A PROPOSAL FOR INVESTIGATING STUDENTS' KNOWLEDGE OF HYPOTHESIS TESTING

ANNE WILLIAMS
Queensland University of Technology

Hypothesis testing is conceptually complex. It builds on a basic knowledge of probability, sampling distributions and the central limit theorem. It includes new concepts such as null/alternative hypothesis, alpha level, Type I and II errors, power, sample size, statistical tests, critical value, p-value, observed value and decisions about hypotheses. The web of interrelationships between these concepts requires their holistic understanding for rational decision making. Firstly, this paper briefly summarises how these concepts may be linked together; and secondly, presents a methodology for research into what students know in these areas, their understanding of simple design concepts, and the sources of their difficulties with hypothesis testing.

Hypothesis testing is a topic in almost all first year university statistics courses. Experience has shown that students often have a great deal of trouble with it. Given the large numbers of students studying such courses from a variety of disciplines, and the importance in research of statistical techniques using hypothesis testing, it is essential to know what students know about hypothesis testing, so that they may be helped, if necessary, through improved instruction or remedial assistance.

While much has been written about hypothesis tests over the past 40 years, little empirical research has been undertaken. Despite generating great controversy, hypothesis tests remain an integral part of traditional research. Many critics accept that the problem lies, not in the tests themselves, but in how the tests are used and interpreted (e.g., Pedhazur & Schmelkin, 1991), and this is a function of the knowledge that the user possesses. The literature on hypothesis testing has implied that many of the difficulties encountered by students are attributable to problems with definitions, notation and terminology; failure to understand prerequisite concepts; the counter-intuitive ideas; the influence of some journals and text books in perpetuating misuse; and underestimation of the complexities.

This paper reports on a study which aims to: (1) summarise the main prerequisite skills, procedures and concepts required for elementary hypothesis testing; (2) identify the main procedures and concepts used in elementary hypothesis testing, and their interrelationships; (3) describe students' knowledge, understanding and difficulties in (1) and (2); and (4) investigate students' ability to make simple design decisions before and after instruction. In the context of this paper, elementary hypothesis testing refers to the concepts and procedures associated with tests of means in parametric statistics. This paper presents a proposal for a methodology for studying the above aims.

HYPOTHESIS TESTING

Typically an introductory tertiary subject in statistics includes descriptive and inferential components, the latter including hypothesis testing. Probability provides a link between the two components. Research studies (e.g., Fischbein, Nello & Marino, 1991; Hansen, McCann & Myers, 1985; Kahneman & Tversky, 1982; Konold, Pollatsek, Well, Lohmeier, & Lipson, 1993; Shaughnessy, 1983) have illustrated the problems and misconceptions that students have in probability. Sometimes these are the result of faulty intuitions which are deep-seated and difficult to change (Shaughnessy, 1983). Alternatively, they may be due to the abstract nature of probability, and difficulties with quantitative and hypothetical reasoning, which in turn may be complicated by their interaction with affective variables (Garfield & Ahlgren, 1988).

The concepts of probability and probability distributions lay the foundation for sampling distributions, with which the central limit theorem [CLT] is closely allied. The CLT has been described as "the cornerstone of statistical inference" (Glencross, 1986, p. 91), and is essential for understanding elementary hypothesis testing. Many articles attest to the difficulty of understanding the concepts of the CLT and sampling distributions (see Glencross, 1986; Johnson, 1986; Konold, Well, Lohmeier & Pollatsek, 1993; Mittag, 1992; Zerbolio, 1989). The literature implies that learning about sampling distributions and the CLT may be improved by teaching via demonstration or simulation: Jones and Lipson (1993) suggest the use of a computer programme which has a dynamic display of the ongoing sampling process. Given that the traditional introduction to the CLT and sampling distributions has been via probability theory, and with the misconceptions that abound in probability, there is reason for concern when students attempt the topic of hypothesis testing without understanding sampling distributions and the CLT.

Some concepts that are associated with hypothesis testing, such as representativeness and randomness, are also associated with sampling design, which will not be discussed in detail here. Others, such as mean and standard deviation, so vital for hypothesis testing, are normally introduced in descriptive statistics, and are assumed to be "known" thereafter. Again, there is evidence that students experience difficulty with these latter two concepts (see Cassell, 1989; Hart, 1984; Loosen, Lioen, & Lacante, 1985; Mevarech, 1983; Pingel, 1993; Pollatsek, Lima & Well, 1981), especially when knowledge beyond substitution into a formula is required. Thus the main prerequisite skills and concepts of interest in this study centre around the mean, the standard deviation, probability, sampling distributions and the CLT, because the literature reveals them as problem areas.

There are four main approaches taken to elementary hypothesis in introductory statistics text books. They are: (i) the critical value method, where the observed and critical values are compared; (b) the p-value method, where the p-value and alpha value are compared; (c) the confidence interval method (conceptually different from the previous two), requiring the construction of an interval in which the mean is expected to belong if the null hypothesis is true; and (d) the Monte Carlo method, where the number of samples with means that refute the null hypothesis is counted, expressed as a probability, and judged as supportive of the null hypothesis or not.

The difficulties of understanding hypothesis testing procedures may be demonstrated by the large number of concepts involved, and their interconnections. For example, the selection of hypotheses entails distinguishing between samples and populations, statistics and parameters, one-tailed and two-tailed tests, and understanding the mean, the null and alternative hypotheses, and the appropriate statistical notation. Additionally, decisions about hypotheses effect the critical value, which is also dependent on the sample size and the level of significance. Underlying these concepts are sampling and design issues such as representativeness, randomness, and the type and method of sampling, on which probability theory depends.

Introductory statistics text books generally discuss the following main concepts in conjunction with elementary hypothesis testing: null hypothesis, alternative hypothesis, alpha, Type I error, Type II error, power, sample size, statistical test (formula), critical value, p-value, observed value, reject/do not reject H_0 and decision. As illustrated above, understanding one of these concepts entails understanding many others, because of the web of interrelationships.

Hence hypothesis testing is an extremely complex process. The conceptual interdependencies involved make it a potentially difficult topic for novices. The methodology described below is one approach to discovering the extent of students' knowledge about these procedures, concepts, and interrelationships.

METHODOLOGY

Selection of Subjects

The study will be based on 32 students from one large introductory statistics subject at a Brisbane university. They must have only minimal prior experience (if any) with statistics, and be volunteers willing to undertake four hours of interviews throughout the semester. Payment may be necessary. If possible, a roughly even number of female and male, younger and older, and high and low achieving students will be selected from the group of volunteers. A statistics course which introduces the topic of hypothesis testing around the middle of the semester will be chosen, so that the longer interviews (discussed below) do not take place in the students' busy time at the end of semester.

Procedure

The study will take place in the first semester of 1995.

Step 1: Interview with the Lecturer

Prior to the commencement of lectures, the researcher will meet with the lecturer involved in the statistics course to discuss: various aspects of the course, such as the role and importance of calculators, computers and text books; the lecturer's approach to the development and use of hypothesis testing in the course; the lecturer's perceptions of the main concepts in elementary hypothesis testing, in terms of what the student is expected to know; and the main difficulties students have with hypothesis testing.

Step 2: Meeting with Statistics Class

At the beginning of the first lecturer of the semester, the researcher will briefly explain the nature of the research project, and call for volunteers to undertake four hours of interviews (each) during the semester. It will be emphasised that this research project is solely the researcher's responsibility, with no involvement from (or feedback to) their lecturer or tutor, and that the interviews are not concerned with assessment for the semester. Names will remain anonymous.

The volunteers will be asked to complete a questionnaire, which requests from the student: biographical information (name, phone number, gender, age group); course-related data (student status, degree of enrolment); achievement information (OP level at Senior, or equivalent); and subject-related data (previous study in statistics, duration of previous study, completion of previous study).

The following steps in the methodology section outline the approaches taken during the three interviews (all recorded on video) with each student.

Step 3 - First Interview with Student

The first interview will occur during the first or second week of the semester. It should take approximately one and one half hours, and involve three activities.

1) Orientation activity: Initial Questions and Discussion

Aims: To gain background knowledge of the student with respect to perceptions about statistics (attitude to the study of statistics, the value of the statistics subject, the role of statistics in everyday life)

2) Introduction Exercise: Concept Mapping

Aims: To introduce the student to the idea of concept mapping; to practice constructing simple maps, such as those from White and Gunstone (1992).

Rationale: This introduction is a response to Novak and Gowin (1984) and White and Gunstone (1992), who concluded that students needed time to learn how to produce a reasonable map on their own. Students will be required to construct a concept map to represent their knowledge of the hypothesis testing concepts, and the relationships between them. Concept maps are considered useful for exploring understanding (White & Gunstone, 1992), and the meaningful relationships between concepts (Novak, 1990).

Method: A brief synopsis of concept mapping will be outlined, noting the main ideas that distinguish them: hierarchical structure, linking words and directional lines. It will be emphasised that: hierarchical structure is not essential if it is not perceived as relevant; directions may be two-way; linking words should improve the meaning of the relationship between the words/concepts; and there is no "right" way to do a concept map. Simple examples, such as White and Gunstone's (1992, p. 17) illustration involving the concepts "living things", "animals", "plants", "dog", "cow" and "grass", will be demonstrated to the student, with the researcher progressively less intrusive.

3) Pretest Problem

Aims: To describe the student's starting knowledge and understanding of prerequisite skills and concepts; to investigate the student's ability to make simple design decisions before instruction

Method: The pretest problem appears as Problem 1 in the Appendix. This activity will take the form of a Talk Aloud interview, with minimal intercession from the researcher. After the student has attempted an answer, additional questions will be asked, such as: If you were to take a another sample of the same size, should your conclusions be similar? Explain.

Step 3 - Second Interview with Student

The second meeting with each student will occur approximately three or four weeks after the introduction of hypothesis testing. It should take approximately two hours to complete the three activities.

1) Concept Mapping Exercise

Aim: To describe the student's knowledge and understanding of the main concepts used in hypothesis testing, and their relationships with each other

Method: This exercise appears as Problem 2 in the Appendix. Students will be asked to "Talk Aloud" while they complete their concept map of hypothesis testing.

2) Hypothesis Testing Exercise

Aim: To study students' attempts at a hypothesis test procedure

Method: This exercise appears as Problem 3 in the Appendix. A single-sample *z* test is expected. Students will be required to "Talk Aloud" while they work through Problem 3. Researcher intercession will be minimal.

3) Follow-up Session

Aim: To further explore research aim (3) with respect to concepts and procedures, and prerequisite concepts such as mean, standard deviation and probability.

Method: Students will be asked to expand on their explanations of the previous two exercises. The extent and direction of the interview will depend on what has been revealed to date. Where possible, they will be asked open-ended questions, for example: Explain the term "alpha"; What would happen if sample size were decreased substantially?

Step 4 - Third Interview with Student

After the last lecture of the semester, a third meeting will be held with each student. It will consist of two activities, which should take approximately one half-hour.

1) Posttest Problem

Aims: To describe the student's knowledge and understanding of the prerequisite skills and concepts required for hypothesis testing at the end of the semester; to investigate the student's ability to make simple design decisions after instruction; and to compare the student's pretest and posttest responses.

Method: The posttest problem appears as Problem 4 in the Appendix. This activity will take the form of a Talk Aloud interview, with minimal intercession from the researcher. After the student has attempted an answer, additional questions will be asked, similar to those asked after the pretest problem.

2) Final Session

Aim: To discuss the main sources of difficulties experienced with hypothesis testing

Method: This section will constitute an informal interview with the student.

ANALYSIS

The analysis of information will be mainly qualitative. The data will be examined for patterns and relationships, as in grounded theory procedures (Strauss & Corbin, 1990), as well as for extentand adequacy of knowledge, as described by White and Gunstone (1992, pp. 91-92). A model of knowledge about hypothesis testing will be developed for each student, with the students' own concept map forming the basis for the model.

REFERENCES

Cassell, D. (1989). What do we mean by the mean? *Teaching Statistics, 11*(2), 38-39.

Fischbein, E., Nello, M. S., Marino, M. S. (1991). Factors affecting probabilistic judgements in children and adolescents. *Educational Studies in Mathematics, 22*(6), 523-549.

Garfield, J., & Ahlgren, A. (1988). Difficulties in learning basic concepts in probability and statistics: Implications for research. *Journal for Research in Mathematics Education, 19*(1), 44-63.

Glencross, M. J. (1986). A practical approach to the central limit theorem. In R. Davidson & J. Swift (Eds.). *Proceedings of the Second International Conference on Teaching Statistics* (pp. 91-95). University of Victoria, BC, Canada.

Hansen, R. S., McCann, J., & Myers, J.L. (1985). Rote versus conceptual emphases in teaching elementary probability. *Journal for Research in Mathematics Education, 16*(5), 364-374.

Hart, A. E. (1984). How should we teach the standard deviation? *Teaching Statistics, 6*(1).

Johnson, D. (1986). Demonstrating the central limit theorem. *Teaching of Psychology, 13*(3), 155-156.

Jones, P., & Lipson, K. (1993). Determining the educational potential of computer based strategies for developing an understanding of sampling distributions. In B. Atweh, C. Kanes, M. Carss, & G. Booker (Eds.). *Contexts in Mathematics Education. Proceedings of the Sixteenth Annual Conference of the Mathematics Education Research Group of Australasia (MERGA)* (pp. 355-361). Brisbane: The Mathematics Education Research Group of Australasia.

Kahneman, D., & Tversky, A. (1982). Subjective probability: A judgment of representativeness. In D. Kahneman, P. Slovic, & A. Tversky (Eds.). *Judgment under certainty: Heuristics and biases* (pp. 32-47). London: Cambridge University Press.

Konold, C., Well, A., Lohmeier, J., & Pollatsek, A. (1993). Understanding the law of large numbers. In J. R. Becker, & B. J. Pence (Eds.). *Proceedings of the Fifteenth Annual Meeting of the North American Chapter of the International Group for the Psychology of Mathematic Education* (Vol. 1, pp. 299-305). San Jose, CA: The Centre for Mathematics and Computer Science Education, San Jose State University.

Konold, C., Pollatsek, A., Well, A., Lohmeier, J., & Lipson, A. (1993). Inconsistencies in students' reasoning about probability. *Journal for Research in Mathematics Education, 24*(5), 392-414.

Mevarech, Z. R. (1983). A deep structure model of students' statistical misconceptions. *Educational Studies in Mathematics, 14*, 415-429.

Loosen, F., Lioen, M., & Lacante, M. (1985). The standard deviation: Some drawbacks of an intuitive approach. *Teaching Statistics, 7*(1).

Mittag, K. C. (1992, April). *Using computers to teach the concepts of the central limit theorem.* Paper presented at the Annual Conference of the American Educational Research Association, San Francisco, CA.

Novak, J. D. (1990). Concept maps and vee diagrams: two metacognitive tools to facilitate meaningful learning. *Instructional Science, 19*, 29-52.

Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn.* Cambridge: Cambridge University Press.

Pedhazur, E. J., & Schmelkin, L. P. (1991). *Measurement, design and analysis: An integrated approach.* Hillsdale, NJ: Lawrence Erlbaum.

Pingel, P. A. (1993). Variability: Does the standard deviation always measure it accurately? *Teaching Statistics, 15*(3), 70-71.

Pollatsek, A., Lima, S., & Well, A. D. (1981). Concept or computation: Students' understanding of the mean. *Educational Studies in Mathematics, 12*, 191-204.

Shaughnessy, J. M. (1983). Misconceptions of probability, systematic and otherwise; Teaching probability and statistics so as to overcome some misconceptions. In D. R. Grey, P. Holmes, V. Barnett, & G. M. Constable (Eds.). *Proceedings of the First International Conference on Teaching Statistics* (Vol. 2, pp. 784-801). University of Sheffield, England: Teaching Statistics Trust.

Strauss, A., & Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques.* London: Sage Publications.

Zerbolio, D. J., Jr. (1989). A "bag of tricks" for teaching about sampling distributions. *Teaching of Psychology, 16*(4), 207-209.

White, R., & Gunstone, R. (1992). *Probing understanding.* London: The Falmer Press.

APPENDIX

Problem 1 (Pretest Problem): Universities have estimated nationwide that their students spend an average of 15 hours per week studying, in addition to their attendance at lectures, tutorials, and so on. How could you investigate the truth of this estimate within your own university?

Problem 2 (Concept Mapping Exercise): On the labels provided are the terms/concepts mentioned in your text book in the chapter on hypothesis testing.

null hypothesis	statistical test
alternative hypothesis	alpha
p-value	decision
confidence level	critical value
calculated value	rejection
Type I error	Type II error

Could you look and reflect on all of them, and use them to try to construct a concept map on the blank page, linking them together wherever you think there is a relationship, and with appropriate words to establish the link. If a term is not entirely familiar to you, either try to fit it onto the map, or leave it aside. Please relax. There is no right answer to this exercise.

Problem 3 (Hypothesis Testing Exercise): Manufacturers claim that a particular brand of light bulb will run for 250 hours, with a standard deviation of 20 hours. To check this claim, a sample of 100 light bulbs was randomly selected from a very large consignment, and the number of hours the light bulbs lasted was recorded. The average number of hours recorded in the sample was 245. Does this sample suggest that the manufacturer's claim may not be correct?

Problem 4 (Posttest Problem): A machine is set to fill empty wine bottles with 750 mls of wine. Explain how you would check that the machine is working correctly.