Applied Statistics (STT) Masters Degree

REPORT PREPARED by: Wang, Weizhen

ACADEMIC YEAR COVERED BY THIS REPORT: 2020-2021

I. PROGRAM LEARNING OUTCOMES

The graduates should master a broad range of modern statistics while establishing a strong ability to solve problems, apply their knowledge to other areas, create rigorous theoretical arguments and interpret the results in plain language. The program prepares the graduates to further their statistical education in a Ph.D. program, to teach at the undergraduate level, or to work in government or business.

II. PROCEDURES USED FOR ASSESSMENT

A. Direct Assessment

Students in the Graduate Statistics Program will be able to produce and judge the validity of rigorous theoretical arguments. i) Four students take the three-part comprehensive exams Part I on Statistics Theory I and II(STT6610/6620), Parts II and III on Statistical Methods I and II(STT 6660/6670). Students who fail in any part may retake them in the following semester. The first two parts of the exams are close-book and each is 3-hour long, and the third part is a one-week take-home project. The students will demonstrate their learning outcomes in probability concepts, sampling distribution, statistical inferences, theory of linear models, regression, ANOVA, designed experiments, and the use of statistical software. All results are evaluated. ii) Students in STT 7620 (Advanced Topics in Linear Models) will be asked to do a project and their performance will be assessed. There are four students and all project reports will be evaluated. iii) Some significant questions will be assigned to students in STT 7440 (Applied Multivariate Analysis) and the answers will be assessed. There are four students and all answers will be evaluated.

B. Scoring of Student Work

A. Assessment based on the Comprehensive exams The students who complete the required courses for the program STT 6610 (Theory of Statistics I), STT 6620 (Theory of Statistics II), STT 6660 (Statistical Methods I), or STT 6670 (Statistical Methods II) should take three parts of the comprehensive written exams. Part i) is for STT 6610/6620, Part ii) is for STT 6660/6670, and the two parts are 3-hour in-class exams, and Part iii) focuses on data analysis using statistical software and is a one-week take-home project. This year Parts i) and ii) were given on 09/10/2021 and 09/17/2021 at 150MM, and the project reports for Part iii) were collected on 09/24/2021. The core topics covered by these exams are listed at
https//science-math.wright.edu/mathematics-and-statistics/applied-statistics-graduate-student-handbook
The exam outcomes were evaluated by a rating scale (0-100) using blind grading, i.e., the student names were not included in their exam papers. The grades were given by the faculty members who prepared the exams and then discussed by the statistical comprehensive exam committee. All four students who took the exams were selected. Their passing rate of the exams was 75% and the students will retake the failing parts of the exams in Spring 2022. B. Assessment based on STT 7620 Advanced Topics in Linear Models The STT 7620 course is an advanced statistical course which focuses on generalized linear models, mixed effects models, and other related topics. The course is a mix of theory and application. In particular, the following concepts will be covered binary response, variations of logistic regression, count regression, multinomial data, generalized linear models, random effects, linear mixed models, and nonparametric regression models. For each regression model, students' learning outcomes are listed below 1. Formulate regression problems and fit models using statistical software R 2. Interpret the R output for regression analysis 3. Perform statistical inference to evaluate the goodness of the regression models 4. Apply important statistical concepts and principles to draw conclusions These learning outcomes are evaluated based on the key questions in the final project of the course. The project includes four parts. In each one, the students are asked to perform an appropriate regression model to analyze a real data problem using R and then answer the questions based on the R output. The project problems were emailed to the students at 5:00 pm on Monday, Dec 6, 2021, and the solutions were collected by the end of the next day. There are four students in the course and all the project reports are evaluated for the assessment. In the project, each of the four parts corresponds to one type of regression model and involves the questions related to the above learning outcomes. In particular, we focus on the logistic regression in the first part, the count regression (the binomial model, the Poisson model, the negative binomial model) in the second part, the multinomial regression in the third part, and the linear mixed model in the last part. For each part, we will select some key problems to evaluate the learning outcomes. One problem may include several questions which correspond to multiple learning outcomes. To evaluate the first learning outcome, we check students' answers and R codes for the questions in the following problems of the project. For each question, the proportion of the students who get the right answers is provided in the following table. Problem 1(a) 2(a) 2(b) 2(f) 3(a) 3(c) 4(a) Prop. 3/4 2/4 4/4 3/4 4/4 4/4 3/4 For Problem

2(a), students need to fit a binomial GLM to fit the count data. Two students did not include the total count variable in the regression model, which leads to the wrong formula. Overall, the students understand how to formulate the regression model and then use the right way to fit the model in R. For the second learning outcome, some key questions from Problem 1(a), 2(b), 2(f), 3(a), and 4(a) are used for the assessment. In these questions, students need to interpret the coefficients in four types of regression models. The performance of the students can be seen in the table below. Problem 1(a) 2(b) 2(f) 3(a) 4(a) Prop. 3/4 3/4 4/4 3/4 1/4 The first four questions are related to the interpretation of the coefficients in generalized linear models. Most of the students got the right answers to these four questions. For 4(a), students are asked to compare the child weights of a 15- and a 25-year-old mother in a linear mixed model. This question is related to the interpretation of the coefficient of the "mage" variable. Only one student got the right answer. For the rest three students, one used the wrong coefficient for interpretation and the other two did not provide answers. In summary, students know how to interpret the coefficients in the generalized linear models, but are less familiar with the interpretation in the linear mixed models. The third learning outcome is assessed using Problem 1(b), 1(c), 2(c), 2(d), 3(b), 3(c), 4(b), 4(c). In these problems, students are asked to check the validity of different regression models using appropriate statistical inference methods. The table below displays the results for these questions. Problem 1(b) 1(c) 2(c) 2(d) 3(b) 3(c) 4(b) 4(c) Prop. 2/4 4/4 2/4 4/4 3/4 3/4 4/4 4/4 In 1(b), an LRT should be used for evaluating the significance of the whole model. Two students used the individual tests and then got the wrong conclusion. In 2(c), two students got the wrong p-value for the Pearson chi-squared test. In general, students have good performance in most of the questions for this learning outcome. In the last learning outcome, students are expected to use important statistical concepts and principles to compare different regression models and then draw conclusions. We select Problems 1(d), 2(f), 3(d), and 4(e) for assessment and present the performance in the following table. Problem 1(d) 2(f) 3(d) 4(e) Prop. 4/4 3/4 3/4 1/4 Students have good performance in the first three questions. In 4(e), students are asked to compare two linear mixed models. A confidence interval could be constructed to check the significance of the random slope in the full model. Most of the students had no clue about this question. Overall, students are more familiar with the model comparison in the generalized linear models than that in the linear mixed models. According to the above analysis, students are capable of fitting an appropriate regression model based on the goal of a study, and interpreting and making inferences for the fitted regression models after taking the course. In this course, the generalized linear model is a main topic, which takes at least 75% of the time in the whole semester. Consequently, students are expected to be more familiar with the generalized linear models than the linear mixed models. In conclusion, these four learning outcomes are generally achieved. C. Assessment based on STT 7440 Some significant questions were assigned to students in STT 7440 (Applied Multivariate Analysis) to evaluate the application of theoretical knowledge and programming to conduct data analysis. There were four students in the class and all answers were assessed based on a series of questions to analyze a specific data set (AutoData in R package ISLR which includes 392 observations and 9 variables). The topics of these questions are 1. Data exploration and statistical visualization. 2. Multivariate linear regression versus multiple linear regression. 3. Sample

statistics and inference in mean vectors. 4. Principal component analysis and principal component regression. 5. Factor Analysis The above five topics (1 - 5) correspond respectively to Questions 3, 4, and 5 in Homework 6 (Due on Nov 23, 2021), and Question 3 and 4 in Homework 7 (Due on Nov 30, 2021). Each question was blindly graded out of 2 points in total, with 2 denoting "all correct" and 0 denoting all incorrect. The four students' performances are Question 3, HW 6 (Scores 2, 2, 2, 2) (Mean 2) Question 4, HW 6 (Scores 2, 2, 2, 2) (Mean 2) Question 5, HW 6 (Scores 1, 2, 2, 2) (Mean 1.75) Question 6, HW 7 (Scores 2, 1.5, 1.5, 2) (Mean 1.75) Question 7, HW 7 (Scores 2, 1, 1.5, 1.5) (Mean 1.5) In summary, these five learning outcomes are achieved in general

C. Indirect Assessment

We have started to conduct exit interviews with graduates and collect feedback from alumni.

III. ASSESSMENT RESULTS/INFORMATION:

We found that A. the students pass 75% of the comprehensive exams B. the proportion of the correct answers in the project of STT 7620 is about 77% C. the average score for 5 homework quesitons in STT 7440 is 90%

According to the assessment plan we expect 75% of our graduates to meet the learning outcome. This is achieved in the three assessments that we have conducted this time.

[Analysis]

IV. ACTIONS TO IMPROVE STUDENT LEARNING

After each comprehensive exam, the graduate statistical faculty members discussed each student's exam paper. From the discussion, we find that the students need more training in statistical theory, statistical programming, and interpretation of statistical results.

V. SUPPORTING DOCUMENTS

Additional documentation, when provided, is stored in the internal Academic Program Assessment of Student Learning SharePoint site.