

## PROPOSAL TO THE SENATE COMMITTEE ON EDUCATIONAL POLICY

**TITLE:** Master's Degree in Statistics: Concentration in Analytics

**SPONSOR:** Jeffrey A. Douglas, Associate Professor, Department of Statistics, Phone: 244-7302, e-mail: jeffdoug@uiuc.edu

**BRIEF DESCRIPTION:** This proposed new concentration within the Master of Science Degree in Statistics is designed to respond to the growing demands in science and industry to make use of large databases for classification, clustering, predictive modeling, and various applications of exploratory data analysis. It differs from the current Master's Degree in that it is more centered around statistical computing, data management, and a field of study being referred to as statistical learning, which encompasses the more statistical topics that fall under the broader title of data mining. The proposed new program is designed to retain the strengths of the traditional MS in Statistics but enhance the computational and data analytic sophistication for those planning careers in information intensive industries.

**Requirements:** The concentration requires completing 35-40 hours, depending on whether the first course in probability and statistics, STAT 410, is waived for students entering with credit for the same or an equivalent course, and whether CS 412 is taken with a 3 hour or 4 hour option. The required courses comprising these 35-40 hours are described below.

### **Mathematical Statistics:**

1-2. STAT 410 and STAT 510

STAT 410 is a course in probability and mathematical statistics and prepares students for STAT 510, which is the first, and most practical, of two courses in mathematical statistics that are required for doctoral programs. This course forms the foundation for statistical inference that is encountered throughout the remainder of the curriculum.

### **Applied Statistics:**

3. STAT 425

4. One of STAT 424, STAT 426, STAT 429, STAT 430, STAT 578

STAT 425 is a thorough course in linear regression and data analysis that is fundamental for further study in analytics. The fourth course is a selection of one of several traditional statistics courses including analysis of variance, categorical data analysis, time series, and special topics.

### **Statistical Consulting:**

#### 5. STAT 427 or STAT 593

It is critical for students to be exposed to team problem solving in a consulting or business environment. That is the purpose of STAT 427. An approved internship that entails applying skills in this same way may be used in place of STAT 427, and credit for the internship is given by registering for STAT 593 with 4 hours of credit. (STAT 593 is currently under the approval process. It has been approved by the LAS.)

### **Statistical Computing:**

#### 6. STAT 440

#### 7. STAT 448

#### 8. One of STAT 428, CS 412

STAT 440 focuses on databases, data management, and sampling and develops skills that are vital for students to succeed in information intensive careers. STAT 448 is a survey of important and common methods of data analysis, all taught emphasizing applications utilizing large-scale statistical software. An additional course in computation is required with choices of computational theory and methods in statistics considered by STAT 428 or CS 412, which is an introductory course in data mining.

### **Advanced Analytics**

#### 9. STAT 542

#### 10. One of STAT 525, STAT 571, CS 512

STAT 542 is a new course in statistical learning that covers state-of-the-art and proven modern methods for classification, clustering, model selection, and predictive modeling in the context of large datasets. A tenth class can be a choice of advanced statistical computing theory, multivariate analysis, and an advanced computer science course in data mining.

**JUSTIFICATION:** Virtually every industry is becoming increasingly data intensive in planning processes and business decision making. Based on feedback from alumni and discussions with industry representatives, it has become clear that there is a strong and increasing demand nationwide for people with training in statistics, data mining and analytics. Our alumni contacts at numerous corporations value the solid foundation in statistics they received from our program, but stress the importance of training in predictive modeling and analysis with complex datasets for problems in risk analysis and demand forecasting, for example. It has become apparent that the substantial and growing proportion of our students who take such career paths would greatly benefit from much more intensive training in data management, data mining, and methodology specifically concerning applications with large databases.

The UIUC partnership with the State Farm research center provides an opportunity to accelerate the development of this concentration. State Farm is developing a larger analytics presence in the research park with ties to UIUC through internships and on-the-job training for select graduate students. This offers an excellent mechanism for recruiting, and as the new program becomes available, we anticipate working with contacts in other industries to attract additional internship opportunities.

This M.S. Concentration in Analytics is designed to be consistent with the traditional M.S. program, having the property that satisfying the requirements of the concentration implies satisfying the requirements for the general M.S. degree. A proposal for a revised M.S. degree in Statistics has been submitted that together with this proposed analytics concentration would satisfy this property and improve the M.S. program in general. One difference between the traditional degree and the concentration in analytics is an extra 3 to 8 hours of coursework for the concentration. This is justified by the special skills being targeted by the concentration, together with the need for a broad understanding of theoretical and applied statistics, which is in common with the general M.S. program.

The Department of Statistics also offers the M.S. in Statistics with Concentration in Applied Statistics, with specializations in a number of areas of application. The department is not proposing any changes to this concentration.

## **EXPECTED ENROLLMENT**

We expect to enroll 9-12 students in this program the first year and then about 15-20 students each year after the first.

## **BUDGETARY AND STAFF IMPLICATIONS**

### **a. Additional staff and dollars needed**

Three new courses (STAT 440, STAT448, and STAT542) have recently been introduced, with each being offered once per year. In addition, an additional section per year of STAT425 will open. To help launch the program, State Farm has pledged funds to increase a current part-time faculty member to full-time so that tenure-track faculty members can be reallocated to cover these courses. Also, each of these four sections will require a TA, so an additional 50 percent TA will be needed for both fall and spring semesters.

### **b. Internal reallocations**

As mentioned above, these four new sections will require some internal reallocations, but the burden will be substantially offset by increasing a part-time instructor to full-time.

**c. Effect on course enrollments in other departments**

The requirements listed above include some CS courses, but only as options where STAT courses are also available. With about 15-20 new students per year, we expect roughly 3-5 will select these CS courses, so the impact should be minimal. Professor Michael Heath, Interim Head of the Department of Computer Science, writes,

Jeff,

I have already consulted relevant CS faculty members about your proposal, and we have no objections, so you have my approval.

-Mike

**d. Impact on the University Library**

We expect any library resources needed for this degree program can be accommodated through the current funding structure, which we understand is tight. The following is the response of Tom Teper, Associate University Librarian for Collections, (email to Kristi Kuntz) to this proposal:

While we believe that most of the titles in this discipline are likely covered either by the Mathematics or Grainger Engineering Library's funding, it is possible that there are some titles that we would not have. In recent years, the Statistics funding has been responsive to changes in the direction outlined by the department (e.g., shifting collections to pick up more in bio-statistical research in recent years), but with no monograph budget and a tight budget for serials, we might have some difficulty picking up additional resources should they be necessary. However, we will do what we can to support the proposal.

**e. Impact on laboratory use**

The addition of two statistical computing classes will increase the demands on the existing statistical computer laboratory and classroom significantly. This laboratory, 113 Illini Hall, houses 30 computers and each would need to run the base SAS system as well as SAS Enterprise Miner. In addition, students would need access to database software for a critical part of the new class in data management. All of the software listed above is currently available in 113 Illini Hall, so there will be no additional budgetary implications.

CLEARANCES

Douglas G. Sims 10/26/2006

Department of Statistics

Ann M. Hunt

College of Liberal Arts and Sciences

Lynda Jappanda 12/17/07

Graduate College

---

Office of the Provost