Statistical Research: Some Advice for Beginners
Michael Hamada; Randy Sitter
The American Statistician; May 2004; 58, 2; ABI/INFORM Global
pg. 93

General

Statistical Research: Some Advice for Beginners

Michael HAMADA and Randy SITTER

Editor's Note: Research is essential to the health and growth of the statistics discipline. The following article discusses some basic strategies for doing and presenting research based on the authors' experience and conversations with other statisticians. The August 2004 issue of *The American Statistician* will feature a discussion on the topic "How to do Statistical Research." All readers are invited to contribute to this special section. Discussion about this article or general perspectives on being a researcher in the discipline of statistics are welcome. Because of space limitations, we ask that your contribution not exceed 500 words. Articles received by the *TAS* editorial office (tas@bgnet.bgsu.edu) by June 4, 2004, will be considered for publication. —James Albert, Editor, *The American Statistician*

For new graduate students, we discuss issues and aspects of doing statistical research and provide advice. We answer questions that we had when we were beginners, like "When do I start?", "How do I start?", "How do I find out what has already been done?", "How do I make progress?", "How do I finish?", and "What else can I do?".

KEY WORDS: Finding problems; Identifying literature; Presenting; Reading papers; Writing.

1. INTRODUCTION

In an academic environment, where most researchers start, it is easy for the beginner to focus too narrowly on a thesis, a paper in a journal, and/or a talk at a conference as the goal of a research effort. It is important, however, to realize that doing research is a continuous process of discovery that is usually not apparent, is difficult to anticipate, and is also difficult to quantify. It is this process you need to learn, and because doing research is a creative process, there is no one way or right way of doing it; you need to discover what strategies work best for you. This article discusses some basic strategies for statistical research based on some of our experiences and those elicited from our colleagues.

This narrative is undoubtedly biased by our personal views, driven by our individual experiences as we progressed through graduate school to eventually become research statisticians. Because the development of one's own research style is unique and personal, we do use some examples of our own successes and failures as students, researchers, and graduate student supervisors. We trust that you will accept these in the spirit in which they are offered and not interpret them as self-promotion or self-deprecation. In any case, we hope that this article helps you think about and work on developing your own research process by identifying some issues, suggesting some activities, and providing a list of resources in the statistical literature to aid you toward this goal.

So where should we start? As a beginning graduate student, you need to begin to understand that graduate school is a time to start the transition from a primary focus on learning basic techniques and methods to curiosity-driven investigation into the unknown. To us, this can be in part described as a transition from "being taught" or "expecting to be told" or searching for a "correct answer" to asking questions like "Why is it done this way?", or even more importantly "Why is it not done another way?". Some students may understand this as a basic principle, but may still not have any idea as to how to go about it. This prompted the first author, while a new graduate student, to ask of one of his professors, "How do you do research?" The professor was kind enough to reply by E-mailing Mosteller's summary of how L. J. Savage did research (Mosteller 1981):

1. As soon as a problem is stated, start right away to solve it; use simple examples.
2. Keep starting from first principles, explaining again and again just what it is you are trying to do.
3. Believe that this problem can be solved and that you will enjoy working it out.

Michael Hamada is a Technical Staff Member in Statistical Sciences, Los Alamos National Laboratory, Los Alamos, NM 87545 (E-mail: hamada@lanl.gov). Randy Sitter is Professor, Department of Statistics and Actuarial Science, Simon Fraser University, Burnaby, BC V5A 1S6 Canada. The first author thanks Gouri Bhattacharyya, George Box, Murray Clayton, Tom Leonard, Rick Nordheim, Kam Tsui, and Jeff Wu, who helped him in developing his research process as a graduate student at the University of Wisconsin–Madison. The second author thanks John Petkau, Jeff Wu, and Jon Rao for their similar and invaluable aid as he progressed from master's student to Assistant Professor, at the University of British Columbia, University of Wisconsin–Madison, University of Waterloo, and Carleton University. We thank George Duncan, Todd Graves, Crystal Linkletter, Sallie Keller-McNulty, Harry Martz, Laura McNamara, David Scott, and Greg Wilson for helpful comments on an earlier version of this article. We also thank the Editor, Associate Editor, and referees for many invaluable suggestions and even their own advice on doing research which we have incorporated. Finally, we thank Art Dempster, Julian Faraway, Scott Grimshaw, Alfred Hero, Valen Johnson, Jon Kettenring, Jerry Lawless, Bill Meeker, John Nelder, Frank Samaniego, and Lynne Stokes for providing helpful comments and additional strategies for doing research.

© 2004 American Statistical Association DOI: 10.1198/000313004329

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

4. Don't be too hampered by the original statement of the problem. Try other problems in its neighborhood; maybe there's a better problem than yours.
5. Work an hour or so on it frequently.
6. Talk about it; explain it to people.

There is a lot of wisdom provided here from a statistical giant, but it assumes that a potential research problem has already been identified. So, in writing this article, we decided to step back to our beginnings and ask the questions that were on our minds as we grew as researchers. These questions are the section titles of this article. First, "When do I start?", which speaks to a mindset for doing research, immediately followed by, "How do I start?", where you begin to identify potential research problems or a problem area. This naturally leads to, "How do I find out what has already been done?" where you refine the research problem and study the problem area. Then we consider "How do I make progress?", to which much of Savage's advice applies, and "How do I finish?" which discusses fleshing out the solution, writing, and presenting. In "What else can I do?", we suggest some useful general activities to enhance your graduate school experience. We conclude with a final salvo of advice for the beginner and a list of issues with references for new Ph.D.'s to think about as they undertake a research career. Note that there is some overlap among the different topics considered and that their order does not imply an order in which they need to be done.

2. WHEN DO I START?

Start now. You might feel that few of the principles or comments of the preceding section apply at an early stage in your graduate studies when courses are being taken, and you should first do your course work, pass comprehensives, find a supervisor, and then begin research. This is a common approach which works well enough, but we feel it is too limiting, thereby missing out on rich opportunities for developing the skills and intuition needed at an early stage in simple situations. For example, we both identified and worked on projects in graduate courses which eventually were extended into portions of our Ph.D. theses; we had no initial intention for this happening and did not have thesis supervisors yet. Of course, there was some luck and prior training involved, but an opportunity for doing research often exists when doing a course project. That is, you should treat it as a research project and adopt some or all of the above points or variations which are amenable to your interests, training, and experience. No matter the specifics of your approach, certain common requirements will always emerge.

Start searching for your project right away to give yourself an opportunity to live it and breathe it as a line of creative investigation rather than treating it like an examination or homework assignment. Try to look for something related to the course but also related to something you know and like. For example, the projects that we alluded to above that eventually were extended into original research both built upon previous courses. The first author was taking a second course in reliability and extended the simple project from his first reliability course to include a covariate. The second author was taking a course in bootstrapping and had just finished a graduate course in sample survey and thus did his project on bootstrapping survey data. This strategy gives you a base from which to jump. Using previous knowledge that interests you ensures an immediate interest and investment in your project and makes it more likely to be compatible with your current skill set and bent, as you will most certainly choose something you liked before; we seldom like what we are poor at. Having an early topic also focuses your learning of the course material itself, as you find yourself constantly relating new topics back to your project, asking questions of yourself (and your course professor) and evaluating aspects with a specific motivation and context in mind.

Of course, you cannot expect to always come upon original research topics in this way. The most important point is to treat your project in this way. By beginning to ask questions and trying to answer them, you begin to understand the fundamentals of addressing unanswered questions. Imagine for instance that you begin such a project and work very hard to discover and/or develop methods beyond what you know. Later, you discover these already exist in the literature or some flaw in your thinking is pointed out by your professor that limits its applicability. You have still begun to learn the process, your process, of doing independent research. Discovering what someone else thought of before should encourage you; after all, you rediscovered it without knowing the result. Having a professor point out limitations in your proposal may be the beginning of a new project; that is, how to overcome these limitations by adapting your approach.

Developing and understanding your own research process need not wait for a project-oriented course, although topics courses that are closer to the frontiers of research certainly provide a richer environment for starting to do research. You can begin at any time by doing simple things such as staying ahead of the professor in a course, reading related material or material that is on the same topic in a variety of textbooks, or reading the papers that are referenced in various sections of the textbook. The primary idea is to change your mindset from one who is told what to do to one who takes the initiative and explores the unknown.

Given the aforementioned benefits of an early start, what if you are several years into your graduate studies and feel that you have not yet begun? Is it too late? Do not be discouraged. Every path is different, and you are in charge of yours. It is never too soon to begin, nor too late. Change your mindset, and begin now. Once underway, you are never certain where the investigation might inevitably lead. With any luck it will take you someplace interesting, unexpected, and fun.

3. HOW DO I START?

This is likely the most daunting of questions for new researchers. What may be surprising to you is that even experienced researchers face this question often when changing directions in their research or opening up new areas of inquiry, or just when they are in a slump. The simple answer can be equally daunting, "Do something. Ask a question and try to answer it." We will elaborate on this.

We think motivation is important. Thus, you should work on something that interests you. For us, working in a new application area not previously considered by statisticians has many benefits; new statistical problems are likely to arise so that any

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

advance you make is likely to be a contribution. In general, solving real problems that have data provide valuable motivation. It is particularly so if there is a subject-matter expert with whom you can work and talk; such collaboration also increases the chances of your research being used.

The most difficult aspect is identifying a problem which is important, unsolved, exciting to you, and within your capabilities. For graduate students, the advice of professors should be heavily relied upon to provide a starting point or point of attack. This can be something quite vague, like direction to a particular recent paper combined with a description of a flaw or limitation in the methods presented there. It can also be quite specific, like direction to a recent method which makes strong assumptions and the suggestion of a small simulation study to investigate the method's robustness to relaxation of these assumptions. Another example is pointing out a recent theorem which makes strong theoretical assumptions and suggesting doing both a numerical and perhaps theoretical investigation of the theorem's validity when the assumptions are relaxed. Note that these are strategies we have used as first points of attack when faced with the same problem and suggestions that we have given to students in courses and/or at the beginning of their thesis work. The professor's main aid to the student lies in his/her experience, developed over many trials and errors, to choose those that are within the capabilities of the student to at least do a good project, with the potential for more. The main thing is to ask a question and, as in Savage's Point 1, begin.

How one proceeds now that a question has been asked seems to vary quite dramatically from researcher to researcher. Two dominant opinions emerged from those we asked, and the remainder were specific examples and compromises between these two. The first is to begin with a thorough literature review; the other is to attack the potential problem yourself first and look to the literature later. We feel that both of these, and thus compromises between them, have strengths and weaknesses. The former results in learning the field and techniques used there and avoids any chance of "rediscovering the wheel" and thus wasting effort. This is a strength. On the other hand, it is time consuming and has the tendency of leading the researcher to view the problem the same way it has been viewed previously. The latter method has the danger of wasting effort re-solving a problem, but avoids too much influence by those who have gone before and may allow "rediscovering the wheel" in a different way and perhaps in a way that has strengths beyond the expected. In any event, a thorough literature review must be done reasonably early so that not too much time is wasted on what has already been done.

We will discuss tools and methods for doing a literature review and for exploring a potential problem in more detail subsequently. A compromise might work best for the beginning graduate student, however. Read a few papers and explore the potential problem. (If a review paper is available, reading it is useful to quickly become familiar with the research area.) Our reasoning goes as follows: (1) you may not have the time for a thorough literature review but reading a few recent related papers is necessary to become versed in the issues and terminology of the area; (2) while discovering something new is always preferable, "rediscovering the wheel" is of nearly the same benefit as discovering something new at such an early stage in a researcher's career if viewed in terms of developing your own research process. It also has the benefit of encouragement. Reading the vast amount of work in an area when you are just starting out can be daunting and discouraging. Young researchers have the habit of reading material out of context and not realizing that many of the results are not obtained in a vacuum but represent the amalgamation of both previous literature, training, and related work by the authors as well as months, and perhaps years, of dedicated effort.

4. HOW DO I FIND OUT WHAT HAS ALREADY BEEN DONE?

Whether you begin to do research by exploring an area of interest or by solving a specific problem first and following it up with a review of the area, a thorough literature review will eventually be necessary. First, relevant work has to be identified and then digested.

4.1 Identifying Relevant Work

For new problem areas, there may be few papers to find. If the problem area is well established, much exploration needs to be done. The relevant literature needs to be identified, including books, journal publications, technical reports, and conference papers. These can be identified by library search engines such as the Current Index to Statistics and SciSearch (the Science Citation Index). Also, subject matter search engines should be used such as the Social Science Citation Index and Engineering Index. Repeated use of these search engines is likely needed, because new key words may be identified as you explore the problem area.

Do not forget the World Wide Web, where such resources as JSTOR (journals stored in electronic form) exist. You should not neglect a straight key word search on the World Wide Web either—you may come up with some interesting finds such as applications in other fields or authors' homepages or conference programs and papers. For example, a search with Altavista and Google using the keywords *how*, *do*, *research*, and *advice* identified a number of interesting Web sites about the topic of this paper. Looking at the references of recent papers helps to identify previous key papers. Also, citation indices can be used to identify other recent papers that cite these previous key papers by doing a "cited search." See Krause (1995) for more hints on using the electronic services that are available today.

In exploring a problem area, it is worthwhile making connections between similar problems in different statistical fields. For example, similar problems arise in reliability and survival analysis.

If the problem area is rapidly developing, recent developments may be reported in talks, so looking at conference programs and proceedings may be revealing. Relying on the published literature alone is problematic because the publication process in statistics can take two or more years; the papers appearing today are likely to be at least that old. Thus, it is important to identify the active researchers and groups of researchers in the problem area. Looking at their Web pages may reveal more recent but unpublished work such as technical reports or overheads from

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

talks. Also, do not overlook researchers in industry and government who may be contributing to the problem area.

Talk to your professors, departmental visitors, and fellow students who may suggest other references, researchers, and connections. Finally, there may be researchers on your campus (outside your department) who are doing relevant work. Besides giving you someone to talk to about applications that motivate statistical work, such contacts provide natural external faculty members for your thesis committee.

4.2 Digesting the Literature

Having identified the relevant work, you need to read and understand it. A new researcher is at a disadvantage because he/she may not have much of a perspective of statistics. So, you have a list of references and now must decide what and how to critically read them, always with your specific problem in mind.

Critical reading and thinking is an acquired skill. You are looking for ideas and the types of problems in the research area. Look for discussion on the ramifications of theorems rather than wading through the details of their proofs (at least at the outset). Read abstracts, introductions, and conclusions as you sort your way through a daunting array of related and somewhat related papers. Are data available and how are they modeled? What assumptions are being made about the models? What are the issues? Are there new issues? Is it an analysis or design (either experimental or survey) problem? What methods are employed (e.g., nonparametric or parametric, frequentist, or Bayesian)? How can the problem be extended? Taking some notes or using a more formalized question-and-answer form may be helpful. See Murphy (1997) for a form that he uses. Gleser (1986), whose focus was on refereeing, provided other questions that you should be asking such as, "Is the solution novel?" or "Can it be used to solve other unsolved problems?". To help understand the ideas or methods, try them out on simple examples. Murphy (1997) provided other useful advice such as stating the problem in your own words and terms.

There may be published literature reviews or bibliographies which can be invaluable (e.g., *International Statistical Review*). If not, we suggest doing a graduated literature review. For example, suppose that there are several hundred papers and a couple of books on the topic in which you are interested. In trying to explore this literature, one might first look to identify what the problems or applications are (e.g., univariate versus multivariate, finite versus infinite population). Then, on the next pass, identify what models and assumptions are used. Next, consider the methods used (e.g., nonparametric versus parametric, Bayesian versus frequentist). In subsequent passes, look at more details as appropriate. Take notes to capture what you are learning. Reading is a lot easier when you are looking for something specific. A graduated literature review provides a specific focus each time you read and reread the papers.

5. HOW DO I MAKE PROGRESS?

Whether you choose to do a thorough literature review first or a less thorough one, the points above on critical reading still apply. The next issue is how to attack your specific problem, having identified it as interesting and worth your time.

5.1 Attacking Your Problem

Savage's six points listed in the Introduction provide good advice on making progress toward solving a problem. Among our favorites are using simple examples, consistently spending time working on the problem, and explaining it to others. Regarding Savage's Point 5, meeting your research supervisor every week whether you have done something or not can provide the needed motivation. The professor who pointed us to Savage's list amplified Point 4 in a follow-up conversation with the encouragement: "Dare to be courageous—make many conjectures. Some may even be right." This last point may be the most important of all. Do not be afraid to try out your own ideas.

Researchers in today's environment have a huge advantage. They have fast computing power. Use the computer! Simulation immediately comes to mind as an important research tool. Generate random configurations to explore the possibility of a counterexample to your conjecture. In experimental design, one can use an optimization algorithm to find the best design according to some criterion; observing geometric patterns in the best designs suggest the possibility of constructing them using combinatorial theory.

We encourage young researchers to use the computer as their laboratory for investigating statistical ideas. Use the computer to try examples to explore whether conjectures are correct or not. It is much easier to prove something when you are confident it is true. The second author recalls a paper he and a co-author had nearly written. One simulation did not seem to back up the theoretical results. After extensive reprogramming and reconsideration of the proof, the proof finally lost out. A disappointment, but as is often the case, one which pointed out some subtleties that eventually led to further research. Neither of us have the computational skills of a professional programmer, and both of us are essentially self-taught. Even such modest skills allow us to explore our own analysis and design ideas, however.

Using the computer as a laboratory forces you to evaluate specific examples or cases when considering a more general idea. Begin with something simple and extend outwards toward the more complex. Try situations near the edges of your assumptions. Do not be discouraged if you discover that your result is not as general as you had hoped. Finding out why may be more important than the original idea itself. For example, imagine that there is a method in the literature that requires a certain assumption. You read about this and have an idea for a new method which is very different but does not seem to require this assumption. So you try some simulations comparing the performance of the two methods in situations where the assumption holds, nearly holds, and does not hold, and to your disappointment your new method never outperforms the existing method. Fortunately, you did not merely view their relative performance but also their actual performance and realize that the existing method does not seem to require the assumption. You have now identified a new research problem: can you prove that the existing method does not require the assumption? Even if you cannot, you can still design a more thorough simulation study that demonstrates the robustness of the method to this assumption and perhaps someone else will be able to prove that the assumption can be replaced by a weaker one . . . and thus science advances.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Having said this, the computer is no substitute for thought and understanding. One should work through some small examples. Try to understand the workings of the method by walking through it with some small made-up example and then with a real and/or simulated dataset. Never forget the computer is merely a tool to speed up your learning process and not a substitute for the need to think.

5.2 The Moment of Discovery

One over-riding and fundamental truth, in our opinion, is that there is no substitute for hard work. We have both experienced long stretches of discouraging attempts with no results, only to solve the problem in an hour. This is a recurring story heard from many. The reason is likely related to Mosteller's points that underly Savage's research process. One needs to live with the problem, have it percolate, think about it over and over until it is always there in the back of your mind. Then one day, that random variable of all random variables, the mind, puts things together in a slightly different way and it is solved. The solution often then seems simple and obvious. Do not be fooled. It only seems this way to one who is intimately immersed in the problem. Certainly, it has been the authors' experience that more often than not the five-minute solution comes only after weeks and months of "banging your head against the wall."

This is the moment which we enjoy the most. It lends itself well to a sports analogy. If you have ever played baseball or a racket sport or golf, it is that stroke when you hit the sweet spot and the ball seems to explode away with almost no effort on your part. It is also the culmination of hard work and practice. At least for most of us, it seemingly happens without any change in how we approach the problem and at random intervals between tries. Enjoy it. You now, however, have to finish the research project or it will be meaningless.

6. HOW DO I FINISH?

6.1 Beginning to Finish

Once the dust settles from your moment of discovery, and you have what you feel is a new result or a set of new results, you must learn the process of developing a finished product or products. This might mean a project write-up and presentation, a thesis or thesis chapter, a conference presentation, or a paper. What is certain is that it will require an organized communication of what you have accomplished, how it fits into what has previously been done, why it is important and interesting, and what problems still exist. In other words, you now have the basic plot of a story and you need to tell the story well.

One should begin by "playing devil's advocate," where you criticize and challenge your own result until you are satisfied with its accuracy and you understand its advantages and disadvantages. Identify competitors. Then explore where your method wins and loses as to relative performance, through application to real examples, through a simulation study, and/or through theoretical comparison. Do not be easy on yourself. Anticipate what others might ask as if it were someone else's idea. Try to defend the method against these mock challenges, but try to be fair, pointing out the strengths and weaknesses of your ideas. Do not expect your method to win everywhere and do not be shy about admitting that this is so. Instead, view these as potential new research problems.

Take it to others. This often begins with informal discussions with a professor or a classmate, or presentation and discussion of the work at informal graduate student presentations. You should talk with others about your research project, whether it is a course project, your thesis work, or merely a paper you are currently trying to read and understand. This is not necessarily for the purpose of getting assistance or even a different viewpoint, although these can be invaluable. There is inherent value in the process of articulating your ideas to someone else. It forces you to clarify, and in doing so refocus, your thinking about your project. More formal collaborations can also develop in this way when another person takes your ideas and runs with them using his/her complementary skill set and experience. There is often a natural reluctance to exposing your original ideas to the inspection of others, but it is a fundamental aspect of the research process.

The writing process can play a similar role. Writing is more formal than verbal communication and requires a more precise structuring of ideas. When you begin to write down your story you may find holes in your plot that need filling and characters which need more development. Some researchers begin writing very early more as a means to organize their thinking than in anticipation that what they write will remain unchanged or be used in its initial form. Others prefer to delay writing to a later stage and use more of a story-board approach until they feel they are ready to write the story fully.

You should also remain ever in search of new and different problems as you flesh out your story. Often, the process of finishing can be the beginnings of something new. As you develop the story you wish to tell, invariably subplots arise that do not entirely fit into the story but could themselves become a new story, an extension to a different context, a potential generalization, or even something entirely different that uses a similar technique. The best papers or theses often pose more new questions than they do answers.

These are merely suggested strategies and general requirements that apply to almost any finishing process. Of greatest importance is that you explore and discover what works best for you. At an early stage this will require self-examination and interaction with others. Finally, you will have a mature story to communicate and are now faced with the task of writing and/or presenting a cohesive final product.

6.2 Writing It Down

Writing is an important and creative part of the research process. It helps to focus and organize the research you are working on. Transferring your ideas to paper has a way of revealing the deficiencies in what you have done and often shows you where to proceed next. Given that, it is important to begin some form of writing soon, but not necessarily to write the whole paper or even formally written sections. We find that it is easier to correct, change, and work from an existing document than to start something new. Consequently, constructing a rough outline and quickly typing your material, disregarding eloquence, provides the "existing document" in short order from which you might work and build. If you have some results, write those up first in a short form. It is a common mistake to start with the first

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

sentence of the first paragraph of the introduction and write the perfect sentence and then move on to the next sentence. This takes a long time and often the resulting draft has to be completely rewritten. One suggestion is to draft the main sections of the paper first before writing the introduction and discussion sections; knowing what is in the main body of the paper makes it easier to introduce and discuss.

Once you have a working draft that contains all the key results and conclusions, you will be faced with the task of creating a final paper that you wish others to read. How does one go about doing this? What are good things and bad things to do, in terms of presenting your work in written form? Although you might have read many papers by this time, you probably have not paid much attention to how they were written. Invariably, if you do so, you will discover that part of the reason you enjoy some papers more than others is the way they are written. When you realize a particular paper is enjoyable for you to read, ask yourself, "Why?". In this way you will learn what you enjoy as a reader, and it should help you also realize that you must think of the reader when writing. You should ask yourself questions in this regard. What are the main results you wish to communicate? How can you help the reader to understand and enjoy what you have to say? How can you capture the reader's interest and hold it? How do you write mathematics?

As with questions concerning the research process itself, there are no pat answers to these questions. You must develop your own style. There are, however, some general guidelines and some references that may assist you in doing so. One such article which we feel captures some key aspects of technical writing that are particularly relevant to young researchers is Ehrenberg (1982). The paper is short and easy to read and gives sound and generally valuable advice. The most important point, in our opinion, is to sequence the writing for the reader and not in the way you did the work. Key to this is presenting your main results and main conclusions first, perhaps even in the introduction. Capture the reader's interest by allowing the reader to ascertain his/her level of interest in your research early on without getting bogged down in notation, literature reviews, and technical proofs. The best papers do this without use of difficult terminology or mathematics. They whet the reader's appetite by saying what they have accomplished while being necessarily less detailed as to how they did it. This leaves readers wanting to discover how, anticipating how they might do so themselves, and wanting to see if you used a similar approach and if they agree with your conclusions. Then those readers who move on to the details and specific methods have a framework and in some sense a spirit of discovery as they forge ahead.

Many young researchers in statistics feel they do not write well, but attribute this entirely to sentence structure and knowledge of the language and less so to organization of thoughts and techniques of pace and sequencing as mentioned above. In writing papers and theses in statistics or other technical fields, the organization and structure is as important as the specifics of prose. It is more important to be brief and concise, to be accessible to the readership, than to display a breadth of language and use stylized prose. A simplistic, clear, and concise writing style that captures the reader's interest in the technical content and presents the new results well is certainly preferable to an eloquent rendering of poorly organized and sequenced topics, methods, and results.

Having said this, one must still learn to write well at a more fundamental level. Some papers helpful to young researchers that discuss technical writing and publishing are: Halmos (1970), Gopen and Swan (1990), Gbur and Trumbo (1995), Smith (1996), and O'Brien (2001). Papers on refereeing (e.g., Gleser 1986) are also relevant as they contain questions that you should be asking about your own writing.

One overriding truth in developing your writing style is that there is really no substitute for writing and rewriting. Practice is required and no amount of study can replace it. Much like the research process itself, begin early and keep at it. Also, set your writing aside, let it percolate, and re-examine it at a later time.

Read your papers out loud as this forces you to read every word and really "hear" what is written. Have others read what you have written. Besides your Ph.D. supervisor, ask your fellow students to look at your writing. Choose a variety of readers such as one who may be strong technically, another who has an applied viewpoint, and another who is a good writer. Also, read others' draft papers as this can help you to see writing problems which you can learn to avoid.

We conclude this subsection with some specific comments on writing mathematics. Remember that equations are independent clauses which require punctuation; for example, end them with a period if they stand alone or follow them with a comma if independent clauses follow that explain terms or symbols in the equation. Try not to start a sentence with a symbol, for example, rather than σ^2 use "The variance σ^2 ...." Finally, use consistent notation throughout the paper making sure to define each new symbol as you use it and taking care to not use the same symbol for more than one thing.

6.3 Talking About It

Presenting your final product is an important part of the research process, as well. You want others to know of your results much earlier than the publication process allows. By presenting, you want to interest others in what you are doing, to gauge the audience's response to your work, and to obtain constructive criticism on where to go and what gaps might exist. You may be pleasantly surprised at how supportive senior researchers are of young researchers' ideas and work.

Preparing a presentation is also creative and is creative in an entirely different way than writing. There have been times when we have prepared a presentation from a paper and were forced to completely reorganize and rethink the material. At times, this is an inherent aspect of the difference in medium; at other times, having done so, we wished that the paper had been rewritten. Thus, much like writing, one should prepare a presentation early on and then another later with the final product. A presentation can even be prepared before a paper is written, which was the case with this article.

You should give careful consideration to the process of presenting research, and how it differs from written communication. Or perhaps a better way to say this is, you should realize that it in fact *should* differ from written communication. Given this, you again must develop your own style. As with writing, and the research process more generally, there are many pitfalls which

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

every new researcher (and some not so new) should avoid. We will discuss only a few and give some related advice. Some good references with other useful comments and suggestions are: Freeman, Gonzalez, Hoaglin, and Kilss (1983), Brillinger (1993), Becker and Keller-McNulty (1996), and Kalicin (2001).

There are now a number of presentation media available to the presenter. The most common are the overhead slide, 35mm slides, and computer-aided projection. These each have strengths and weaknesses. But most pitfalls lie more in what each slide contains and how one progresses from slide to slide, than in the particular choice of medium. One should use the medium which is best for the presentation and not the "flavor of the month." Do not use fancy computer-aided presentations unless you know you have all the necessary equipment, the knowledge to use it, and the technical support to handle any problems . . . and always have a low-tech backup. The audience will only be distracted by any technical difficulties. Also, if the software is so new that it is slicker and of more interest to the audience than the content of your paper, you will lose them to the technology.

What do you put on each slide? This depends on the talk, of course, but as a general rule one should dedicate each slide to a single idea which requires you a modest amount of verbal explanation. Text should be sparse, easily read from a distance, and easily understood, given the verbal explanation. One should not read from the slide. The slide can contain some key written points, but these are for focusing the audience's attention on the key aspects of what you are saying to them. If possible, one should use graphs rather than tables. Tables of numbers are often impossible for the audience to take in; even when they try, they may not focus on the small corner you want them to focus on. When graphs are used, they should be easy for you to guide the audience through and should be used to make one point, or perhaps two. A slide should not be filled with mathematics or proofs. There is never time for the audience to actually take in such material and the bulk of them will typically have little interest in trying.

Carefully consider the time which you have and the points above when deciding how much material you can realistically cover in the presentation. It is a common failing of inexperienced presenters to attempt to summarize everything in a paper or the entire contents of a thesis into a single 20-minute presentation. It would serve them far better to take one key result and present the problem solution and perhaps a sketch of the novelties of the required proof or the key aspects of the simulations, and do this well.

At the other extreme, one colleague stressed the importance of learning to give a very short (one- to five-minute) summary of their research which a nonexpert can understand and appreciate. Poorly presented research no matter how good it is has a diminished chance of being practiced.

All of these suggestions and the bulk of those in the key references center around one fundamental idea: think as carefully about presenting to an audience as you do about every other aspect of the research process. Do not take it lightly. Treat it as a separate and equally important aspect of your overall research endeavor.

7. WHAT ELSE CAN I DO?

This section lists some activities we feel can enhance your academic experience and the development of your research skills. Many are often mentioned as ways to make your graduate experience more fun and rewarding. We agree, but we also feel that they represent a set of activities that should help you become a good researcher, provided you view them as such.

- Depending on your interests, try taking some science, engineering, social science, or business courses and look for opportunities for statistical thinking and research. Attend seminars in other departments and even conferences in other disciplines looking for the same things. Also, read the other disciplines' journals.
- Talk to your friends in other disciplines about their problems. Working with a business person, a scientist, or an engineer as part of a statistics course can be enlightening.
- Become a project and/or research assistant. It will allow you to work with professors and become involved in research and the practice of statistics.
- Be a summer intern in industry, business, or government to learn about and work on real problems.
- Periodically browse journals to see what is being published. Do not forget the past. It is quite informative to start with the early issues of *Technometrics*, the *Journal of the Royal Statistical Society Supplement*, or *Biometrika*. Look for trends.
- Participate in selected activities as a group. Many departments have a graduate student group that reads and presents papers and practices their thesis proposals on each other. If your department does not have such a group, organize one.
- Meet with department visitors and prepare some questions beforehand. Visitors on extended stays will probably not mind being invited for a home-cooked meal.
- Participate in writing grant proposals. Learn about potential sources of funding for your research.
- Help a professor referee a paper.
- Accompany a professor on a consulting trip. If your advisor is driving to visit a department, ask to go with him/her. Just think of the hours of uninterrupted time, with no phone calls and no knocks at the door, that you will have to discuss research with your advisor.
- Have a subject-matter specialist as an active member of your thesis committee.
- Use electronic services such as newsgroups and e-mail, but be circumspect. It is easy to abuse e-mail and try the patience of those you are contacting. The first author received quite a revealing response from John Nelder when he asked by e-mail how generalized linear models arose. That e-mail led to further discussions and collaboration on a paper about the application of generalized linear models.
- Attend some conferences. Drive to nearby ones or ask your department to send you. Your university or department may have funds to help you.
- Attend seminars. It is likely that no one seminar will produce an epiphany, but the cumulative effect of attending seminars provides perspective. Develop a set of seminar questions and then ask them. Ask that some seminars be directed toward the graduate student audience with some emphasis on the research process.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

• Organize a seminar and get your professors to talk about how they do research. We have used the following format. The faculty were e-mailed beforehand asking for their participation. The students were asked to anonymously provide questions. In the seminar, we gave a 35-minute talk based on this article which was followed by the students' questions. The faculty then responded to their questions.

• Use statistics to solve seemingly simple-minded problems. Years ago, the first author and his young daughter tried unsuccessfully to make a bubble solution (a mixture of water, dishwashing soap, and glycerin) for her toy applicator; the failure was blamed on the local hard water. This spurred interest in mixture experiments which eventually led to work considering robustness to variables that one had little control over, such as water hardness (Steiner and Hamada 1997).

• Make a map, a physical representation, of your research. It can help organize your thinking and determine where you are. This is not unlike a crib sheet prepared by a student who has organized a semester course's material. You need to develop a representation that works best for you, for example, lists, flowcharts, and so on. One of our colleagues mapped and tracked his Ph.D. thesis research on a large sheet of butcher block paper tacked up on his office wall. Where do the material from the courses you take, the papers you read, the talks you attend, the discussions you have, etc., fit into your map?

• For perspective, read recent books on the history of statistics and probability (Hald 1990; Stigler 1986), biographies of statisticians such as Fisher (Box 1978) and Neyman (Reid 1982), key papers [*Breakthroughs in Statistics* by Kotz and Johnson (1982); see also Savage (1970)], interviews of famous statisticians in *Statistical Science* (e.g., D. R. Cox by Reid 1994) and famous statisticians' views of statistics (Box 1976; Rao 1993). Learn about the impact of statistics [*Chance* magazine; *Statistics: A Guide to the Unknown* by Tanur (1978), and where the profession is going ("Statistics in the Year 2000: Vignettes," *Journal of the American Statistical Association*, Volume 95, on the life and medical sciences, social science, business, physical sciences, and engineering, and theory and methods.]

8. SUMMING IT UP

In this article, we have given strategies for doing statistical research. Here is a short summary of our advice.

• When do you start? Right away. It is never too soon.

• How do you start? Do something. Ask a question and begin.

• How do you find out what has already been done? Hunt it down with every available weapon, but consume it carefully and digest it slowly.

• How do you make progress? Live it and breathe it. Months of banging away yields a moment of discovery.

• How do you finish? Sharpen your story under fire. Then tell it well.

• What else should you do? Anything and everything that can help you have fun exploring the unknown.

See Kempthorne, Mukhopadhyay, Sen, and Zacks (1991), Bolker (1998), and Paydarfar and Schwartz (2001) for more discussion and advice.

There are many issues that we have not addressed. These include finding an advisor and working with him/her (Bolker 1998), the role of ethics (Vardeman and Morris 2003), the publishing process, managing your time, making professional contacts, collaborating as a member of a cross-disciplinary team, bringing research into practice, developing a taste in problems, and what happens after graduation. After your Ph.D., we predict that you will still be developing your research process for several years. Some helpful references for things to expect after graduation include Sindermann (1962), Medawar (1979), Trumbo (1989), Altman et al. (1991a, 1991b), Pendergast (1993), Stasny (2001), and Perl and Meyer (2002).

When this material was first presented, one of the authors' undergraduate students attended. After the session he came up and commented, "Gee, you can apply this to anything!" We are not sure if we want to make such a sweeping claim although we sure felt good. Nevertheless, we think this article is relevant as well to master's students who are just beginning to develop their view of statistics and even undergraduate statistics majors.

Finally, take responsibility for developing your own research process and work at it!

[Received July 2003. Revised February 2004.]

REFERENCES

Altman, N., Banks, D., Hardwick, J., and Roeder, K. (1991a), *New Researchers' Survival Guide*, Institute of Mathematical Statistics.

Altman, N., Banks, D., Chen, P., Duffy, D., Hardwick, J., Leger, C., Owen, A., and Stukel, T. (1991b), "Meeting the Needs of New Statistical Researchers," *Statistical Science*, 6, 163–174.

Becker, R. A., and Keller-McNulty, S. (1996), "Presentation Myths," *The American Statistician*, 50, 112–115.

Bolker, J. (1998), *Writing Your Dissertation in Fifteen Minutes a Day*, New York: Henry Holt and Company.

Box, G. E. P. (1976), "Science and Statistics," *Journal of the American Statistical Association*, 71, 791–799.

Box, J. F. (1978), *R.A. Fisher: The Life of a Scientist*, New York: Wiley.

Brillinger, D. R. (1993), "Why I Prefer to Present Posters at Conferences," *Biometric Bulletin*.

Ehrenberg, A. C. S. (1982), "Writing Technical Papers or Reports," *The American Statistician*, 36, 326–329.

Freeman, D. H., Gonzalez, M. E., Hoaglin, D. C., and Kilss, B. A. (1983), "Presenting Statistical Papers," *The American Statistician*, 37, 106–110.

Gbur, E. E., and Trumbo, B. E. (1995), "Key Words and Phrases—The Key to Scholarly Visibility and Efficiency in an Information Explosion," *The American Statistician*, 49, 29–33.

Gleser, L. J. (1986), "Some Notes on Refereeing," *The American Statistician*, 40, 310–312.

Gopen, G. D., and Swan, J. A. (1990), "The Science of Scientific Writing," *American Scientist*, 78, 550–558.

Hald, A. (1990), *A History of Probability and Statistics and Their Applications Before 1750*, New York: Wiley.

Halmos, P. R. (1970), "How to Write Mathematics," in *P. R. Halmos, Selecta Expository Writing*, eds. D. E. Sarason and L. Gillman, New York: Springer-Verlag, pp. 157–186.

Kalicin, S. (2001), "Ughh!! Technical Presentaions!," *Amstat News*, September, 26–28.

Kempthorne, P., Mukhopadhyay, N., Sen, P.K., and Zacks, S. (1991), "Research—How to Do It: A Panel Discussion," *Statistical Science*, 6, 149–162.

Kotz, S., and Johnson, N. L. (eds.) (1992), *Breakthroughs in Statistics*, Volumes I, II, and III, New York: Springer Verlag.

Krause, A. (1995), "Electronic Services in Statistics," *Computational Statistics and Data Analysis*, 19, 595–604.

Medawar, P. B. (1979), *Advice to a Young Scientist*, New York: Harper and Row.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

Mosteller, F. (1981), "Memorial Service Tributes" in *The Writings of Leonard Jimmie Savage—A Memorial Selection*, American Statistical Association and Institute of Mathematical Statistics, pp. 25–28.

Murphy, J. R. (1997), "How to Read the Statistical Methods Literature: A Guide for Students," *The American Statistician*, 51, 155–157.

O'Brien, R. G. (2001), "Discover the Science of Technical Writing," *Amstat News*, September, 33–36.

Paydarfar, D., and Schwartz, W. J. (2001), "An Algorithm for Discovery," *Science*, 292, 5514, 13.

Pendergast, J. F. (1993), "Issues in Grantsmanship or I want Money . . . Lots and Lots of Money . . . I want the Pie in the Sky . . .," *Proceedings of the Section on Statistics and Education*, Alexandria, VA: American Statistical Association, pp. 42–47.

Perl, M. L., and Meyer, M. A. (2002), "The Practice of Experimental Physics—Recollections, Reflections and Interpretations," *Theoria et Historia Scientiarum*, Torun: Nicolas Copernicus University, 6, 205–240.

Rao, C. R. (1993), "Statistics Must Have a Purpose, The Mahalanobis Dictum," *Sankhya*, Ser. A, 55, 331–349.

Reid, C. (1982), *Neyman from Life*, New York: Springer-Verlag.

Reid, N. (1994), "A Conversation With Sir David Cox," *Statistical Science*, 9, 439–455.

Savage, L. J. (1970), "Reading Suggestions for the Foundations of Statistics," *The American Statistician*, 24, 23–26.

Sindermann, C. J. (1962), *Winning the Games Scientists Play*, New York: Plenum Press.

Smith, W. B. (1996), "Publication is as Easy as C-C-C," *STATS*, 16, 12–14.

Stasny, E. A. (2001), "How to Get a Job in Academics," *The American Statistician*, 55, 35–40.

Steiner, S., and Hamada, M. (1997), "Making Mixtures Robust to Noise Factors and Measurement Errors," *Journal of Quality Technology*, 29, 441–450.

Stigler, S. (1986), *The History of Statistics: the Measurement of Uncertainty Before 1900*, Cambridge: Belknap Press of Harvard University.

Tanur, E. (ed.) (1978), *Statistics: A Guide to the Unknown* (2nd ed.), San Francisco: Holden-Day.

Trumbo, B. E. (1989), "How to Get Your First Research Grant," *Statistical Science*, 4, 121–150.

Vardeman, S. B., and Morris, M. D. (2003), "Statistics and Ethics: Some Advice for Young Statisticians," *The American Statistician*, 57, 21–26.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.