Program Ideas for the Workshop on Statistical Computing in Academia

March 5, 2001

The topic of furthering research in statistical computing in academia is complex and involves many inter-related topics. However, we have divided them into 4 general areas for discussion in the workshop. These are

- Evaluating and rewarding research in statistical computing;
- Promoting inter-disciplinary collaboration and exchange of research;
- Teaching statistical computing;
- Increasing the number of students involved in statistical computing.

1 Workshop Schedule

The current proposal is to divide the day into 5 sessions. In each of the first 4 sessions, we will discuss one of the above topics. The final session will be a discussion about what concrete and feasible actions can be undertaken in response to the earlier discussions.

In each of the 4 topic sessions, the workshop participants will discuss the perception, promotion, use and evaluation of statistical computing and graphics. These are less technical and specific concepts than most workshops aim to discuss, and thus we plan a series of dialogues instead of a succession of technically focused presentations about people's work in statistical computing.

Ideally, 1 or 2 people will lead the discussion in each session, starting by presenting an overview of the issues and problems, and emphasizing their experiences (both personal and observed) which suggest an approach to the problems. These presentations would be approximately 10 minutes long and establish the content for the discussion. We hope that these speakers can coordinate with each other and the program committee in the months before the workshop, and prepare a short document that describes the issues. Then, in the short presentations, they can provide more focus, background and suggestions.

The content of the final section will emerge from the earlier discussions. We will discuss practical steps and actions that we can take to promote statistical computing and support those doing research in the area. This will probably include discussion about how the ASA and other organizations can help and also how we can use provide or improve services such as on-line journals, educational packages, software announcements and reviews, hosting services, licensing and copyright information, etc.

There are two things that we need to do soon. These are to identify topics for discussion and to arrange for people to lead these discussions. We suggest some topics in the next section, but strongly encourage people to mail topics that *they* would like to hear discussed.

We are looking for volunteers and for others to suggest people to invite for this role. We will want these individuals to prepare a page outlining their thoughts on the particular topic they choose to moderate. This paragraph will be made available to the participants a few weeks before the workshop. We are not looking for people with any particular expertise in these areas, but people who can identify and present the issues.

2 Outline of the Topics

Some of us feel that statistical computing – that is visualization/graphics, environments and languages, and algorithms – are important areas both for our discipline and others, but that they do not receive sufficient attention or focus. While many benefit, there are few people actively working in statistical computing as a research area. This workshop aims at helping to understand what practical steps we can take to promote and sustain these areas of research.

We feel statistical computing is important for a variety of reasons that include the following.

- It is probably the most common medium by which people come in contact with statistical ideas and methodology (e.g. Excel). This trend can increase with suitable technology for publishing interactive and dynamic research papers.
- Many different disciplines use statistical computing concepts. It is efficient to centralize this specialized expertise so that other disciplines can profit from it rather than recreating it.
- Many problems are not amenable to closed form solutions and require extensive, flexible computing approaches and tools. Similarly, handling increasing volumes of data in an efficient manner (in human terms) is a significant challenge facing the field of statistics.

In the following paragraphs, we touch on some of the potential issues that we might discuss at the workshop in each of the four areas. These are by no means intended to be definitive or exhaustive, but rather to set the scene and tone; feel free to suggest others.

Inter-disciplinary Collaboration Statistics has a lot to offer other disciplines, from ideas and methodology to software. Yet we find that other disciplines are re-inventing ideas and software long available to statisticians. We should be attempting to understand why the work we have done is not being used and how we can adapt to the needs of these other disciplines. Fields such as bioinformatics, astronomy and geography are ripe for interaction with statistical computing. How can we facilitate and encourage such interactions and include researchers in other areas of statistics?

Also, there are many new and very interesting areas in which statistics, and in particular statistical computing, can embrace and exploit ideas from other disciplines. While computer science is an obvious example, what are others?

Promoting & Rewarding Research Academics pursuing research in statistical computing have frequently received less recognition for their work than others working in more classical areas. This is partially because statistical computing is less mainstream and most academics have little background in the area. Hence they cannot comfortably evaluate statistical computing work and may end up making "safe" decisions instead. It would be valuable to develop ways to help academic committees responsible for search, promotion and tenure to evaluate people's work in statistical computing. We should encourage review of software and computing ideas with a critical eye that filters the "glitz" from substance.

We should also try to encourage the idea that software is a key element of the research, and that it is comparable to published papers. Indeed, software allows many others to discover and interact with new ideas quickly and flexibly. And developing high quality software is frequently more time consuming than authoring papers.

Syllabus The teaching of statistical computing tends to be done in an ad-hoc fashion as part of more classical statistics courses. Often the focus is on learning the tools of the computer rather than ideas. One reason for the lack of more focussed courses is the small number of people that are prepared to teach them.

It would be valuable to understand what topics might be usefully included in a statistical computing course and what we would like students to learn. Should we teach them LaTeXor Word; algorithms for linear algebra; programming in C or Java; or focus on environments such as database technology?

Encouraging New Students In order for the field of statistical computing to flourish, we must encourage students to study the subject and consider it to be a valid, valuable and sustainable research area for them. Some people dissuade students from engaging in statistical computing research because of the difficulties of being rewarded for it.

We also have to present it as an interesting and exciting area. This will most likely involve collaboration with other disciplines such as computer science and biology. An increasing number of students are studying statistics from inside other departments and thus we are "losing" students to more attractive presentation of very similar material. While this may in fact be a good thing, we should discuss our response to it: try to arrest it, or embrace it?

We might discuss what are general areas of interesting and "good" statistical computing research.