Assessment of the Understanding of Statistical Concepts

Flavia R Jolliffe - Surrey, England

1. Introduction

Much has been written about methods of teaching statistics and about how to assess students' knowledge of statistics, but almost nothing on the extent to which assessment procedures measure whether students understand statistical concepts, or what is involved in the application of techniques, and little research has been done on the development of instruments to measure statistical understanding. This paper reviews work in related areas.

Measurement of understanding ideally forms part of most assessment procedures, but it can be hard to isolate the understanding of concepts from knowledge of basic definitions and from computations. A student who has produced a correct "solution" does not necessarily understand the solution, or even the "question" behind it. It should be possible to assess understanding through the practical skills of doing and communicating statistics.

2. Classification schemes

When developing a syllabus, a method of instruction, and associated assessment methods, it is useful to draw up a framework of objectives. A 5-point classification scheme of teaching objectives for mathematics (see Wood, 1968) adapts readily to statistics. The five objectives are: (1) Knowledge and information, (2) Techniques and skill, (3) Comprehension, (4) Application, and (5) Inventiveness. This scheme is similar to Bloom's taxonomy of educational objectives in the cognitive domain. The main categories are supposed to be ordered along a simple to complex dimension and higher objectives build upon and depend on lower objectives.

Knowledge and information means basically the recall or recognition of such items as definitions, notation, and facts, for example quoting the formula for a standard deviation, or being aware that the median is a measure of location. Answers to questions testing knowledge depend mainly on memory and can often be copied from textbooks. Techniques and skills includes computation, algebraic manipulation, and construction of diagrams. Examples are maximising a likelihood function, drawing a box plot, and use of a statistical computer package. No understanding is needed at these lowest two levels. Even skills can be demonstrated in inappropriate situations as in the calculation of an arithmetic mean of numerical codes given to categorical data.

Comprehension is not complete understanding. There are three types of Comprehension behaviour - Translation, Interpretation, and Extrapolation (which includes Interpolation). Translation is an activity requiring the change of form of a communication, such as describing the contents of a table of data in words or vice versa, or being able to interpret a diagram at a descriptive level. Interpretation involves a re-arrangement of material, for example, "making inferences from data presented in tabular or graphic form" (Wood, 1968). Extrapolation is an extension of Interpretation to include statements about the consequences of a communication.

Application is selecting and using appropriate techniques, and to be successful, all of the lowest three levels are needed. However, testing whether a student has attained the required level of Application creates some problems, as there is a limit to the number of genuinely new situations which are within reach of a student's general level of attainment. Doing questions is a recognised teaching method and a necessary part of learning, but if the application has been met before, the student might obtain a solution by recall.

Inventiveness involves a student in making discoveries, and perhaps improvising. It might require an approach which is new to the student as opposed to application where previously-met methods are applied in new situations. It can be demonstrated by a student who extends practical and project work beyond the basic requirements, or by one who finds out about topics in which formal instruction has not been given. Inventiveness is the highest level of skill and has to involve understanding of concepts.

As a step in finding out what qualities were being cultivated in undergraduate mathematics courses, Griffiths and McClone (1979) analysed a sample of timed examination papers set in 1976 in 10 British universities. They used ten qualities: (1) Procedure, (2) Objectives, (3) Jargon, (4) Mathematical content other than bookwork, etc., (5) Definitions, Bookwork, Stock example, (6) Abstract/ Concrete, (7) Mathematical manipulation, (8) Logical manipulation, (9) Sustained thinking, and (10) Openness of solution.

The two qualities which come nearest to consideration of understanding are *logical manipulation*, that is the powers of reasoning required for the successful solution, and *open solution* which measures the extent to which a question determines the nature of the solution. Questions with such instructions as "Analyse the data" or "Discuss your conclusions" have a high degree of openness. Note that any classification of examination questions is subjective to some extent, and ideally has to be relative to the preparation students have received and to the solutions they produce. Even the open statement "Analyse the data" might well be recognised by candidates as their lecturer's way of asking for a specific set of routines.

Of the subject areas considered by Griffiths and McClone, they noted that

Statistics assesses differently from the norm. In the sample of 228 statistics questions many were "applicable", there was a relatively rare use of bookwork and stock examples, and more was called for in the way of mathematical content and understanding in setting up the question. Candidates were in the main not guided through the solution by being asked to carry out a succession of small steps each requiring only limited thought.

A model of the problem-solving process central to statistical reasoning with three stages, (1) Comprehension, subdivided into six steps, (2) Planning and execution (two steps), and (3) Evaluation and interpretation (two steps), is described by Chervany et al. (1977). Comprehension refers to recognition of the task required. The two steps in Evaluation and interpretation are, verifying the solution from knowledge of similar problems, and, stating results using paraphrases. Part of Chervany et al.'s concern was the need of a model for constructing measurement instruments to determine the degree to which the teaching objectives of introductory statistics courses have been achieved, and, insofar as courses are evaluated partly by assessing students' performance, this model should be useful also when considering the assessment of students' understanding of statistical concepts.

3. Newer methods of teaching and assessment

In recent years there have been many exciting suggestions for teaching statistics in ways that are likely to increase students' understanding and which have been successfully implemented in the classroom, but development of corresponding methods of assessment has been slow.

Replacement of timed written examinations by alternative methods of assessment can improve assessment of understanding. An example is using a six-week case study in an introductory statistics course (Mortensen, 1988). Griffiths and McClone (1979) give examples of two mathematics courses assessed entirely by coursework. Some aspects which could be adapted for statistics are group work on a "real-life" problem carried out over several months, and giving reference to reading material on extended essay topics selected by the students.

The author has some experience of teaching social scientists an introductory statistics course graded on a pass/fail basis and assessed entirely by coursework (Jolliffe, 1976). To obtain a pass grade on an assignment a student had to demonstrate at least one of "capability of using statistical techniques" and "thought as regards the meaning of the data", higher grades going to those who demonstrated both. Students were set either an essay or an open-ended assignment every two to three weeks. Each piece of work was assessed individually. There were never more than twenty-five students in a group so that marking was not a burden.

Examples of essay titles are "Is defence a major item of government expenditure?" and "A study of trade union membership". In the open-ended assignments students were usually given a table of data, for example a set of income distributions for several social class groups, and an instruction of the type "Discuss fully, using diagrams and summary measures as appropriate". Students were given some guidance as to the kinds of things which were expected. They had to have some understanding in order to do assignments of this nature successfully, and a real benefit was that they were motivated to learn and to do statistics.

4. Attitude scales

An alternative to asking students statistics questions to find out the extent of their understanding is to explore a variety of dimensions related to understanding by means of an attitude survey. For example, the strength of students' feelings concerning whether or not they are happy about the statistics course, whether lectures make them feel inadequate, whether they can cope with the coursework, and so on, can be investigated.

As part of a statistics project done under the author's supervision, Kenney (1989) conducted a survey among Brunel University students. She adapted statements from attitude scales relating to various subjects of study and wrote some to form a set of twenty statements, ten favourable to statistics and ten unfavourable, to investigate attitudes towards statistics. A five-point scale ranging from strongly agree to strongly disagree was used. Examples of statements relating to understanding are *Often, when I solve a statistics problem, I am uncertain how to interpret the results; I find it easy to explain statistical problems to someone else;* and *I find statistical problems confusing.* In the random sample of 164 undergraduates, most had favourable attitudes towards statistics and, the more statistics they had used or studied, the more favourable the attitude.

5. Studies of understanding

There have been only a few studies of the understanding of statistical concepts and some of these have used tests containing mainly probability questions. There are many good probability concepts questions, but methods for probing into responses to find out something about understanding are not practicable for assessment in a typical classroom situation. Some of the questions used to investigate whether people adjust correctly in situations involving, for example, a weighted mean or regression towards the mean, are suitable for testing understanding.

The papers by Nisbett et al. (1983) and Fong et al. (1986) on thinking about everyday problems contain some excellent examples, based on scenarios, of how to ask about statistical understanding in ways that do not depend on knowledge or require the application of techniques. However, it is not obvious how to grade responses to these problems, and in assessment of students some testing of knowledge and skills would usually be required as well. Many of the newer textbooks include some open-ended questions where some thinking is required and which would test understanding.

Kenney's questionnaire (1989) contained the following four questions designed to assess intuitive ideas about statistics and to some extent understanding; the correct answer is denoted by an asterisk.

1. To avoid biasing the results of a survey, a sample selected from a population should be
   (a) Large
   (b)* Random
   (c) Representative
   (d) Don't know

2. Conclusions drawn from sample data about a population are subject to uncertainty because
   (a) The data are not reliable
   (b) Calculations are not accurate
   (c)* Only part of the population is available
   (d) Don't know

3. In doing a statistical test we
   (a)* Draw conclusions about populations from sample data
   (b) Draw conclusions about populations and then collect sample data to support these conclusions
   (c) Collect sample data and use the data to make assumptions about a population
   (d) Don't know

4. In a town every family with six children was surveyed. In 70 families the pattern of boy (B) and girl (G) births was GBGBBG. What is your estimate of the number of families surveyed in which the order of births was BGBBBG?
   (a) Less than 70
   (b)* 70
   (c) More than 70
   (d) Don't know

Respondents were told that no calculations were required and they simply had to tick the one answer they thought was most correct, and they were asked not to look up the answers. Asking for the most correct answer covers both the possibility that respondents might feel there is more than one correct answer and that there might be no correct answer. In Question 3, which was the least satisfactory, the modal response was (c) which was incorrect. This could have been due to poor wording or responses might have been affected by the previous question. In Question 4 23% answered "Don't know". Perhaps they thought more information was needed, or that a calculation was required. Many respondents asked for an explanation of this question.

6. Conclusions

The large classes faced by many statistics lecturers undoubtedly influence both the teaching and assessment procedures in undesirable ways. When the group of students is large it is hard to interact with students and to find out informally if they understand concepts, and it is difficult to implement those methods of teaching which are thought to teach understanding but which really require a small-group approach. When it comes to assessment of students, lecturers may well prefer to set questions at the levels of Knowledge and information and Techniques and skill which can be marked quickly in semi-automatic mode. Such questions rarely assess understanding and when used in unsupervised assessments there is a risk that students might copy one another. There is a real need for the development of questions which assess understanding but where the marking can be kept within reasonable bounds.

References

Chervany, N L, Collier, R O Jr, Fienberg, S E, Johnson, P E and Neter, J (1977) A framework for the development of measurement instruments for evaluating the introductory statistics course. *The American Statistician* **31**, 17-23.

Fong, G T, Krantz, D H and Nisbett, R E (1986) The effects of statistical training on thinking about everyday problems. *Cognitive Psychology* **18**, 253-292.

Griffiths, H B and McClone R R (1979) *Qualities Cultivated in Mathematics Degree Examinations.* A research report for the Social Science Research Council.

Jolliffe, F R (1976) A continuous assessment scheme for statistics courses for social scientists. *International Journal of Mathematical Education in Science and Technology* **7**, 97-103.

Kenney, T (1989) *Attitude Scale Construction and a Survey of Attitudes Towards Statistics of Students at Brunel University.* Undergraduate project, Brunel University, Department of Mathematics and Statistics.

Mortensen, P S (1988) On a new approach to an introductory course in statistics for business and economics including some experiences. In: R Davidson and J Swift (eds) *Proceedings of the Second International Conference on Teaching Statistics.* University of Victoria, Canada, 415-424.

Nisbett, R E, Krantz, D H, Jepson, C and Kunda, Z (1983) The use of statistical heuristics in everyday reasoning. *Psychological Review* **90**, 339-363.

Wood, R (1968) Objectives in the teaching of mathematics. *Educational Research* **10**, 83-98.