-700), Voorburg, The Netherlands: International Statistics Institute.

Tobias, S. (1994). *Overcoming math anxiety*. New York: Norton.

Tversky, A., and Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin*, 76(2), 105-110.

Tversky, A. and Kahneman, D. (1982). Judgement under uncertainty: Heuristics and biases. In D. Kahneman, P. Slovic, and A. Tversky (Eds.), *Judgement under uncertainty: Heuristics and biases* (pp. 3-20). Cambridge: Cambridge University Press.

Vinner, S. (1983). Concept definition, concept image and the notion of function. *International Journal of Mathematics Education in Science and Technology*, 14(3), 293-305.

von Glaserfeld, E. (1987). Learning as a constructive activity. In C. Janvier (Ed.), *Problems of representation in the teaching and learning of mathematics* (pp.3-17). Hillsdale, NJ: Lawrence Erlbaum Associates.

Vygotsky, L.S. (1978). *Mind in society*. Cambridge, Massachusetts: Harvard University Press.

Wallace, J.D. and Mintzes, J.J. (1990). The concept map as a research tool: Exploring conceptual change in biology. *Journal of Research in Science Teaching*, 27(10), 1033-1052.

Weldon, L. (1986). *Statistics: A conceptual approach*. Englewood Cliffs, New Jersey: Prentice Hall.

Weiss, N., and Hassett, M. (1987). *Introductory statistics*. (2^{nd} Edn.). Reading, Massachusetts: Addison-Wesley.

Well, A.D., Pollatsek, A. and Boyce, S.J. (1990). Understanding the effects of sample size on the variability of the mean. *Organizational Behaviour and Human Decision Processes*, 47, 289-312.

White, R. and R. Gunstone (1992). *Probing understanding*. London: The Falmer Press.

Wilkox, S.K. (1998). Another perspective on concept maps: Empowering students. *Mathematics Teaching in the Middle School*, 3(7), 464-469.

Williams, A. (1993). Is a pass good enough in tertiary statistics?. In W.Atweh (Ed.) *Proceedings of the Mathematical Education Research Group of Australia 16th Annual Conference* (pp.587-591), Brisbane: MERGA.

Williams, A. M. (1998). Students' understanding of the significance level concept, In L.Pereira-Mendoza, L. S. Kea, T. W. Kee and W. K. Wong (Eds.), *Proceedings of the Fifth International Conference on Teaching of Statistics* (pp.743-749). Voorburg, The Netherlands: International Statistics Institute.

Williams, C. G. (1995). *Concept maps as research tools in mathematics*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Yin, R. K. (1989). *Cast study research: design and methods* (Revised Edn.), California: Sage.

Appendix 1 Marking Schemes used in the preliminary studies

Appendix 1A: The *t*-test Task

Step	Mark allocation
Identify test	1
Hypotheses in μ_d	1
Equal to 0	0.5
Alternate directional	0.5
$\bar{d}$	1
s_d	1
t- formula substituted	1
t-value	1
degrees of freedom	1
P-value	1
Decision	1
Conclusion	1

Appendix 1B: The Explanation Task

Steps in your hypothesis test	Explanation
Hypotheses: H_0: $\mu = 0$ H_1: $\mu < 0$	Some clear indication that we are trying to decide if the reduction in nicotine content observed in the sample might be generalisable to the population. (1 mark) We are to choose between two alternatives, one which states the reduction is real, the other that it is explained by sampling variability. (1 mark)
directional test	Comment that we are assuming only a reduction is possible. (1 mark)
Significance level: $\alpha = 0.05$	The probability of incorrectly concluding that a relationship exists in the population when it does not has been set to a maximum of 5%. or We will conclude that there is a relationship in the population if the probability of the result observed in the sample occurring by chance alone is no more than 5%. (2 marks)
Test statistic $z = \dfrac{\bar{x} - \mu}{\sigma / \sqrt{n}}$	The decision will be made on the basis of the size of the reduction observed in the sample, standardised to a number of standard deviations below the mean. (1 mark)
Calculate the value of the test statistic $\bar{x} = 13.0$ $\mu = 14$ $\sigma = 4$ $n = 64$ $z = \dfrac{13.0 - 14}{4/\sqrt{64}} = -2.0$	Substituting the values for the sample mean, the hypothesised mean, the standard deviation and the sample size we find that a sample mean of 13.0 is two standard deviations below the accepted mean of 14.
P-value P-value $= P(z < 2.0)$ $= 0.0228$ or 2.28%	We need to determine how likely a sample mean of 13.0 is if the population value is 14. (1 mark) From the tables we can see that such a value is not likely (occurs less than 2.28% of the time). (1 mark)
Decision & conclusion As $p < 0.05$, reject H_0 and conclude that there is has been a reduction in nicotine levels.	Since we would be very unlikely to get a sample mean of 13.0 if there really is no reduction in nicotine content, we will conclude that sampling variability is not the reason for this reduction. (1 mark) Rather, we think that the new process is really an improvement. (1 mark)

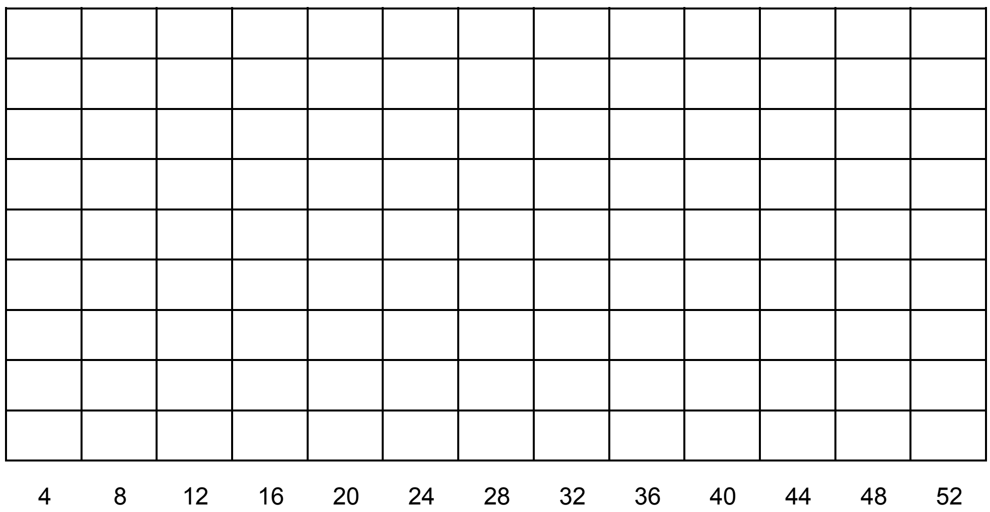
Appendix 2 Sampling Activity

1 What is the proportion of coloured beads in the box? Have a guess.
proportion of coloured beads in box =

2 Select some beads from the box (without looking) using the sampling shovel, replacing the selected bead each time so that the number of beads in the box remains constant.
number of coloured beads in sample =

3 For your own sample(s) of twenty-five beads, calculate the proportion $\hat{p}$ of coloured beads in the sample.
proportion of coloured beads in sample =

4 Collect the values of sample proportion $\hat{p}$ from every person and display the distribution of the proportion of coloured beads in the samples in a histogram.



5 Describe the pattern of the distribution of the sample proportion, in particular noting the shape, centre and spread.

Appendix 3 Prior Knowledge Questionnaire

It would help us to have some indication of the extent of your prior studies in statistics. Please tell us how familiar you are with the following terms, by ticking one of the statements given: never heard of it, heard of it (but couldn't actually tell you anything about it), some knowledge (but fairly sketchy), or feel you have a reasonable understanding of the term.

	never heard of it	heard of	some knowledge	understand
alternate hypothesis				
chi square test				
confidence interval				
critical value				
hypothesis testing				
level of significance				
normal distribution				
null hypothesis				
p-value				
parameter				
population				
probability				
random sampling				
randomness				
sample				
sample statistic				
sampling distribution				
sampling variability				
standard error				
standardisation				
statistical estimation				
statistical inference				
survey				
type 1 error				
t–test				
z–score				
z–test				

Appendix 4: Concept maps, week prepared and terms suggested

Map	Week	Terms
Sampling distribution for the sample proportion	9	centre computer generated constant distribution normal model p $\hat{p}$ population population parameter population proportion repetitions sample proportion sample size sample statistic sample(s) sampling distribution sampling variability spread variable
Sampling distribution for the sample mean	10	centre computer generated constant distribution normal model μ $\bar{x}$ population population mean repetitions sample mean sample size sample statistic sample(s) sampling distribution sampling variability spread variable

Map	Week	Terms
Sampling distribution	11	centre constant μ p $\hat{p}$ population population parameter r ρ sample size sample statistic sample(s) sampling distribution spread variable $\bar{x}$
Hypothesis testing	12	alternate decision hypotheses null P-value population population parameter sample statistic sample(s) sampling distribution sampling variability significance level test statistic
Estimation	13	confidence interval estimation interval estimates point estimates population sample sample statistics sampling distribution

Map	Week	Terms
Statistical Inference	15	confidence interval decision & conclusion estimation hypotheses hypothesis testing inferential statistics interval estimates P-value point estimates population sample sample statistics sampling distribution significance level statistical significance test statistic

Appendix 5 Marking Schemes used in the main study

Appendix 5A: The *t*-test task

Step	Mark allocation
Hypotheses in μ	1
Equal to 3.6	0.5
Alternate directional	0.5
$\bar{\boldsymbol{x}}$	1
s	1
t- formula substituted	1
t-value	1
degrees of freedom	1
P-value	1
Decision	1
Conclusion	1

Appendix 5B: The sample size task

Step	Mark allocation
Formula	1
Substitution in formula	1
Correct answer	1

Appendix 5C: Two group *t*-test task

Step	Mark allocation
Hypotheses in μ_d	1
Equal to 0	0.5
Alternate directional	0.5
t-value	1
P-value	1
Decision	1
Conclusion	1

Appendix 5D: The Modelling Task

Step	Mark allocation
Part (a)	
Hypotheses in *p*	1
Equal to 0.5	0.5
Alternate non-directional	0.5
$\hat{p} = 0.61$	1
P-value from sampling distribution	1
P-value correct	1
Decision	1
Conclusion	1
Part (b)	
Hypotheses in *p*	1
Equal to 0.5	0.5
Alternate non-directional	0.5
$\hat{p} = 0.61$	1
z-value	1
P-value	1
Decision	1
Conclusion	1

Appendix 5E: The Correlation Task

Step	Mark allocation
Part (a)	
Hypotheses in ρ	1
Equal to 0	0.5
Alternate non-directional	0.5
P-value	1
Decision	1
Conclusion	1
Part (b)	
Calculate r^2	1
Interpret r^2	1
Comment on practical significance	1

Appendix 5F: The Chi-square Task

Step	Mark allocation
Correct null hypothesis	1
Correct alternate hypothesis	1
Choice of test	1
Expected values	1
Chi-square value	1
Degrees of freedom	1
P-value	1
Decision	1
Conclusion	1

Appendix 5G: The Unknown Test Task

Step	Mark allocation
Hypotheses in σ	1
Hypotheses correct	1
Decision	1
Conclusion	1

Appendix 5H: The Explanation Task

Steps in your hypothesis test	Explanation
Hypotheses: H_0: $\rho = 0$ H_1: $\rho \neq 0$	Some clear indication that we are trying to decide if the relationship observed in the sample might be able to be generalised to the population. (1 mark) We are to choose between two alternatives, one which states there is no relationship, the other which states there is a relationship (1 mark)
non-directional test	Comment that intelligence might be higher or lower for mother and eldest child. (1 mark)
Significance level: $\alpha = 0.05$	The probability of incorrectly concluding that a relationship exists in the population when it does not has been set to a maximum of 5%. or We will conclude that there is a relationship in the population if the probability of the result observed in the sample occurring by chance alone is no more than 5%. (2 marks)
Test statistic $r = 0.5135$ for $n = 30$ pairs of data values	The decision will be made on the basis of the strength of the relationship observed in the sample. This is measured with a statistic called Pearson's *r*. (1 mark)
***P*-value** P-value $= 2 \times P(r > 0.5)$ $= 2 \times 0.0025$ $= 0.005$ or 0.5%	We need to determine how likely a sample value of 0.5135 is if the population value is 0. (1 mark) From the tables we can see that such a value occurs only rarely. (1 mark)
Decision & conclusion As $p < 0.05$, reject H_0 and conclude that there is a relationship between the intelligence of children and their mothers in the general population.	Since we would be very unlikely to get 0.5135 if there really is no relationship, we will conclude that sampling variability is not the reason for the high correlation. (1 mark) Rather, we think that there is a correlation between these two variables in the population. (1 mark)

Appendix 5H: The Radio Task

Mark	Criteria
1	Student recognises that the problem could be with data collection, in that the sample is not random.
1	An example of the way the data collection problem could have occurred is given.
1	The student recognises alternatively that there may be no problem with the data collection, the difference may be due to sampling variability (eg mention p and $\hat{p}$).
1	Students recognises that there is a need to know the sample size before any conclusions may be drawn.
1	The student uses these arguments to discuss the hypothesis in the question.

Appendix 6 Original Student concept map and schematic representation of that map.

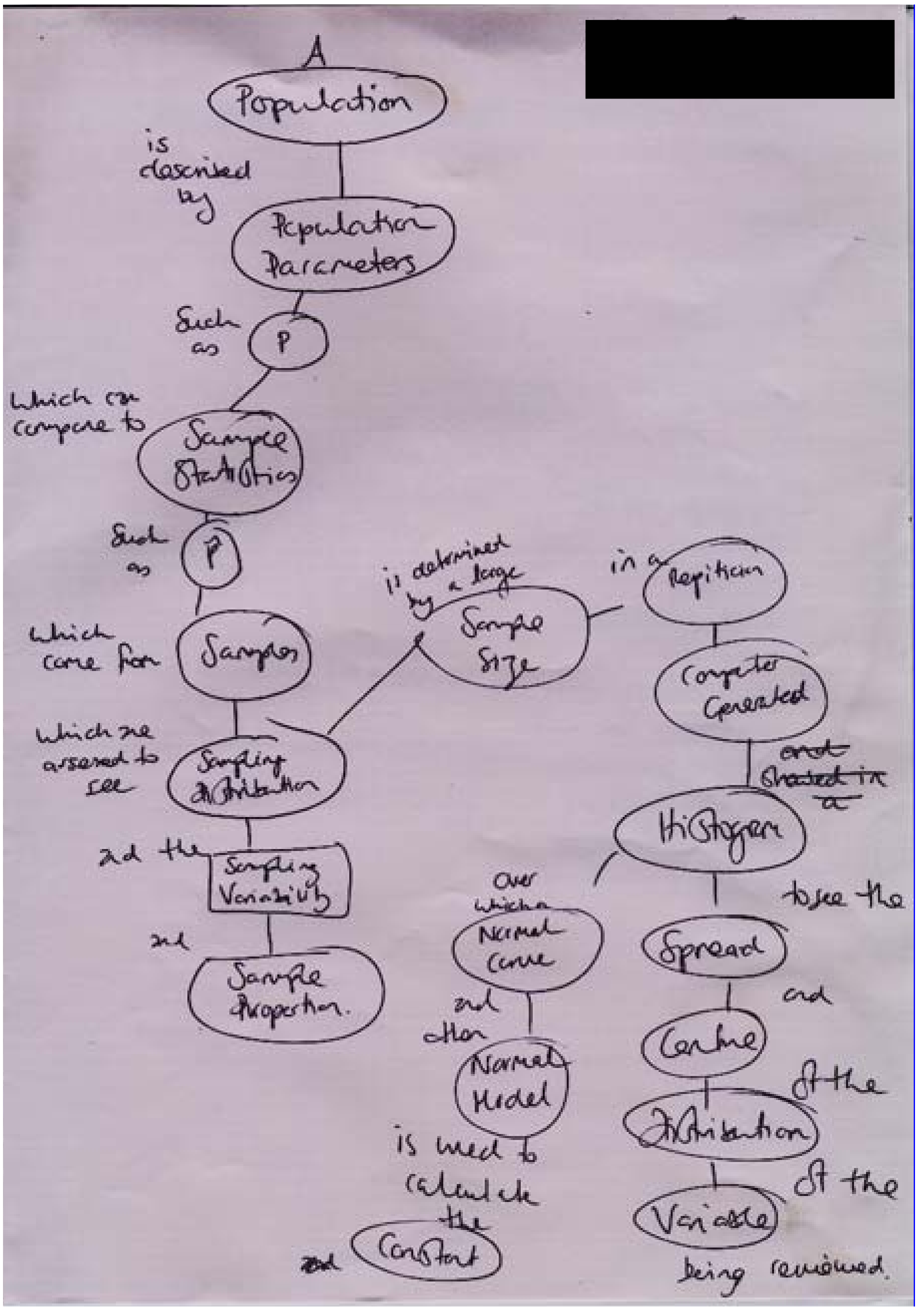


Figure A1: Concept map prepared by a student for the sampling distribution of the sample proportion

Appendix 6 (continued)

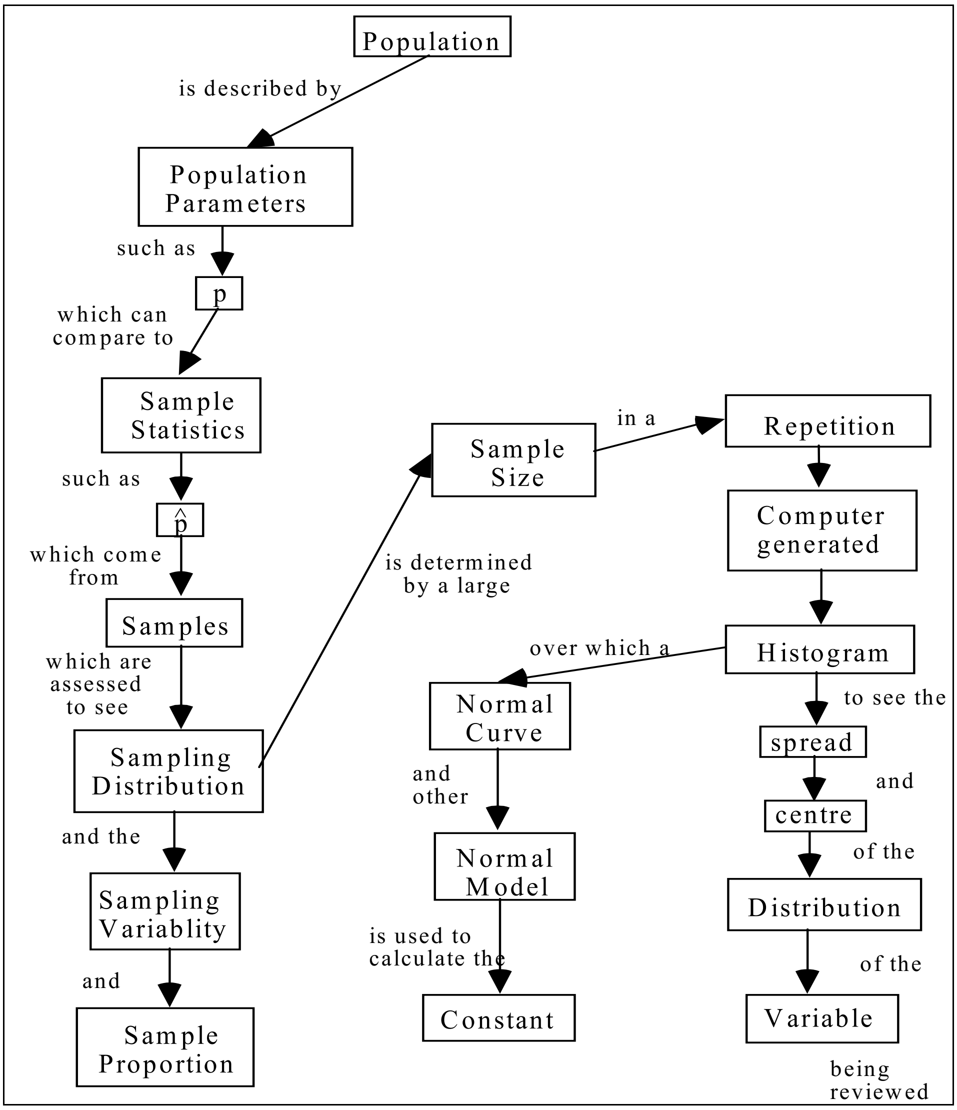


Figure A2: Schematic representation of the concept map shown in Figure A1

Appendix 7 Group differences on actual scores

The means and standard deviations of the actual score for the procedural and conceptual understanding measures for each of the three groups are shown in the table below.

Summary Statistics for the Procedural and Conceptual understanding actual scores for each group

Group		Procedural understanding		Conceptual understanding	
	N	Mean	SD	Mean	SD
Sampling distribution concept developed and linked to inference	7	43.7	2.4	12.1	3.7
Sampling distribution concept developed but not linked to inference	10	40.0	11.5	9.1	2.3
Sampling distribution concept not well developed	6	35.7	12.9	5.1	2.6

As expected, analysis of group differences with the actual scores (as opposed to the factor score used in the previous analysis) showed the same results as before (see Section 5.5). A oneway Analysis of Variance revealed that there was a significant difference in the scores in conceptual understanding between the three groups, $F(2,20) = 9.836$, $p = 0.001$. Planned comparisons showed that the mean score for conceptual understanding in Group 1 was significantly higher than that for Group 2, $t(20) = 2.156$, $p = 0.022$, and that the mean score for conceptual understanding in Group 2 was significantly higher than that for Group 3, $t(20) = 2.717$, $p = 0.007$. A further oneway Analysis of Variance showed that the differences in the scores in procedural understanding were not statistically significant, $F(2,20) = 1.013$, $p = 0.381$.

Appendix 8 List of Publications

Jones, P. and Lipson, K. (1994a), Introducing the Logic of Statistical Inference through Computer Intensive Methods. In *Proceedings of the Fourth International Conference on Teaching Statistics*, International Statistics Institute, Marrakech, pp. 130-137.

Jones, P., Lipson, K. and Phillips, B. (1994), A Role for Computer Intensive Methods in Introducing Statistical Inference. In L. Brunelli and G. Cicchitelli (Eds.), *IASE: Proceedings of the First Scientific Meeting* (pp.255-263). Perugia, Italy: University of Perugia.

Jones, P. L. and Lipson, K. L. 1993, Determining the educational potential of computer based strategies for developing an understanding of sampling distributions. In W. Atweh (Ed.), *Proceeding of the Mathematical Education Research Group of Australia 16th Annual Conference*, Mathematical Education Research Group of Australia, Brisbane, pp. 355-360.

Lipson, K. (1992), Exploring Statistical Concepts with Visualisation Software. In M. Horne and M. Supple (Eds.), *Mathematics: Meeting the Challenge*, Mathematics Association of Victoria, Melbourne, pp. 406-409.

Lipson, K. (1994), Understanding the Role of Computer Based Technology in Developing Fundamental Concepts of Statistical Inference. In *Proceedings of the Fourth International Conference on Teaching Statistics*, International Statistics Institute, Marrakech, pp. 65-72.

Lipson, K. (1995), Assessing Understanding in Statistics. In J. Garfield (Ed.), *Fourth International Conference on Teaching Statistics: Collected Research Papers*.

Lipson, K. (1996). What do students gain from computer simulation exercises? Examining the Educational Potential of Computer Based Technology in Statistics. In C. Batanero (Ed.), *Proceedings of the International Association for Statistical Education Round Table on Research on the Role of Technology in Teaching and Learning Statistics* (pp. 131-144). Spain: University of Granada.

In those papers published jointly with my supervisors Prof Peter Jones and Mr Brian Phillips, they fulfilled a normal supervisory role in commenting and contributing to the conceptual framework which I had established.