Developing a Second Course in Statistics for Students in the Social Sciences (Poster)

Prof. Sharon Lane-Getaz
St. Olaf College, 1520 St. Olaf Ave., Northfield, MN 55057

Abstract:
The question is often asked among statistics educators, "What should we teach after the introductory course?" The objective for this paper is to describe the development, implementation and evaluation of a second course in statistics designed to meet the needs of students in social science disciplines. This "Intermediate Statistics for Social Research" course was developed collaboratively with input from professors in social science disciplines and a student double majoring in Mathematics and Psychology. The course topics are unique at an undergraduate level addressing measurement, design, and analysis, including: ANOVA designs, Classical Test Theory, Principal Components Analysis (PCA), and Exploratory Factory Analysis (EFA). Students explored three essential questions: "How do we reliably measure something?," "How do we design valid research?," and "How do we analyze research?".

Key Words: statistics education, social science, second course in statistics, principal components analysis, exploratory factor analysis

1. Introduction

“What should we teach after the introductory course?” The objective for this project was to develop, implement and evaluate a second course in statistics designed to meet the needs of students in social science disciplines who may have limited mathematical backgrounds but want to pursue a concentration in applied statistics. Intermediate Statistics for Social Research (ISSR) was developed collaboratively with social science professors’ and students’ input; most notably from Psychology professors, a summer researcher double-majoring in Mathematics and Psychology, and five students, further described below, taking the experimental course. “How do social scientists reliably measure something; design valid research, and analyze results?” ISSR case studies go beyond one- and two-way ANOVAs. Students learn techniques that allow them to explore relationships among correlated variables. They are introduced to Classical Test Theory and data reduction techniques: Principal Components Analysis and Exploratory Factor Analysis. Students also research and present a topic to the class.

2. Methods and Materials

2.1 Enrollment

Five students enrolled for the January term (2 males, 3 females) –all with different majors; four were statistics concentrators. One senior (political science), three juniors (economics, psychology, and exercise science), and one undeclared sophomore enrolled. All the students had taken at least one introductory course (3 AP Statistics, 2 Algebrabased intro courses, and 2 took an intro course with a Calculus pre-requisite); four had also completed a second course in statistics which emphasizes regression (statistical modeling), and one student had also taken advanced statistical modeling. There was a wide range of ability among the five students.

2.2 Course Description

The first three weeks of the course consisted of instructor led lectures and lab activities:

I. ANOVA design and interpretation
- One-way ANOVA, decomposition of Sums of Squares
- Two-way ANOVA: Main effects and interaction
- Tukey's Honestly Significant Difference (HSD)
- Fisher's Least Significant Difference (LSD)
- Scope of inference

II. Basic Measurement Statistics / Psychometrics:
- Classical Test Theory
- Bias, discrimination, and difficulty
- Reliability and validity

III. Advanced Measurement Topics
- Eigenvalues and Eigenvectors of the Covariance Matrix
- Principal Components Analysis
- Exploratory Factor Analysis

The fourth week of the course consisted of elected topics developed and presented by students. The topic of the student presentation was to deepen their classmates' understanding of a topic touched upon in the first three weeks of the course. The grading rubric for the student presentations was developed collaboratively by the students in the course. Topics selected included: planned comparisons and contrast coding, two different presentations using the Analysis of Covariance (ANCOVA), and a lesson and lab utilizing the Randomization F-test simulation.

2.3 Lesson Objectives by Topic

I. Introduction to Design & Analysis using ANOVA (Days 1-4)
   Students will be able to: (1) Plan for sufficient power, effect size, and sample size for ANOVA designs, (2) Distinguish fixed versus random effects models (3) Apply one- or two-way ANOVA designs with and without interaction, (4) Interpret ANOVA and ANCOVA designs, (5) Plan and present an intermediate level experimental design topic.

II. Introduction to Measurement (Days 5-7)
   Differentiate true score as described by Classical Test Theory from observed: $X = T + e$. Differentiate nominal, ordinal, interval and ratio scales.
   1. Reliability: Students will be able to: (1) Identify key assumptions of Classical Test Theory, (2) Predict scale length, (3) Describe the concept of reliability in intuitive and statistical terms, (4) Estimate reliability (e.g., Cronbach's Coefficient alpha) and compare with common reliability measures, and (5) Compute and discuss concepts of inter-rater reliability.
   2. Validity: Students will be able to: (1) Discuss types of validity evidence, (2) Describe conceptual and quantitative relationship of reliability to validity, and (3) Differentiate a true correlation from an observed correlation.
   3. Bias, discrimination, and difficulty: Students will be able to: (1) Interpret item analysis output, including measures of bias, discrimination, and difficulty for a measured variable, and (2) articulate how these measures may impact a research study

III. Dimension Reduction and Latent Variable Analysis (Days 8-14)
Differentiate issues and features using Principal Components Analysis & Exploratory Factor Analysis.
1. Principal Components Analysis (PCA)
Students will be able to: (1) Identify when PCA is appropriate and when not, (2) run a PCA analysis using R (or SPSS), (3) Use rotation to reduce dimension(s) and graphically display and interpret data relationships using scree plots, biplots, score plots, and (5) Describe processes underlying PCA.
2. Exploratory Factor Analysis (EFA)
Students will be able to: (1) Identify when EFA is appropriate and when not, (2) Run multi-factor analyses to extract underlying factors, (3) Differentiate why an orthogonal or oblique rotations is used, (4) Interpret output including regression score estimates and score plots, and (5) Describe processes underlying Exploratory Factor Analysis.

IV. Student Presentations (Days 15-20)

2.4 Sample Case Study

A brief description of the final case study, objectives, and lab is described. During the third week of the course the class explores the Moral Schema Scale (MSS) case study. Previous research on developing a Moral Schemas Scale (MSS) generated 73 items and a 3-dimensional scale, with sacredness, care, and justice as sub-scales (Hanson & Huff, 2001). In this extension, a fourth sub-scale (self-interest) was theorized. The 41 MSS items in this study include eight each for sacredness, care, and justice and 17 for self-interest. The MSS lab objectives, sample student output and analysis, and major teach points from the class lab and discussion follow:

MSS Research questions include:

- Does evidence support that “self-interest” is an MSS sub-scale?
- To what degree do assigned roles (clergy, business consultant, counselor or judge) correspond with high ratings that the character uses the schema (sacredness, self-interest, care or justice, respectively) to make a moral decision?

MSS Lesson Objectives

- Utilize Key Vocabulary: Eigenvalues & Eigenvectors, Covariance & Correlation Matrices, Orthogonal, Scree plots, Factor Loadings, Score plots, Biplots
- Articulate and differentiate conceptual foundations of PCA and EFA
- Interpret PCA and EFA results
- Apply ANOVA techniques used in previous lessons

MSS Lesson Outline

- Discuss Moral Schema Scale (MSS) case study
- Produce PCA and EFA output; examine evidence for MSS sub-scales
- Use boxplots and ANOVA to examine MSS sub-scale ratings by Role

2.4.1 Principal Component Extraction.

Seven components were extracted that explained 72% of MSS score variation. Only four components account for $\geq$ 10% variation (Table 1). Furthermore, given the large drop in variation explained by the fifth component (Figure 1) and the MSS theory being investigated, four components were retained with 58% of the score variation explained in total.

Table 1. Standard Deviation and Proportion of Variance Explained by each Principal Component (PC)

	PC1	*PC2*	*PC3*	*PC4*	*PC5*	*PC6*	*PC7*
Standard deviation	2.1	1.7	1.6	1.5	1.1	.9	.9
Proportion of Variance	.21	.14	.12	.11	.06	.04	.04
Cumulative Proportion	.21	.35	.47	.58	.64	.68	.72

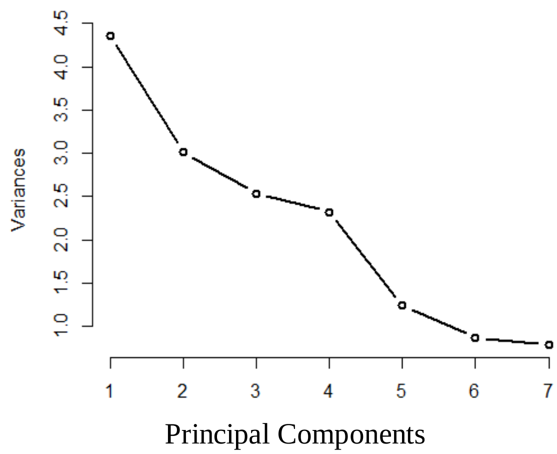

Figure 1. Scree plot for seven Principal Components

2.4.2 Principal Components Analysis

Principal Components Analysis characterizes all the sample score variation in a new coordinate space. MSS scores are well-dispersed across PC-1, “self-interest;” however, the sample noise inherent to PCA seems to conflate “justice” and “care.” Both justice and care appear to load ambiguously on both biplot y-axes in Figure 2.

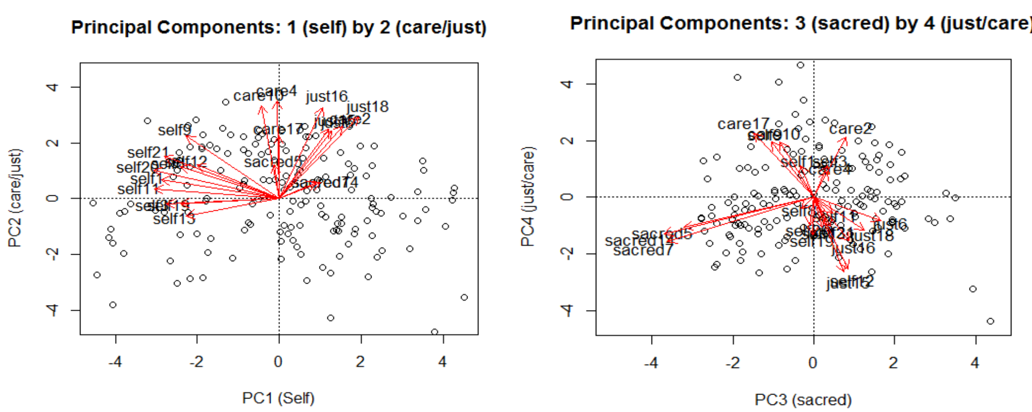

Figure 2. Principal component biplots for PC-1 by PC-2 and PC-3 by PC-4.

2.4.3 Exploratory Factor Analysis

Exploratory Factor Analysis extracts only common variation in MSS scores to eliminate variation unique to the sample. Factor loadings show “self-interest” items loading on Factor-1, “justice” loading on Factor-2, and both “sacredness” and “care” clustered at the near origin (Figure 3). Similarly, “sacredness” is characterized by Factor-3, “care” by Factor-4 with the other schemas clustered at (0, 0). These orthogonal loadings support the theory that "self-interest" is a unique moral schema. As one student summarized, "The groupings … provide support for the moral schema theory."

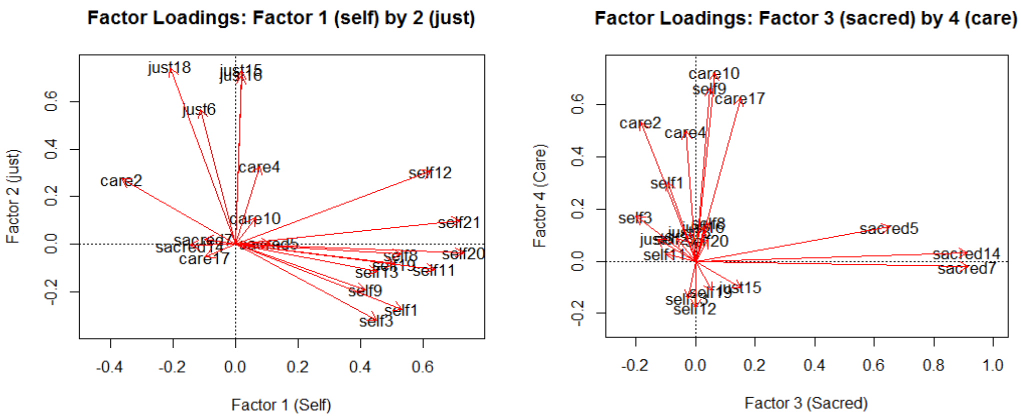

Figure 3. Loading plots for Factor-1 (self) by -2 (just) and Factor-3 (sacred) by -4 (care).

2.4.4 ANOVA

ANOVAs were conducted assuming roles were randomly assigned. Comparative boxplot (Figure 4) and Tukey's Honestly Significant Differences support the MSS theory. Those in a clergy role believed a sacredness schema would be employed (n_1=41, p < .001); the business consultant role had high ratings on a self-interest schema (n_2=43, p < .001); the counselor role rated a care schema highest (n_3=42, p < .001); and those in the judge role, the justice schema (n_4=42, p < .01). Reliability a = .68, n = 21 items, N = 168

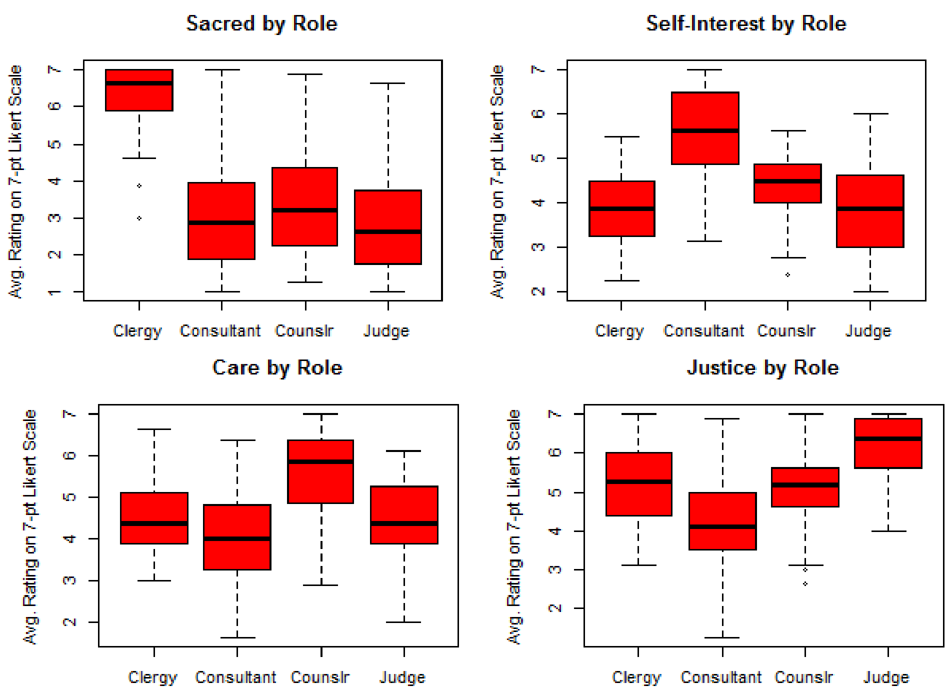

Figure 4. Comparative boxplots of the average Likert scale rating for the four item schemas (sacredness, self-interest, care and justice) by assigned role.

3. Evaluation

3.1 Students' Favorite Aspects of the Course

- "Learning PCA and EFA. I had never done anything like this before, and I was really excited to learn it because it is so applicable in psychology research! I think we learned it in a way that made it comprehensible at the undergraduate level."
- "EFA and all of the in class activities."
- "I like EFA, ANOVA and the randomization F-test and contrast coding best"
- "The ANOVA/ANCOVA and the randomization tests."
- "ANOVA is probably the most useful …. PCA and EFA are a little tougher to grasp, but are a really interesting way of looking at relationships."

3.2 Students' least Favorite Aspects of the Course

- "The lessons got a bit long especially the eigenvectors and eigenvalue lesson."
- "The lectures … a better way would be have the quiz and HW first and then the lecture or …questions, but trouble shooting is the best way to learn stats and *R*."
- "The context of each problem wasn't always very interesting. But often keeping the context simple … allowed … focus more on the functionality of the concepts, so there was a trade-off."
- "I don't have any least favorite aspect of the course, but I would recommend to add randomized F test and contrast coding…."
- "We never really discussed the reading from the DeVellis book."

4. Future Directions

Develop an additional social science-related case study. New case study will evaluate effects of repeated *Alternatives to Violence Project (AVP)* anger management trainings on inmates' anger expression. AVP surveyed participants before and after initial training and after any subsequent sessions using the reliable and valid STAXI-2 scale (Dalton, et al., 1998). Comparative data were also gathered from nonparticipants. Data analyses may include repeated-measures ANOVA on STAXI-2 scores, principal components analysis and factor analysis of STAXI-2 state and trait anger subscales. (Classical Test Theory may be reduced by 1-2 days in order to add this content.)

Include "Randomization F-test Simulation" activity. One of the student presentations included a Randomization F-test simulation. The learning module was well-received and will make a good addition to the course with some additional development. (To add this, the first day "Gummy Bear Launch Activity" can be introduced and set up during class time for completion as homework in teams.)

Scaffold "Planned Comparisons and Contrasts" project. One student researched and presented planned contrasts. The learning module generated as many questions as were answered. Per student requests this topic may need more instructor guidance, even if it remains a student-led activity.

5. Conclusion

Intermediate Statistics for Social Research will be offered biannually as a statistics concentration seminar. Enrollment will be limited to 15 students to scaffold, as needed, and to facilitate discussion and cross-disciplinary group activities. The ISSR course was "at the right level." Four students thought the course would be accessible to students after completing an introductory course. Two students, with limited or no prior use of R statistical software, successfully used the menu-driven "R Commander" as a sheltered starting point. All students stated that they would recommend ISSR to others and envision using the skills they obtained in this course for future research.

Acknowledgements

Chuck Huff (St. Olaf Psychology Professor) for his Moral Schema Scales research. Terry Kayser (Alternatives to Violence Project) for the anger management research. Luke Obermann (St. Olaf '12) for collaborating on PCA and EFA labs. St. Olaf Center for Interdisciplinary Research for summer research funding from the National Science Foundation Grant #0354308.

Selected References

Cannon, A. R., Cobb, G, Hartlaub, B., Legler, J., Lock, R., Moore, T., Rossman, A., Witmer, J. (2011). *Stat 2: Building models for a world of data.* Freeman. (Textbook)

Dalton, J. E., Blain, G. H., & Bezier, B. (1998). State-trait anger expression inventory scores of male sexual offenders. *International Journal of Offender Therapy and Comparative Criminology, 42*(2), 140-147.

DeVellis, R. F. (2012). *Scale development: Theory and applications.* Thousand Oaks: Sage Publications. (Textbook)

Hanson & Huff, C. (2001). The moral schemas scale: Care, justice, sacredness, and self-interest. For the 14th Annual Meeting of the *American Psychological Society*, June 2002, New Orleans.