ASSESSING UNDERSTANDING OF THE CONCEPT OF INTERACTION IN ANALYSIS OF VARIANCE

GREEN, Kathy E.
University of Denver
USA

Students have difficulty understanding the concept of interaction in analysis of variance, implications of the presence of a significant interaction for interpretation of main effects, and appropriate follow-up procedures. While grasping computational procedures and how to get software to generate the appropriate terms, interpretation and communication of results poses a challenge. In this paper exercises for teaching interpretation of interaction are presented. The context was an anova course designed for doctoral students in a college of education. These students are typically interested in substantive research areas such as counseling or educational administration, so the course is applied in nature. The approach taken to teaching interaction was scaffolding or layering of skills. Students first listened to explanations and examples of what interaction is and answered multiple choice questions oriented to comprehension and application outside of class. They then worked in small groups in class to calculate interaction terms both by hand and using statistical software. Next students wrote both technical summaries of results suitable for submission to a journal and one-paragraph "press releases," both of which were critiqued by the course instructor and graduate teaching assistant. These exercises were standard across students and were highly structured. Students then received different data sets and were asked to repeat the analysis, technical, and brief write-up of results for the new data as a homework exercise. The "press releases" were the basis of 5-minute in-class presentations done to encourage verbal communication of results. After the presentation, classmates were given a 3-item quiz about their understanding of the results just presented. The quiz responses were used as feedback to the presenter about the effectiveness of their communication. Successes and failures of this approach are discussed as are thoughts about future directions in teaching about interaction.

INTRODUCTION

Assessment in statistics is changing from a product to a process deeply integrated with teaching and learning that benchmarks student accomplishments as a means of improving student learning (e.g., Garfield, 1994; Mathematical Science Education Board, 1993). The impetus for this change is described by Garfield (1994), Hubbard, (1997), Garfield and Gall (1999), and Garfield and Chance (2000) among others. This change parallels a shift in higher education generally from assessment as a product to assessment as a process driven by clear learning outcomes with a goal of continuous improvement (Higher Learning Commission, 2007). Assessment in introductory statistics has moved from the dominance of computation and multiple choice tests to inclusion of a broader range of methods (e.g., projects, reports, journals, portfolios) that fall under the heading of authentic assessment (see Chance, 1997). Assessment in intermediate-level statistics courses (e.g., correlation and regression, analysis of variance, multivariate analysis), while less likely to have relied on multiple choice tests, also requires alignment with instructional goals in support of student learning.

Statistical literacy is among the goals of statistical education at all levels (Reston, 2005), including intermediate statistics courses. Statistical literacy has been defined in multiple but similar ways. Phillips (2001) stated that in general it refers to "people's ability to interpret, critically evaluate, and communicate statistical information, data-related claims, or chance-related phenomena which they may encounter in diverse life contexts" (Slide 2). Goodall (2005) quoted the definition provided by the Royal Statistical Society which includes the idea of drawing substantively useful conclusions about the functioning of the world from which they [quantitative data] are derived. The Royal Statistical Society report recommended acquisition of statistical literacy via the experience of handling data within the appropriate mathematical frameworks. Shield (2004) stated statistical literacy focuses on the ability to read, interpret, and communicate arguments that use statistics as evidence.

Writing assignments have been used in statistics courses to encourage statistical literacy via communication in formal and natural language. Smith, Miller, and Robertson (1992) assigned microthemes to students in undergraduate business statistics classes, requiring the result expressed “in such a way that a manager with little or no statistical training could understand it” (p. 24). Results of this quasi-experiment suggested attitudes toward statistics were enhanced by the writing assignments and writing was seen by students as helpful. Test performance did not differ between the writing and no writing groups. Beins (1993), however, found writing (in the form of nontechnical ‘press releases’) resulted in an increase in computational and interpretive but not conceptual scores on a final examination. His quasi-experiment was conducted with undergraduate psychology students. Forster, Smith, and Wild (2005) concluded that assignments to write ‘technical notes’ and ‘executive summaries’ with structured examples provided to students resulted in students’ appreciation of the importance of interpretation and communication of findings. Radke-Sharpe (1991) noted the potential benefits of writing across the entire statistics curriculum. Though reports of the effectiveness of writing assignments in graduate level intermediate statistics classes are lacking, it is likely that the effects would mirror those found with undergraduates, particularly for those students lacking extensive background or interest in statistics.

One intermediate statistics class typically required in social science graduate programs is analysis of variance (anova). While a number of anova concepts pose conceptual challenges to students (e.g., error terms in mixed designs, incomplete designs, split-plot or nested designs), this paper is concerned with the concept of interaction.

The concept of interaction in analysis of variance, implications of the presence of a significant interaction for interpretation of main effects, and appropriate follow-up procedures present students with a conceptual challenge. While grasping computational procedures and how to use statistical software, interpretation and communication of results are generally difficult. Rosnow and Rosenthal called interaction effects the “universally most misinterpreted results in psychology” (1989b, p. 1282) and “an enormous source of confusion” (1991, p. 574). Zuckerman, Hodgins, Zuckerman, and Rosenthal (1993) found about a third of 551 researchers in psychology failed to identify the correct way to understand interaction while Umesh, Peterson, McCann-Nelson, and Vaidyanathan (1996) found 75% of articles in marketing and consumer behavior research had errors involving interactions. Schaefer (1976) stated instructors seem to feel interaction is difficult to both teach and learn. Recommendations for teaching the concept of interaction include using multiple “languages” (e.g., spoken word, tables, graphs, arithmetic: Richardson, 1998; Schaefer, 1976), and lectures and handouts (Khanna, 1996). A specific recommendation for assessing understanding of interaction is asking students to generate tables of values showing main and interactive effects (Rosnow & Rosenthal, 1995).

Rosnow and Rosenthal (1989, 1991, 1995, 1996) presented explanations of calculation and interpretation of interaction, with a rejoinder by Meyer (1991). Jaccard (1998) presented both a difference between mean differences and a residualized approach to expressing interaction. The view taken in this paper (and in the anova class) follows that of Keppel and Wickens (2004), who stated that in the presence of a significant interaction, conclusions about factor means must be tempered by consideration of level of the other factor(s), graphs are almost essential, simple effects are usually looked at, simple contrasts might be calculated, and finally interaction comparisons might be used. Keppel and Wickens noted that simple effects confound the main effects and interaction sources of variation.

The understanding of interpretation of interaction was assessed with an emphasis on writing assignments. The context was a 10-week, five quarter-hour anova course offered to doctoral students in education, psychology, biology, human communication, and social work at a mid-sized university in the United States. Classes met for 2 ½ hours twice a week. The enrolled students are typically interested in substantive research areas such as counseling or educational psychology, and the course is applied in nature as opposed to theoretically oriented. The approach taken to teaching interaction was scaffolding or layering of skills, with conceptual instruction prior to procedural instruction and supplemented by extensive use of tables and diagrams. Hong and O'Neil (1992) found this sequence effective in instruction in inferential statistics.

ACTIVITIES RELATED TO INSTRUCTION AND ASSESSMENT

The concept, data analysis, and interpretation of results with respect to statistically significant interactions were introduced via lecture in a 2 ½-hour class with powerpoints available for students prior to class. During this presentation, students viewed tables of means and graphs reflecting presence and absence of significant interactions. During the next 2 ½-hour class, students worked in small groups to calculate main and interactive effects by hand for a small data set and to generate results from a real data set using SPSS software. They verbally described their results in class. ("It looks like the individual treatment worked better for people living at home than in assisted living but people in the group treatment didn't do so well.") After this week of instruction, students were required to take a 10-item, multiple choice, graded quiz on-line, outside of class time. Items assessed comprehension and application of the concept of interaction. Students then selected one of five created data sets of some interest to them and used that data set to complete five graded homework assignments. Two of the homework assignments involved analysis of data where a significant interaction was present, one for a between-subjects factorial design and a second for a mixed (repeated measures) design. These two (and two other) homework assignments required students to produce a formal and informal write-up of results. The concept of interaction was reinforced via review at least six times subsequent to the initial presentation. Finally, students completed a final exam with two problems (12/30 points) yielding interactions that had to be correctly interpreted to receive full credit.

Writing Exercises

Students were asked to report results in a *formal* write-up of results supplemented by an analysis of variance summary table and figures as necessary. This was to be no longer than 1-2 paragraphs plus appropriate tables and figures. Examples were placed on-line and students were referred to a refereed journal of their choice to find exemplars. Students were also asked to report results in an "executive summary" that was *informal* in nature and 3-4 sentences in length maximum. (Two examples of reporting were again placed on-line for each homework exercise.) Executive summary results were given verbally in class in 2-minute presentations (no handouts allowed). After the verbal presentation, classmates were given a 3-item quiz about their understanding of the results just presented. The quiz responses were used only as feedback to the student-presenter about the effectiveness of their communication.

CLASS PERFORMANCE

Students did well on the multiple-choice items, with a typical score of 9/10 correct. Students generally did well on narrowly defined homework and test questions (e.g., identify concerns in interpretation of an unbalanced design, define the error term for sources of variance in a nested design with A fixed and B and C:B random). Students on the whole clearly understood the implications of an interaction for results and follow-up analyses. However, after two homeworks with formal and informal write-ups of results followed by verbal presentations and class feedback, students were clearly *not* able to simply and effectively describe what they found. Students' formal write-ups improved substantially from the first to the last of five homeworks and were generally good on the final examination. However, their executive summaries continued to be muddled and their verbal presentations confusing. Student comments were on the order of "I know what I found but I don't know if I said it right." Students did *not* report finding their classmates' feedback on verbal presentations useful.

DISCUSSION

Students generally performed better on more structured work. The tension arises in providing sufficient structure but not so much that the problem becomes a completion exercise rather than an exercise in interpretation and communication. With three iterations of writing formal results summaries involving interactions along with examples of appropriate wording, students approached a reasonable level of proficiency. These formal summaries are what students read in journals in their fields and what they understand will be needed in dissertation defenses. Students were likely motivated to learn this form of communication.

The executive summaries were less prescriptive and did not improve noticeably over the course of the quarter, even with instructor and classmate feedback. Clearly, the examples and wording provided for the executive summaries would need to be adjusted to become less flexible. Also, students' verbal presentations lacked focus and were barely understood by many of their sympathetic classmates. While verbal executive summary presentations were modeled by the instructor, this was insufficient. The executive summaries were required but received less credit than the data analysis and formal write-ups. The verbal presentations were not graded. The failure of this portion of the communication exercise to meet the learning goal may be due to any or all of the following: lack of student motivation since point values were low or zero, confusion about the task, misunderstanding of the interaction concepts, lack of practice in talking in anova language, and difficulties with public speaking in English generally. (This was an especially difficult task for the non-native English speakers.)

Changes that will be made in the next iteration of the course will be to drop the requirement for an executive summary. The assignment will instead be presented as a mini conference talk. Students may be more motivated to practice presenting results of interaction analyses if there is a clearer link to their professional practice. More extended time for the verbal presentation will be scheduled though with less frequency. Students will be required to do a mini-presentation only once during the course, but it will be graded. Notes and materials for their presentation will also be critiqued prior to presentation. More requests are also planned for students to talk in anova language in class as opportunities arise.

REFERENCES

Beins, B. C. (1993). Writing assignments in statistics classes encourage students to learn interpretation. *Teaching of Psychology, 20*, 161-164.

Chance, B. L. (1997). Experiences with authentic assessment techniques in an introductory statistics course. *Journal of Statistics Education, 5*. Retrieved online from www.amstat.org/publications/jse/.

Dunn, D. S., Smith, R. A., & Beins, B. (Eds.). (2007). *Best practices in teaching statistics and research methods in the behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.

Forster, M., Smith, D. P., & Wild, C. J. (2005, April). Teaching students to write about statistics. Paper presented at the IASE/ISI Satellite Conference, Sydney, Australia.

Garfield, J. B. (1994). Beyond testing and grading: Using assessment to improve student learning. *Journal of Statistics Education, 2*. Retrieved online from www.amstat.org/publications/jse/.

Garfield, J. B., & Chance, B. L. (2000). Assessment in statistics education: Issues and challenges. *Mathematical Thinking and Learning, 2*, 99-125.

Garfield, J. B., & Gal, I. (1999). Assessment and statistics education: Current challenges and directions. *International Statistical Review, 67*, 1-12.

Goodall, G. (2005). News and notes. *Teaching Statistics, 27*, 96.

Higher Learning Commission (2007). *Criteria for accreditation*. Downloaded from www.ncahlc.org, March 23.

Hong, E., & O'Neil, H. F., Jr. (1992). Instructional strategies to help learners build relevant mental models in inferential statistics. *Journal of Educational Psychology, 84*, 150-159.

Hubbard, R. (1997). Assessment and the process of learning statistics. *Journal of Statistics Education, 5*. Retrieved online from www.amstat.org/publications/jse/.

Jaccard, J. (1998). *Interaction effects in factorial analysis of variance*. Thousand Oaks, CA: Sage.

Khanna, R. (1996). *An analysis of the teaching, understanding and interpretation of interaction effects in a factorial design*. Dissertation Abstracts International Section A, 56(7-A), 2653.

Mathematical Science Education Board. (1993). *Measuring what counts: A conceptual guide for mathematical assessment*. Washington, D.C.: National Academy Press.

Meyer, D. L. (1991). Misinterpretation of interaction effects: A reply to Rosnow and Rosenthal. *Psychological Bulletin, 110*, 571-573.

Phillips, B. (2001, August). Statistics literacy. Presentation at the IASE Satellite Conference on Statistics Literacy, Seoul, Korea.

Radke-Sharpe, N. (1991). Writing as a component of statistics education. *The American Statistician, 45*(4), 292-293.

Reston, E. D. (2005, April). *Assessing statistical literacy in graduate level statistics education*. Paper presented at the 55th session of the International Statistical Institute, Sydney, Australia.

Richardson, W. K. (1998). Teaching analysis of interaction in the 2x2 factorial design. *Teaching of Psychology, 25*, 297-299.

Rosnow, R. I., & Rosenthal, R. (1989a). Definition and interpretation of interaction effects. *Psychological Bulletin, 105*, 143-146.

Rosnow, R. I., & Rosenthal, R. (1989b). Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, *44*, 1276-1284.

Rosnow, R. L., & Rosenthal, R. (1991). If you're looking at the cell means, you're not looking at *only* the interaction (unless all main effects are zero). *Psychological Bulletin, 110*, 574-576.

Rosnow, R. L., & Rosenthal, R. (1995). "Some things you learn aren't so": Cohen's paradox, Asch's paradigm, and the interpretation of interaction. *Psychological Science, 6*, 3-9.

Rosnow, R. L., & Rosenthal, R. (1996). Contrasts and interactions redux: Five easy pieces. *Psychological Science, 7*, 253-257.

Schaefer, V. H. (1976). Teaching the concept of interaction and sensitizing students to its implications. *Teaching of Psychology, 3*, 103-114.

Schield, M. (2004). Statistical literacy and liberal education at Augsburg College. *Peer Review, 6*, 16-18.

Umesh, U. N., Peterson, R. A., McCann-Nelson, M., & Vaidyanathan, R. (1996). Type IV error in marketing research: The investigation of ANOVA interactions. *Journal of the Academy of Marketing Science, 24*, 17-26.

Smith, C. H., Miller, D. M., & Robertson, A. M. (1992). Using writing assignments in teaching statistics: An empirical study. *Mathematics and Computer Education*, *26*, 21-34.

Zuckerman, M., Hodgins, H. S., Zuckerman, A., & Rosenthal, R. (1993). Contemporary issues in the analysis of data: A survey of 551 psychologists. *Psychological Science, 4*, 49-52.