Michigan Tech

Office of the Provost and
Senior Vice President for Academic Affairs

Phone: (906) 487-2440
Fax: (906) 487-2935

TO: Richard Koubek, President

FROM: Jacqueline E. Huntoon, Provost & Senior Vice President for Academic Affairs   *Jacquelin E Huntoon*

DATE: April 24, 2020

SUBJECT: Senate Proposal 45-20

Attached is Senate proposal 45-20, "Establishment of a New Graduate Certificate in Applied Statistics," and a memo stating the Senate passed this proposal at their April 22, 2020 meeting. I have reviewed this memo and recommend approving this proposal.

I concur____✓____   do not concur_____   with this recommendation.

_____   04/27/2020
Richard Koubek, President   Date

DATE:	April 23, 2020
TO:	Richard Koubek, President
FROM:	Michael Mullins University Senate President
SUBJECT:	Proposal 45-20
COPIES:	Jacqueline E. Huntoon, Provost & Senior VP for Academic Affairs

At its meeting on April 22, 2020, the University Senate approved Proposal 45-20, "Establishment of a New Graduate Certificate in Applied Statistics". Feel free to contact me if you have any questions.

The University Senate of Michigan Technological University

Proposal 45-20

Establishment of a New Graduate Certificate in Applied Statistics

Submitted by: Department of Mathematical Sciences

1. **Proposal Date**:
   March 25, 2020
2. **Proposing Contacts and Departments:**
   Mark Gockenbach, Professor and Chair, Department of Mathematical Sciences, msgocken@mtu.edu
   Ying Sha, Graduate Director, Department of Mathematical Sciences, qsha@mtu.edu
3. **Sponsor Department Approvals:**
   N/A
4. **General Description and Characteristics of Certificate:**
   The certificate requires total 9 credits: one core course and two electives. The core course (namely Statistical Methods) provides a basic understanding of the foundations of statistics and an introduction to the design, conduct, and analysis of statistical studies, including statistical computing and the preparation of statistical reports. In addition, it requires two elective courses that can be chosen from a variety of traditional and modern topics: Regression Analysis, Experimental Design, Times Series Analysis, and Predictive Modeling, to name some of the choices. A graduate student earning this certificate will be well prepared to apply statistical techniques to other disciplines (such as Forestry, Economics, Social Sciences, Psychology). The certificate will be offered both on-campus and (entirely) online.

   Recipients of this certificate will be able to:
   1) Choose a statistical method that is appropriate for a given problem, can apply that method, and can draw appropriate conclusions. All students must take Statistical Methods which introduces them to the basic theory of statistics. In addition, each student must take two elective courses on applied statistics which allow them to go into more depth in classical techniques such as Regression Analysis or Experimental Design, or in modern approaches such as Predictive Modeling.
   2) Use popular statistical software to solve realistic problems. Programming and the use of statistical software are introduced in Statistical Methods (MA5701). All of the applied statistics courses (thus two of the three courses required for the program) require the use of software. Students desiring to go into more depth can select Computational Statistics (MA5761) as one of their electives.
   3) Work with real data. They can clean the data, deal with missing data values, and generally appreciate the complexities of handling real-world data. Almost of all of the applied statistics courses assign problems or projects involving real data.

The certificate will be available to degree-seeking students enrolled at Michigan Technological University as well as non-degree seeking students employed in industry and at federal facilities and laboratories. The certificate will be offered both on-campus and (entirely) online. Students must earn a cumulative B average counting toward the certificate. Following the University policy, a maximum of 3 credits is allowed in courses offered at the 4000-level. The Director of Graduate Studies in the Mathematical Sciences will oversee this certificate program.

5. **Rationale for the Certificate**:
   Throughout the economy, companies are making increased use of data to inform their business decisions. Evidence of this is seen at Michigan Tech as companies that have long recruited our engineering graduates are now also looking for graduates with the ability to analyze data. Specific examples include Ford Motor Company (through its Global Data Insight and Analytics group) and Caterpillar (which has a Data Analytics group and also maintains a Data Innovation Lab). Many individuals who end up in these analytics groups are not trained in data analysis, but rather come out of engineering, IT, or business units when they find they enjoy working with data. This points to one of the intended audiences for this certificate: working professionals who have or want to move into a data-centric part of their company. The certificate will be offered both on-campus and (entirely) online, making it accessible to such individuals.

   Anecdotal evidence also suggests that companies would like to be able to hire graduates from STEM programs that already a knowledge of data analytics. More than one recruiter has told our Career Center words to the effect, "What I need is an electrical engineer who can analyze data." Therefore, another intended audience for the certificate consists of Michigan Tech graduate students in STEM disciplines who want to broaden their marketability by adding data analysis to their skill set.

   Moving beyond anecdotes, numerous reports have predicted a shortage of workers trained in data analysis. For example, last year IBM released the report The Quant Crunch: How The Demand For Data Science Skills Is Disrupting The Job Market,[1] which predicted that demand for data scientists will increase by 28% by 2020. The report also points out that, at the time the report was issued, advertised positions for data scientists remained open longer than the average for other professional positions. IEEE Spectrum, which reports on technology, engineering, and science news, recently published an article entitled "Desperate for Data Scientists",[2] reporting on a study from LinkedIn on the shortage of data scientists. Since data science is described as the intersection of computer science, statistics, and domain-specific knowledge, the proposed certificate can help an individual move into the data science field, where jobs are plentiful.

6. **Related Programs:**
   (a) Michigan Tech offers a graduate certificate in Data Science that requires five courses: a survey course on the methods and tools of data science, one statistics course (Predictive Modeling), a computer science course (Data Mining), a course from SBE (Information Systems Management and Data Analytics), and one graduate elective. The proposed

certificate will offer more depth in data analysis, while the certificate in Data Science offers more breadth in techniques, tools, and application areas.

(b) At other institutions (three examples):

    I. The Penn State World Campus offers a 12-credit Graduate Certificate in Applied Statistics. It requires a survey course on applied statistics (comparable to our requirement of MA5701), a course on regression methods, and two graduate electives.

    II. North Carolina State University offers a 12-credit Graduate Certificate in Applied Statistics and Data Management that requires a two-semester sequence in statistical methods and a two-semester sequence in statistical programming.

    III. Purdue University offers a 12-credit Graduate Certificate in Applied Statistics. It requires a two-semester sequence in probability and statistical inference and two electives chosen from a short list.

7. Projected Enrollments Provide projected:

We have capacity for approximately ten on-campus students (the limiting factor is space in the required course MA5701). The provost has established a revenue-sharing plan for online tuition that will allow us to offer more online sections as enrollment expands. Therefore, we should be able to meet any reasonable demand among online students (up to perhaps 100 students).

8. Scheduling Plans:

No change in the regular scheduling of the existing courses is anticipated. All of the courses are offered regularly, including the online courses.

9. Curriculum Design:

The certificate requires total 9 credits: one core course and two electives. The core course is **MA5701** Statistical Methods (3 credits, Fall and Spring, graduate standing required) and the two electives can be chosen from among the following courses. The core course needs to be taken first. Courses typeset in bold face will be available online. Most courses require an introductory statistics course as a prerequisite; any of MA2710, MA2720, MA3710, MA3715, or MA4700 is sufficient.

MA4710 Regression Analysis (3 credits, Fall, introductory statistics)
MA4720 Design & Analysis of Experiments (3 credits, Spring, introductory statistics)
MA5750 Statistical Genetics (3 credits, every other Spring, graduate standing required)
MA5751: Statistical Data Mining (3 Credits, Spring, MA5701)
MA5761 Computational Statistics (3 credits, Fall, (MA4700 and MA5701) or MA4770)
MA5781 Time Series Anal. & Forecasting (3 credits, Spring, intro. statistics and (MA3720 or MA4700))
MA5790 Predictive Modeling (3 credits, Fall, MA4700 and MA5701)
MA5791 Categorical Data Analysis (3 credits, every other Spring, graduate standing required)

Following the university policy, at most one 4000-level course (three credits) is allowed.

10. Course Descriptions:

MA5701: Statistical Methods (3)

Introduction to design, conduct, and analysis of statistical studies, with an introduction to statistical computing and preparation of statistical reports. Topics include design, descriptive, and graphical methods, probability models, parameter estimation and hypothesis testing.

MA4710: Regression Analysis (3)

Covers simple, multiple, and polynomial regression; estimation, testing, and prediction; weighted least squares, matrix approach, dummy variables, multicollinearity, model diagnostics and variable selection. A statistical computing package is an integral part of the course.

MA 4720: Design/Analysis of Experiments (3)

Covers construction and analysis of completely randomized, randomized block, incomplete block, Latin squares, factorial, fractional factorial, nested and split-plot designs. Also examines fixed, random and mixed effects models and multiple comparisons and contrasts. The SAS statistical package is an integral part of the course.

MA5750: Statistical Genetics (3)

Application of statistical methods to solve problems in genetics such as locating genes. Topics include basic concepts of genetics, linkage analysis and association studies of family data, association tests based on population samples (for both qualitative and quantitative traits), gene mapping methods based on family data and population samples.

MA 5751: Statistical Data Mining (3)

Course will cover various topics in statistical data mining, including linear model selection and regularization, regression and smoothing splines, unsupervised learning, resampling methods, tree-based methods, and deep learning. This course will introduce modern statistical data mining techniques and their applications.

MA5761: Computational Statistics (3)

Introduction to computationally intensive statistical methods. Topics include resampling methods, Montes Carlo simulation methods, smoothing technique to estimate functions, and methods to explore data structure. This course will use the statistical software R.

MA5781: Time Series Analysis (3)

Statistical modeling and inference for analyzing experimental data that have been observed at different points in time. Topics include models for stationary and nonstationary time series, model specification, parametric estimation, and time regression models.

MA5790: Predictive Modeling (3)

Application, construction, and evaluation of statistical models used for prediction and classification. Topics include data pre-processing, over-fitting and model tuning, linear and nonlinear regression models and linear and nonlinear classification models.

MA5791: Categorical Data Analysis (3)

Structure of 2-way contingency tables. Goodness-of-fit tests and Fisher's exact test for categorical data. Fitting models, including logistic regression, logit models, probit and extreme value models for binary response variables. Building and applying log linear models for contingency tables.

11. **Model Schedule Demonstrating Completion Time**:
Online courses will be offered on a seven-week schedule (Track A and Track B), with the required course offered every semester (Fall, Spring, and Summer). Thus an online student will be able to complete the certificate in as few as two semesters, while starting the program in any semester. On-campus students are expected to pursue the program on a part-time basis, and thus there is no model schedule. However, the required course (MA5701) must be taken before the elective courses.

12. **Library and other Learning Resources**:
Students in this program will need only the Library resources presently available to all enrolled students.

13. **Faculty Resumes**:
CVs for Ray Molzon, Yeonwoo Rho, Qiuying Sha, and Kui Zhang are attached at the end of the document. All of them have worked with Keypath for the online MS Applied Statistics program and obtained online teaching certificates.

14. **Equipment:**
No additional equipment will be required.

15. **Program Costs**:
The courses are presently being taught on a regular basis. Software is already through IT at Michigan Tech

16. **Space:**
No additional space is required.

17. **Policies, Regulations, and Rules**:

18. **Accreditation Requirements:**
Michigan Tech is accredited by the Higher Learning Commission (HLC) (https:/www.mtu.edu/provost/accreditation/hlcomission/). The proposed certificate will not require additional accreditation. The proposed certificate will meet HLC criteria 3 and 4.

19. Planned Implementation Date:
   Fall 2020.

20. Assessment:
   Students will demonstrate proficiency in the subject matter through the successful completion of the required and elective courses. A portion of the students will be degree-seeking, while others will be non-degree seeking pursuing only the certificate. The Graduate Learning Objectives (GLOs) of the proposed certificate are:
   1) Demonstrate basic understanding of the foundations of statistics
   2) Demonstrate proficiency using advanced statistical methods to analyze real data.
   3) Demonstrate effective written and oral communication skills
   4) Demonstrate effective responsible conduct of the profession

   The assessment of the GLO for the Graduate Certificate in Applied Statistics will encompass required and elective courses listed on the student's Degree Schedule – Graduate Certificate in Applied Statistics (GCAS).

[1]https://www-01.ibm.com/common/ssi/cgi-bin/ssialias?htmlfid=IML14576USEN&
[2]https://spectrum.ieee.org/view-from-the-valley/at-work/tech-careers/desperate-for-data scientists