

DRAFT

Undergraduate Committee Report on Computational and Communication Skills for Undergraduate Majors in Statistics

1 Introduction

The American Statistical Association recently developed Curriculum Guidelines for Undergraduate Programs in Statistical Science (see www.amstat.org/education/Curriculum_Guidelines.html). There are five broad areas listed under skills needed to be effective statisticians. They are; Statistical, Mathematical, Computational, Non-mathematical, and Substantiative Area. The undergraduate program in statistics at Iowa State University does a very good job in providing background in most of these areas. However, some areas have weaknesses that should be addressed. Specifically, the Computational and Non-mathematical skills of our undergraduates can, and should, be strengthened. Recent changes in the Department, the change from Minitab to JMP as the introductory level statistical computing package, will have an effect on what computing packages will be used in higher level courses. Efforts across the University to strengthen communication skills (e.g. AgComm and ISUComm) provide guidance for improving some of the Non-mathematical skills. Given these developments, the Undergraduate Committee decided to review the undergraduate curriculum and to make suggestions for improvements.

2 Current Undergraduate Program

The current undergraduate program in Statistics at Iowa State University matches the ASA Curriculum Topics for Undergraduate Degrees in Statistical Science in many areas. The mathematics support courses we require (Math 165, 166, 265 and 307 or 317) cover calculus through multivariate calculus and applied linear algebra as recommended by the ASA. Our core courses (Stat 101, 341, 342, 401, 402 and 421) cover most of the topics under Probability and Statistical Topics listed in the ASA Guidelines. Students are exposed to professional statistical software in a variety of courses. Up until Fall 2000, students were introduced to SAS in Stat 401 and continued to use it in Stat 402. Extensive work with SAS is done in Stat 479 and with other computing programs (e.g. Excel, Minitab, JMP, etc.) in Stat 480. The program requires a computer applications course (Com Sci 103), a technical writing or business communications course (Engl 314 or 302) and an oral communications course (Com St 102 or Sp Cm 212). Together with a requirement of 2 additional statistics courses and a recommendation to have a minor in some substantive area, provides our undergraduate majors with virtually all of the components contained in the ASA Guidelines.

Even so, there are areas mentioned in the ASA Guidelines that are weak or lacking in our current program. Specifically,

- **Exposure to programming concepts.** Up until the 1999-2001 Catalog, the undergraduate program required a computer programming course. This was a FORTRAN course up until Computer Science stopped offering it. It was then a C++ course more recently. This re-

quirement was dropped because the Computer Science courses were not adequately preparing our students. Currently, students are exposed to programming concepts in Stat 479 and 480, although more should be done.

- **Data base concepts/management.** There is virtually no exposure to data base concepts/management within the current courses for our undergraduate majors. This topic is one of the ones cited most often by employers of our graduates as an important topic.
- **Exposure to advanced modeling topics.** The addition of Stat 415, Advanced Statistical Methods for Research Workers to the 2001-2003 Catalog may have an impact on this area.
- **Graphical data analysis.** The addition of Stat 415, Advanced Statistical Methods for Research Workers to the 2001-2003 Catalog may have an impact on this area.
- **Synthesis of theory, methods and applications.** Some programs, like the one at the University of South Carolina, have a capstone course or experience that provides students with the opportunity to tie together many ideas of statistics.
- **Frequent opportunities to develop communication skills.** An outcome for our undergraduate students is the ability to communicate statistical ideas and results. Currently there is no systematic way of assessing this.
- **Team work and collaboration.** There are team projects in some of our core courses but there is no systematic means of assuring all our undergraduate students have these opportunities.

Lending support for these items is a recent job announcement for a Data Analyst at the Iowa Foundation for Medical Care/ENCOMPASS. The requirements for this job are representative of those our undergraduate majors must satisfy. Among the requirements are;

- Four year degree in statistics, mathematics or related field and/or equivalent training and/or experience.
- Demonstrated experience writing SAS code in a Windows environment for purposes of data management, data cleaning, data analysis and reporting.
- Demonstrated experience working with relational databases such as Oracle.
- Demonstrated experience using MS Access as an analysis tool, data reporting system and data management system.
- Demonstrated ability to formulate and communicate effective and scientifically sound methods of sampling, collecting, analyzing and displaying data.
- Demonstrated ability to succeed in a team environment.

3 Recommendations

3.1 Computing

- Make JMP the statistical analysis package for the undergraduate sections of Stat 401 and 402.

JMP is the statistical analysis package for all introductory level statistics courses. Because of JMP's flexibility and wide distribution on campus (it is free and easy to download from Scout), it is very popular with students. JMP has virtually the same functionality as SAS and has superior dynamic graphic capabilities. This spring in the undergraduate section of Stat 402, although SAS is the default analysis program, virtually all of the students are using JMP. In Fall 2001, Profs. Duckworth and Stephenson will use JMP in their sections of Stat 401. Although SAS output will be presented in class, students will use JMP to do labs and homework.

- Introduce undergraduate students to Splus in Stat 341-342.

Splus is a modern statistical programming package that is widely used both in academia and industry. Undergraduate statistics majors need to be exposed to Splus in order to be prepared for graduate school and/or employment. Prof. Froelich has incorporated some use of Splus in Stat 341-342 during the 2000-2001 academic year. She wishes to continue and expand this in future years. The ability to simulate easily and graph results makes Splus a good choice for presenting some of the ideas of probability and statistical theory.

- Introduce the use of SAS as a data management tool in Stat 479.

Much of the work our B.S. graduates do with SAS in the workplace revolves around the use of SAS as a data management tool. As indicated in the job description above, this is a skill our students should have. Since students will not be exposed to SAS as much in Stat 401 and 402, it seems appropriate that their introduction to SAS should begin with data entry, data cleaning, and data management. The use of SAS as a data management program also allows one to introduce programming concepts such as DO loops, IF-THEN-ELSE statements, sorting and the like. One can also use queries to SAS to construct subsets of a larger data set. SAS will also be used as a data analysis tool in Stat 479. Because of the changes in content and intent, Stat 479 would become a prerequisite for Stat 480.

- Include exposure to other data base management tools such as Microsoft Access and Oracle in Stat 480.

This is a natural follow up to the use of SAS as a data management system.

3.2 Communication

- Include several opportunities to develop communication skills in the undergraduate sections of Stat 401 and 402.

Examples of this are:

- Data on the increase of deaths due to HIV/AIDS is given in a newspaper article. Students are asked to model the increase in the number of deaths exploring several different models (linear, quadratic, cubic, logarithmic). Results of the analysis are written up in the style of a newspaper article.
- Data on the amount of organic compounds at various depths in a section of river downstream from an abandoned dump site are given. Students are asked to analyze the data and report their findings in a memo to the Environmental Protection Agency.

- Have team projects in the undergraduate sections of Stat 401 and 402.

This is currently being done when Profs. Duckworth and Stephenson teach these sections. We would like to make this a formal part of these sections. Examples of projects are:

- Stat 401: Teams of students collect data on the number of bedrooms, number of bathrooms, square footage, number of garage stalls, size of lot and the selling price of homes. The teams would then analyze the data and come up with a multiple regression prediction model for the selling price of homes. The results are presented in a written report.
 - Stat 402: Teams of students design an experiment with at least two factors and with at least one of the factors having three or more levels. The teams decide whether to conduct the experiment as a completely randomized, randomized complete block, Latin square, or split plot design. The results are presented in a written report.
- Have Stat 480 serve as a “capstone” experience through the use of individual and team assignments.

Over the past three years Profs Duckworth and Stephenson have revamped Stat 480 to give students more exposure to the types of computing they will see in the workplace. This includes; work with spread sheets, verifying data accuracy, transferring data between systems, exposure to microcomputer statistical software, macro programming, simulation, software reliability, and interfacing with the World Wide Web. By its nature, this course lends itself to individual and team projects in pulling together data collection, data analysis and reporting results.

- Explore the possibility of having Stat 415 fill in the gaps in our current offering.

Statistics 415: Advanced Statistical Methods for Research Workers is an outgrowth of an NSF funded grant that partially paid for the computers in Snedecor 322. It has been team taught as an experimental course the last three springs. In the new (2001-2003) catalog it will be offered every other spring. We anticipate that future offerings of the course will have fewer individuals involved in the team teaching. This will offer the opportunity to truly field test the modules by having someone teach material that he/she has not written. We see this course as a vehicle for filling in the gaps in our coverage of topics. Specifically, graphical data analysis, non-linear regression (including logistic regression), bootstrap techniques and possibly a new module on log-linear models would cover many of the advanced modeling techniques mentioned in the ASA Guidelines.