What: Sabbatical Proposal
Who: Dr. Laxman Hegde, Professor of Mathematics
Title: A Revisit to Canonical Correspondence Analysis (CCA)
Term Requested: Spring 2012
Purposes: Scholarly or creative endeavor; Improvement of skills in the discipline being taught
Category: A special project or research problem planned with specific objectives

Abstract

The purpose of this proposal is to request a sabbatical leave for the spring 2012 semester to enhance my academic research and teaching skills. A main driving force behind this proposal is my research paper (jointly with Dr. Naik of ODU) published in 1999, entitled "Canonical Correspondence Analysis in SAS Software." This paper has received world-wide attention of the researchers in ecology and it needs update to enhance its value to the future users. There are two goals to the proposed work. First is to undertake research activities in the field of statistical inference (estimation and testing of hypotheses) related to CCA and adding descriptive interpretations to mathematical quantities generated by our SAS program, and second to organize a workshop (jointly with Biology Department of FSU and Appalachian Laboratory) for the researchers who use CCA. Execution of the proposed sabbatical activities will enhance my research skills and statistical consulting and computational abilities. Importantly, the proposed workshop during my sabbatical leave will be of immense help to future researchers who apply CCA and our SAS program in their work. Some outcome to these activities: a refined research paper in CCA for a possible publication, better teaching, and **overall prestige and national reputation** for Frostburg State University.

Introduction and Background

If we do a simple Google Search with the key words "canonical correspondence analysis", the Google will return the following site as a result about my research paper.

Jointly with Dr. Dayanand Naik, Professor of Statistics at Old Dominion University (ODU), VA, I published this paper entitled "Canonical Correspondence Analysis in SAS Software" in SUGI 24 in 1999. A main concern of this paper was to bring out a SAS program to do Canonical Correspondence Analysis (CCA). Please see the abstract of our paper (Attachment A) or click the link given in Google Search to see the complete paper. We wrote this program mainly for the following reasons.

- CCA is a very popular data analysis tool among ecologists and Canoco was the only program to do the statistical analysis.
- CANOCO is a very expensive program, and not widely known among the researchers outside the field of ecology.
- SAS (Statistical Analysis System) is a very popular, widely used, powerful, and readily available statistical software in most research schools and institutions.

Consistent with the reasons stated above, ever since our paper was published in 1999, we have received numerous requests (more than 50) for our SAS program from researchers all over the world. See a list of researchers from whom we have received requests (Attachment C). The following two sample emails I received recently would give me supporting evidence for the proposal. We just received a request from South Korea.

The FSU Profile (winter 2004) published a brief write-up about the success story of our work (Attachment B). These increasing requests for our research became a motivational factor for me to seek the sabbatical leave in the spring of 2012 to conduct more advanced research on the topics of CCA and enhance the utility of our SAS program.

Objectives

- To acquire research skills and knowledge to more successfully address CCA
- To provide more insight/interpretations to statistical output from SAS program
- To conduct a workshop for researchers using CCA and SAS

Activities

1) **Research**

   In our 1999 SUGI paper, we mainly translated the mathematics of CCA originally developed by Ter Braak (1986) to SAS codes. Dr. Naik of ODU and I are of the strong opinion that the theory of CCA can further be advanced by looking into some important statistical inferential issues (estimation and hypothesis testing). Currently our SAS program mainly provides some descriptive statistics (Eigen values, Correlations, and Biplot) and it does not perform any inferential procedures like constructing confidence intervals or computing p-values. During my sabbatical leave, I would like to investigate these topics and enhance the capabilities of our SAS program. Dr. Naik is a versatile researcher in the areas of Multivariate Statistics. I also intend to spend a part of my sabbatical period with Dr. Naik of ODU to improve my research skills. *This is one reason I will not be able to perform this activity during a regular semester given my full teaching load of 4 courses, and perhaps an overload.*

2) **SAS Program**

Similar to the output from CANACO (1987), our SAS program also outputs several pages of statistical quantities and a Biplot that are generally difficult to interpret and feel the insight for users of CCA who are not well conversant with Linear Algebra concepts such as eigenvalues and eigenvectors. In its existing format, our SAS program simply provides numerical output and a Biplot. During my proposed sabbatical period, I intend to add interpretations of important mathematical/statistical quantities to our SAS output wherever it is applicable and meaningful. This will greatly enhance utility of our program to the end users. Note that this feature is not currently available in CANOCO output also.

3) **Workshop**

In the last 15 years of my tenure at FSU, I have provided statistical consultancy and computational services in SAS to several of our graduate students in biology and ecology as part of their thesis work. In fact, quite a few of these projects dealt with CCA. It is important to note that the FSU Biology Department is closely associated with the Appalachian Laboratory (AL) which is an environmental research facility of the University of Maryland Center for Environmental Science (UMCES), the principal institution for advanced environmental research and graduate studies within the University System of Maryland. The CCA type applications are quite common in research activities of AL. During the last week of my sabbatical period, jointly with AL and FSU Biology Department, I intend to organize a workshop for researchers who are interested in applications of CCA and our SAS program. I will work out the logistics (how to organize this workshop in consultation with AL) if my sabbatical proposal is approved. Given that my sabbatical is approved, I also intend to write a proposal for Workshop/Speaker Grants from the Faculty Development and Sabbatical Subcommittee to bring distinguished speakers for the proposed workshop.

Benefits and Outcomes

I will gain increased competence in providing statistical consultancy and computational services to graduate students in Biology Department and AL. In fact, I am teaching Math 680 (Research Statistics) this semester with 13 graduate students, all from Biology Department. It is quite likely that some of these students will be analyzing their data using CCA and SAS program under my guidance in the coming years.

Future users of our SAS program will appreciate it more, and in turn, FSU will earn more reputation.

Increased creative ability to do research.

Possible seminar course for graduate students.

Possible research publication.

Workshop should add to the overall prestige and national reputation of Frostburg State University.

I thank you in advance for your consideration of this sabbatical proposal. Naturally, I would be pleased to answer any questions you may have concerning any aspect of the proposal.

Attachment A: Abstract of our 1999 SUGI paper

Ecologists analyze species-environment relations from data on biological communities and their environment. Generally, the data of occurrence or abundance of each species of a taxonomic group is collected at several sites. Also data on a set of environment variables that are important in explaining the variation in species are collected. A "site" is a basic sampling unit separated in space or time from other sites, e.g. a quadrant, a woodlot or a light trap. Canonical Correspondence Analysis (CCA) is a multivariate technique to relate composition of a species when species have bell-shaped response curve with respect to environmental gradients. Note that statisticians interpret CCA as canonical correlation analysis in standard multivariate statistical analysis. Although canonical correspondence analysis and canonical correlation analysis are closely related, there are some subtle differences between the two techniques. Ter Braak (1986) has shown a complete derivation and applications of CCA techniques. Also, Ter Braak (1988) has developed a computer program CANOCO to perform CCA and several other multivariate statistical techniques to analyze species-environmental relations. CANOCO is perhaps a widely used program by ecologists and biologists, but not so well known for the researchers in other fields. In this paper, we review mathematical contents CCA as discussed in Ter Braak (1986) and develop a SAS program to perform CCA and biplots to present a species environmental relationship.

Attachment B: Article in FSU Profile (Winter 2004)

Mathematical Finding Brings International Attention

Laxman Hegde, professor of mathematics, has recently seen a jump in interest in a paper he and co-author, Dayanand Naik of Old Dominion University, published in 1999 in the Proceedings of the SAS Users Group International (SUGI) Conference. The paper discusses a different type of data analysis used by ecological scientists who study species-environmental relationships, and readers are offered electronic copies of the SAS (statistical analysis system) program code.

Word has finally gotten out, and, in the past several months, Hegde and his co-author have suddenly received dozens of requests for the paper and the program code from environmental scientists worldwide. It's particularly popular because the previous method of analyzing large amounts of environmental data was through the use of a commercial – and expensive – computer program.

Hegde and Naik are in the process of submitting the paper to "Ecology," a major journal.

The program has already assisted a number of FSU biology graduate students with their projects. It's helpful with interpreting results because it provides a geometric interpretation of their data.

— LDM