

Towards a Pedagogy of Inferential Statistics in Graduate Education Programs: Insights from Cognitive and Educational Research

Cengiz Alacaci
Florida International University, USA

Abstract: This study attempts to understand the nature of knowledge base that supports the ability to select statistical techniques for research situations. Findings showed that the largest component of such knowledge was related to research design. One implication is that techniques should be taught in relation to features of research design.

Inferential statistics is an important tool of research. The purpose of the study described herein was to investigate the nature of knowledge base that can support the ability to choose appropriate statistical techniques for applied research situations and to draw implications on statistics education based on the findings. The knowledge that statistical experts (statistical consultants and mathematical statisticians) and novices (graduate students in research methods and mathematics education) use in performing statistical tasks was explored and compared. Subjects compared research scenarios from the perspective of choosing a statistical technique. The knowledge that experts and novices use to perform statistical tasks requiring the selection of appropriate statistical techniques for applied research situations was the focus of the study. Implications of findings for statistics instruction to help novices develop expert-like knowledge bases were sought.

Theoretical Framework

Statistical analysis is an integral part of the research process, and it can be considered as a form of human problem solving. Theories of human problem solving in academic domains postulate that a critical part of the problem solving process is *problem representation*. Problem representation is a personal model of a problem based on domain-related knowledge and its organization (Chi, Feltovich & Glaser, 1981). In inferential statistics, problem representation refers to a person's statistical modeling of a research situation based on his or her knowledge.

One important conjecture is that problem representation relies on the elements of an individual's domain-related knowledge and *their organization* (Chi et al., 1981). Not only the content, but also the inner organization of the elements of an individual's knowledge in a domain is important to represent a problem for successful solution. Studies of knowledge structures in physics and medicine have led to two important conclusions about knowledge structures: (a) elements of knowledge structures can vary in their generality (or size), and (b) knowledge structures can be more efficiently used when they are interconnected (Boshuizen & Schmidt, 1992; Eylon & Reif, 1984).

Knowledge representation

There is a need for a theory to represent knowledge structures in inferential statistics. Schema theory can be used for this purpose. There is also a need for a technique to elicit knowledge structures. Further, the underlying assumptions of the technique should be in congruence with the theory used in representing which is the schema theory. Repertory Grid

Technique of Personal Construct Psychology was used for eliciting knowledge structures of individuals in inferential statistics (Kelly, 1955).

Since knowledge resides in human memory, schema theory--a theory of human memory--can be reasonably used to think about human knowledge structures in a given domain. The basic premise of schema theory is that there exist structures (specified as *schemas*) in memory for situations experienced recurrently. Although schema theory has primarily been developed in the context of understanding prose text, it is consistent with and can embody the three important qualities of knowledge structures listed above--size, organization and goal-directedness (Anderson & Pichert, 1987).

In summary, the ability to choose a statistical technique for a research situation is underlined by the knowledge structure of an individual in inferential statistics. In particular, selection skills of an individual in inferential statistics are supported by well-developed schemas corresponding to statistical techniques. Repertory Grid Technique provides us the tools (additive tree structures) to represent the schema of individuals in a way that can capture the important characteristics of their schemas.

Methods and Data Analysis

Methods

The participants were chosen from two kinds of statistical experts--3 statistical consultants and 3 mathematical statisticians, and two kinds of statistical novices--3 graduate students from mathematics education and 3 from research methods programs. These subjects represented different levels of experience and different specialization in inferential statistics.

Subjects were interviewed by using the procedures of *Repertory Grid Technique* to elicit the knowledge used in tasks based on two kinds of elements: five carefully constructed research scenarios, and names of five statistical techniques. Research scenarios were developed so that they represent a variety of research design situations, corresponding to five techniques: two-way ANOVA, t-test, chi-square test, test of Pearson's moment correlation coefficient, and Sign Test. These research situations were designed to reflect a cross section of theoretical characteristics of the appropriate techniques that reasonably fit them (i.e., parametric and non parametric tests) as well as a variety of the research designs.

Data were collected in two sessions: In the first, subjects were asked to compare and contrast the research scenarios presented in dyads from the perspective of "the kinds of things they consider while choosing an appropriate statistical method of analysis" for them. Six such dyads of research scenarios were used for this purpose. The second session involved rating the research scenarios separately with respect to the aspects of similarity or difference (or the constructs) they had suggested in the previous sessions. Subjects completed the rating by using a Likert type scale.

Data analysis

Data were analyzed to delineate two qualities of knowledge bases: content and organization. The contents of the knowledge base of individuals were inferred from a thematic analysis of the aspects of similarities and differences suggested by the subjects. The themes were classified into four categories: research design, theory, procedures and non technical aspects. The combined number of themes addressed in the research design, theory and procedural categories were taken as a measure of the extensiveness of the content of an individual's knowledge. The number of themes correctly related to all five of the scenarios tasks were taken as a measure of the

connectedness of a person's knowledge, as these themes provided the conceptual connections among the knowledge of individual statistical techniques.

Results and Implications

Results showed that, of the four types of knowledge used for statistical modeling of applied research situations, research design knowledge comprised more than 50 % of the knowledge experts used. Theoretical and procedural elements of the knowledge constituted relatively smaller portions. Although there were no significant differences in the extensiveness of the knowledge used by experts and novices, experts' knowledge use was found to be much more inter-connected.

This study furthered our understanding of expert-novice differences in the field of inferential statistics. More specifically, the results did not support the generalizations about the extensiveness of experts' knowledge use compared to those of novices, but the better organization and connectedness of experts' knowledge use compared to those of novices' was confirmed.

There were two implications of these findings for statistics education: Statistical techniques should be taught in relation to features of research design with which they can be used. Second, statistics education should stress the potential conceptual connections (that is, the similarities and differences) among the techniques from the perspective of associated research design features, as well as the theoretical and procedural aspects. To put these implications into practice, instructors may use a number of strategies: (a) explicating the research design features of the techniques, (b) teaching the techniques in relation to well-designed research scenarios, (c) asking the students to compare techniques from the perspectives of the conditions within which they can be used, and finally (d) asking the students to create flowcharts that can be used for choosing techniques.

From an educational perspective, the results of the study informed us of the kind of knowledge needed for the valued skill of choosing appropriate statistical techniques for applied research situations. The importance of selection skills is highlighted by the increasing availability and use on campuses of powerful statistical software that can perform complicated computations in seconds.

References

- Anderson, R.C., & Pichert, J.W. (1987). Recall of previously unrecalable information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior*, 17, 1-12.
- Boshuizen, H.P.A., & Schmidt, H.G. (1992). On the role of biomedical knowledge in clinical reasoning by experts, intermediates, and novices. *Cognitive Science*, 16, 153-184.
- Chi, M.T.H., Feltovich, P.J. & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121-152.
- Eylon, B. & Reif, F. (1984). Effects of knowledge organization on task performance. *Cognition and Instruction*, 1(1), 5-44.
- Kelly, G.A. (1955). *A theory of personality: The psychology of personal constructs*. New York, NY: Norton.