SOA Research INSTITUTE

Actuarial Innovation and Technology

# Inspiring Actuarial Education through Learning Communities and Research Experiences

October | 2022

# Inspiring Actuarial Education through Learning Communities and Research Experiences

**AUTHORS** Jeffrey Beckley, FSA

Mark Daniel Ward, PhD

Jianxi Su, FSA, PhD

Lingsong Zhang, PhD

**SPONSOR** Society of Actuaries Research Institute

# CONTENTS


**Executive Summary** ........ **4**

**Section 1: Introduction** ........ **5**

**Section 2: Description of the learning community model** ........ **6**

- 2.1 Academic coursework ........ 6
- 2.2 Residential life ........ 6
- 2.3 Professional Development ........ 7
- 2.4 Faculty-mentored research ........ 7

**Section 3: Implementation** ........ **7**

- 3.1 Student Recruitment ........ 8
- 3.2 Courses provided ........ 8
- 3.3 Seminar provided ........ 10
- 3.4 Research provided ........ 12
- 3.5 Faculty-student interactions ........ 12

**Section 4: Impacts of COVID pandemic: From disruption to recovery** ........ **13**

**Section 5: Lessons learned** ........ **13**

- 5.1 First year versus second year students ........ 13
- 5.2 Working in Groups ........ 14
- 5.3 Use of R ........ 14
- 5.4 Data Cleaning ........ 14
- 5.5 Flipped Classroom ........ 15
- 5.6 Research ........ 15
- 5.6 Corporate Partners ........ 15

**Section 7: Acknowledgments** ........ **17**

**Appendix A: Learning Objectives for STAT 490, Data Validation for Actuarial Science** ........ **18**

**Appendix B: Learning Objectives for STAT 490, Predictive Modeling in Actuarial Science** ........ **18**

**Appendix C: Technical details of the mortality modeling project** ........ **19**

- C.1 Data Cleaning ........ 19
- C.2 The Cox porportional Hazard model ........ 20
- C.3 Model Estimation and interpretation of the results ........ 21
- C.4 Confidence intervals for the death count estimates ........ 22

**References** ........ **24**

**About The Society of Actuaries Research Institute** ........ **25**

# Inspiring Actuarial Education through Learning Communities and Research Experiences

## Executive Summary

This project pertained to the development of a new Actuarial Science Learning Community model, involving four types of students' engagement: academic coursework, residential life, research, and professional development. The model provided students an immersive exposure to Data Science from the Actuarial perspective, which will result in Actuarial students greatly expanding their skill set and future job opportunities, at the intersection of Actuarial Science and Data Science.

After implementing the proposed model at Purdue University for three years, we learned

- Actuarial students were enthusiastic about participating in the learning communities.
- The integration between Actuarial Science and Data Science drew substantial interest from students.
- Academic-industrial collaborative research was suitable for second-year students given that a proper level of faculty supervision was provided.
- Lessons about how to better
  - Design study groups in order to encourage all students' participations;
  - Use R in classes by setting up a proper expectation on the students' learning pace and providing sufficient technical supports;
  - Implement a flipped classroom and when it may not be suitable for highly technical courses;
  - Facilitate students' research via advanced planning and flexible schedules.

## Section 1: Introduction

*Learning Communities* (LC's) are organized communities formed with a group of students having common academic interests or majors. Typically, students within a LC take classes together and house together. The LC's model has been growing in importance and popularity among higher education scholars over recent years.

A vast number of empirical studies have indicated that the adoption of LC's can effectively facilitate students' academic success (Ward, 2017; Zhao and Kuh, 2004), retention (Ward, 2015; Pomerantz and Norleen, 2000), and graduation rate (Hill and Woodward, 2013). Furthermore, the LC approach has shown to be strongly correlated with positive students' behaviors such as promoting openness to diversity, social tolerance, and personal and interpersonal development (Zhao and Kuh, 2004). Purdue University, as a pioneer in developing innovative education models, hosts hundreds of LC's on campus each year.

Despite the existence of a great variety of LC's, we are not aware of any models that are specializing in Actuarial Science education. The unique landscape in Actuarial education presents the needs for developing Actuarial Science specific LC's that may be different substantially from many of the existing models. In the first Mid-West Actuarial Students Conference hosted by the Purdue Actuarial Science program in 2013, we surveyed students from 18 universities about ``What are the things you find most useful from your school years?'' The three elements that were most attractive to the students was credential exams preparation, internship/job opportunities, and the development of professional skills such as communication, leadership and networking. Thereby, we believe a successful Actuarial training model should encompass the following components:

1. In-depth coursework for covering the necessary economics, mathematics and statistics concepts.

2. Appropriate interplay between the direct applications of theoretical contents in an idealized business environment (e.g., solving questions in the SOA exams) and the ``messier'' applications faced in the real world professional practice.

3. Opportunities for professional development with an emphasis on verbal and written communication and working in teams.

Integrating the three above components into the Actuarial education curriculum will benefit students greatly when they are seeking internships, future research experiences, graduate school, and permanent employment.

With the ambitious goal of strengthening the education system for the next generation of actuaries and broadening the role of actuaries in nowadays society, we aimed to develop an Actuarial Science Learning Community (AS-LC) model that was tailor-made for students who were majoring in Actuarial Science or interested in the Actuarial field. Though the AS-LC model was implemented at Purdue University, in this report we thoroughly document the educational objectives, the approach utilized, and the lessons learned during our implementation experience. In so doing, we hope that other AS-LC's can get started on other campuses, and thus extend the reach of the AS-LC model beyond Purdue.

The rest of this report will be organized as follows. The learning objectives and the curriculum of the proposed AS-LC are described in Section 2. The implementations of the AS-LC in 2019-2020, 2020-2021, 2021-2022 academic years (AY's) are elucidated in Section 3. Admittedly, the COVID pandemic had posed unprecedented challenges to both learners and educators. Section 4 discusses how the pandemic impacted the implementation of AS-LC. Finally, Section 5 concludes the report by summarizing the lessons learned during the 3 years of implementation of the proposed model.

## Section 2: Description of the learning community model

The proposed AS-LC model involved four types of students' engagement: academic coursework, residential life, professional development, and research. The theme of the AS-LC was Actuarial and Data Sciences, which is consistent with the present trend of Actuarial education in which much attention is paid toward the development of predictive analytics skill set among candidates. The proposed AS-LC provided students an immersive exposure to Data Science from an Actuarial perspective which resulted in Actuarial students expanding their skill set and future job opportunities into the intersection of Actuarial Science and Data Science.

The targeted students were mainly sophomores who would benefited from the experience when seeking internships. Moreover, the sophomore year is often called the “sophomore slump” because of the corresponding lower performance of students as sophomores. This happens exactly when students are starting to take the foundational courses in the major which are harder than the freshman courses. However, support and guidance provided by student peers and faculty may be limited. Students who experienced failures of SOA exams may drop out of the major, which results in the Actuarial profession suffering from student retention. By living and working together, the AS-LC model helped the students feel more confident, comfortable, and valuable. This in turn further enhanced the sophomore transitions and retention of Actuarial students.

During the first-year implementation of the AS-LC, we allowed first year students into the AS-LC. A lesson learned was that we should have limited the class to just the second year students (or occasionally, the third year students) and not allowed the first year students in the class. Subsequently, we have not allowed the first year students in the AS-LC. This is discussed further below in Section 5.

### 2.1 ACADEMIC COURSEWORK

Students in the AS-LC learned together in at least two courses that were relevant to Actuarial Science and Data Science. Both academia and the marketplace have recognized the value of teamwork. Despite recent innovations in the Actuarial education, it remains the case that most students experience universities as isolated learners whose learning is disconnected from that of others. The students in the AS-LC share common experiences, which helped them to build a sense of community. For instance, by taking courses as a cohort, the students naturally studied together and helped each other to reinforce the understanding of materials and promote cognitive development. Via study groups, the AS-LC also resulted in a better model to help the students to pass the SOA exams. Moreover, learning as a community inevitably makes the academic experience more enjoyable. This part of the proposed AS-LC model facilitated the necessary component 1 outlined in Section 1.

### 2.2 RESIDENTIAL LIFE

Students in the AS-LC lived on the same residence hall floor, with activities to enhance the student engagement. These activities included weekly dinners with LC participants, and social gatherings with Actuarial Science faculty, campus leaders, and practicing actuaries and data scientists. Students in the AS-LC got to know each other quickly and fairly intimately in a way that is part of their academic experience. A great deal of learning took place in the residence halls as the students had more opportunities to discuss complementary ideas outside the classroom. Their own knowledge was enhanced when other voices were part of the learning experience. The residence hall became the place where the students spent time solving problems in Actuarial Science, working on their mathematics, and honing their computational skills. The social activities created opportunities for the students to develop networking skills in a professional manner. This part of the proposed AS-LC facilitated the necessary components 1 and 3.

### 2.3 PROFESSIONAL DEVELOPMENT

The students participated in a series of professional development seminars to broaden their skill set. The seminar series aimed to provide an introduction of some fundamental Data Science topics that are highly relevant to Actuarial practice. Through the seminar series, the students also had opportunities to meet with alumni, actuaries, statisticians, data analysts, faculty, graduate students, and staff members from professions that were not limited to Actuarial Science. The seminars were dedicated to opening students' eyes to the many ways that analytics techniques can be used in the workforce. The seminar series also helped prepare students for post-college opportunities, in graduate school, and in different kinds of employment. This part of the proposed AS-LC facilitated the necessary components 1 and 2.

### 2.4 FACULTY-MENTORED RESEARCH

The students had opportunities to engage in faculty-mentored, team-oriented, Actuarial research. All the students were encouraged to participate in these research opportunities. Research activities allowed students to connect with faculty mentors, to learn from peers, and to get a sense of the type of work involved in the Actuarial workforce. The research component in AS-LC created an immersive environment for students to touch on dirty data, deal with the real word applications of textbook knowledge, and handle practical constraints and limitations. This should result in the development of critical thinking skills for the students. Through participation in cooperative research experiences, the students were mutually dependent on one another and required to share responsibilities. During the process, the students naturally learned to address intra-team conflict, manage individual and team accountability, and provide constructive feedback to team members. Student were asked to present the research project to their peers and encouraged to present at university conferences or regional conferences. The students were also encouraged to submit overviews of their research projects in academic newsletters. In so doing, the students learned how to succinctly explain the motivation and impact of their research projects to a broad audience. This helped to improve the students' communication and presentation skills. This part of the proposed AS-LC facilitated the necessary components 2 and 3.

Collectively, the diversity of experiences and the sense of community that the students gained in the AS-LC should be essential aspects of the students' success.

## Section 3: Implementation

The AS-LC at Purdue was kicked off in Fall 2019 as one of the twenty cohorts in a campus-wise learning community initiative---the Data Mine, which taught Data Science to participating students from all majors, regardless of their previous experience. Students in the Data Mine were not necessarily pursuing a major in Statistics or Actuarial Science. Instead, they were united by a strong interest in how Statistics and Data Science would be used in their future careers. The Actuarial Science cohort was the largest cohort in the Data Mine, and Actuarial Science students had the priorities to enroll into the cohort.

Implementing the proposed AS-LC through the participation of Data Mine allowed us to leverage university recourses that are cost-effective and sustainable. For instance, it became convenient for us to arrange students to live in the same residential hall and take part in campus-wise professional activities. It also allowed students to interact with peers from other relevant disciplines such as agriculture, health and human sciences, computer science, etc., so that the students naturally learned the values of diversity.

## 3.1 STUDENT RECRUITMENT

The recruitment of students to the AS-LC started in February of each year. A web site[1] was created to provide information for perspective students. The information includes

- General learning objectives
- Eligibility
- Residential information
- Coursework curriculum
- Events and activities

In addition, the AS-LC was advertised broadly via a public email list for all Actuarial students. Faculty also introduced the AS-LC opportunity to students in their classes.

Further, information sessions (callouts) were held bi-weekly where Actuarial Science faculty would present the AS-LC to students, followed with a Q&A component. Information sessions were usually held at different time of a weekday in order to facilitate the different course schedules of students. There were about ten to twenty students and one or two faculty who attended each callout.

The application window started at the same time and lasted till the end of April. Applicants were reviewed concurrently by Actuarial faculty so that students could make plans for their coursework and residential arrangement for the upcoming academic year as early as possible. Students needed three calculus classes to be ready for the Probability class involved in the AS-LC curriculum. However, the applicants were not assumed to have backgrounds or prerequisites in actuarial mathematics, probability, statistics, or programming. The criteria for selection was holistic: inclusion of females, minorities, and persons with disabilities; interest in Actuarial Science or Data Science; interest in an applied field. Other qualitative criteria included leadership, teamwork, motivation/likelihood to make a research contribution. Since the students were selected early in their college career, and they might have struggled during the transition into college, we de-emphasized their first-semester GPA. Table 1 summaries the enrollment of the AS-LC in the past three academic years.

Table 1
NUMBER OF ENROLLMENTS OF THE AS-LC IN THE PAST THREE ACADEMIC YEARS.

| | AY 2019-2020 | AY 2020-2021 | AY 2021-2022 |
|---|---|---|---|
| The number of enrolled students | 64 | 16 | 28 |

## 3.2 COURSES PROVIDED

We scheduled at least three Actuarial Science and/or Data Science courses in blocks, so students in the AS-LC would take courses together as a cohort while still having plenty of time for individual electives. Table 2 outlines these courses, followed by more detailed elaborations of the course objectives and designs.

[1] https://www.datamine.purdue.edu/actuarial-science/

Table 2
COURSES TAKEN BY STUDENTS IN AS-LC AS A COHORT.

| Semester | Course Code | Credit Hours | Course Title |
|---|---|---|---|
| Fall | STAT 416 | 3 | Probability |
| Spring | STAT 417 | 3 | Statistical Theory |
| | STAT 473 | 3 | Introduction to Arbitrage-Free Pricing of Financial Derivatives |

- *STAT 416 Probability*. The course introduces the basic mathematical model of randomness and examines the fundamental notions of independence and conditional probability. It covers the mathematics used to calculate probabilities and expectations and discusses how random variables can be used to pose and answer interesting problems arising in nature. The course is required in most science programs at Purdue University including Mathematics, Statistics, Computer Science, but it is also offered to students in e.g., Psychology and Engineering. For Actuarial students, this course is suitable as preparation for the Exam P of the Society of Actuaries (SOA).

  In the past, Actuarial Science students had taken STAT 416 in different general sections where there were other STEM or non-STEM major students enrolled. Given the class size and the diversity in the mathematical backgrounds of participants, the task to meeting learning objectives was somewhat challenging. Tedious proofs were avoided when it came to the non-STEM students, while Actuarial Science students desired a higher level of mathematics training.

  Owing to the initiation of AS-LC, we introduced a new section of STAT 416 designated for Actuarial Science students. This allowed us to improve STAT 416 to better align with the learning objectives of SOA Exam P. Insurance related examples and homework questions were given in order to trigger students' imagination to comprehend the potential real-life applications of the probability theories learned in class.

- *STAT 417 Statistical Theory*. This course is an introduction to the mathematical theory of statistical inference, emphasizing inference for standard parametric families of distributions. Specifically, the course covers properties of estimators, Bayes and maximum likelihood estimation, sufficient statistics, properties of test of hypotheses, distribution theory for common statistics based on normal distributions. For Actuarial Science students, this course fulfills the SOA VEE requirement on Mathematical Statistics.

  Similar to STAT 416, this course was open for students from different majors including Actuarial Science, Statistics, Mathematics, Computer Science, etc. With the development of AS-LC, an Actuarial Science focused section of STAT 417 was created, and a Statistics professor who has research experience in Actuarial Science taught this section. This arrangement provided more in-depth statistics exposures to Actuarial Science students with connections to insurance applications for motivating students' learning.

- *STAT 473 Introduction to Arbitrage-Free Pricing of Financial Derivatives*. This course covers a variety of fundamental results related to arbitrage-free pricing and hedging of derivatives in the Binomial and the Black-Scholes Market Models. This course stresses financial intuition by extensively exploiting the concepts of hedging and arbitrage. The course covers the financial derivatives portion of Exam IFM of the SOA. We will continue to teach this material even though the SOA will no longer examine the material after 2022. This course was taught by Actuarial faculty with extensive industry experience.

  The placement of this course in the AS-LC has resulted in many students taking the IFM exam following their second year of college. As a result, many students have three exams by the start of their third year.

### 3.3 SEMINAR PROVIDED

The students in the AS-LC could take five Data Science seminar courses. These courses were elective so that the students could still have plenty of flexibilities to plan for their other course schedules. Generally, the courses were only available to students in the AS-LC. When space was available, we opened the seminar courses to all Actuarial Science students. These seminar courses were offered once a year. Their information is summarized Table 3.

Table 3

DATA SCIENCE SEMINAR COURSES AVAILABLE TO STUDENTS IN THE AS-LC.

| Semester | Course Code | Credit Hours | Course Title |
|---|---|---|---|
| Fall | STAT 190 | 1 | The Data Mine (R focused) |
| | STAT 490 | 1 | Data Validation for Actuarial Science |
| Spring | STAT 190 | 1 | The Data Mine (Python focused) |
| | PHIL 208 | 1 | Ethics in Data Science |
| | STAT 490 | 1 | Predictive Modeling in Actuarial Science |

- *<u>STAT 190 The Data Mine.</u>* This one credit hour course was offered in both the fall and spring semesters, with the fall semester section focused on R implementation while the spring semester section focused on Python implementation. The students were encouraged to take both sections to learn to use the two computation platforms that have gaining in both popularity and importance in the Actuarial practice.

  By taking the course, the students gained hands-on experience with computational tools for representing, extracting, manipulating, interpreting, transforming, and visualizing data, especially big data sets, and in effectively communicating insights about data. Topics include: the R environment, Python, visualizing data, UNIX, bash, regular expressions, SQL, XML and scraping data from the internet.

  The course was project oriented, and the students worked in teams. The students were expected to spend about 3 hours per week doing work for the class. There were thirteen programing projects available to the students. Generally, every week from the very beginning of the semester, a new programming project was released to the students on a Thursday, and it was due 8 days later on the Friday.

  The instructor typed code directly, using the computer’s projector, so he could adapt to questions. The students saw how the instructor recovered from errors. This cannot be learned from a book. After each lecture, the instructor uploaded all of the day’s code, including comments he added after class, to supplement the discussion.

- *<u>STAT 490 Data Validation for Actuarial Science.</u>* In the Fall semester, we offered a class entitled “Data Science for Actuarial Science”. The class was based around a data set provided by insurance companies. For 2019, the students worked in teams to complete a mortality experience study. For 2020 and 2021, the students worked through an exercise of calculating loss reserves and setting premiums based on the data set.

  From a Data Science standpoint, this was primarily an exercise in learning to manage and validate large data sets. The students also learned techniques for dealing with incomplete or inaccurate data. From an Actuarial Science standpoint, students learned and applied various techniques to real data sets. Students worked in teams to complete the work. Students also worked with various software tools such as Excel and R.

  More details about the learning objectives of this class are reported in Appendix A.

During the class process, the students worked in teams of 5. Teams were assigned and the students did not have inputs into the team assignment. The teams were diverse with a mixture of genders and domestic versus international students. Additionally, to the best of the instructor's ability, students with strong technical skills were dispersed throughout the groups.

Each team was assigned a teaching assistant (TA) who were undergraduate junior or senior Actuarial Science students who were knowledgeable in R. Each TA was responsible for two teams.

Each team met in the classroom once a week for one hour and then were also required to meet for one hour outside the classroom as a group with their assigned TA.

Besides the undergraduate TAs, we also utilized two graduate students as lead TAs. One was from the Mathematics Department and one was from the Statistics Department. These graduate TAs were chosen primarily for their knowledge of R and interest in Data Science.

The class was intended to be a "flipped" classroom. While a flipped classroom means different things to different people, the intent here was that students would read material, watch videos, and so on between classes. During the class, additional context would be provided, and questions would be answered for 15 to 20 minutes and then the students could work on the next step of the project during the rest of class times. It turned out that this did not work well as expected and will be discussed in Section 5.

- *PHIL 208 Ethics in Data Science.* In addition to the analytic Data Science skills, the ethics involving in Data Science are important for Actuarial students to understand and consider now and during their future careers. We thus seek out experts in the fields of Philosophy and Political Science to assist with this education.

  In this seminar course, the students explored the ethical issues in big data and contemporary digital technologies. This course addressed these issues by providing (1) a conceptual framework for ethical reasoning, particularly in the professional setting, and (2) a procedure for case-study analysis designed to give students practice in employing this conceptual framework to real world cases in big data ethics.

- *STAT 490 Predictive Modeling in Actuarial Science.* Owing to the Data Science revolution, the Actuarial Science discipline has changed dramatically in the past ten years. This course aimed to provide basic introductions of state-of-the-art statistical learning techniques to Actuarial Science students. The course also covered some important components in the two SOA credential exams---Statistics for Risk Modeling and Predictive Analytics, introduced in 2018. At a high level, the course introduced a handful of predictive analytics techniques that are useful in modeling insurance losses. The topics taught in this course included maximum likelihood estimation, numerical optimization, generalized linear models, kernel density estimation, and decision trees for regression analysis. These fundamental Data Science skills, which have not traditionally been a major component of university curriculum in Actuarial Science, are increasingly valued by employers in the insurance industry. More details of the learning objectives of this class are reported in Appendix B.

  The students in this course met twice every week during the first or second half of the Spring semester. This course was project oriented. Programing projects were offered to the students to practice on the data analytic skillset learned in class. The projects were based on real life data obtained from the Wisconsin Local Government Property Insurance Fund and an Australian insurance company. These projects provided an immersive environment for students to touch on real life data, deal with the messier applications of textbook knowledge, and handle practical constraints and limitations of statistics models.

### 3.4 RESEARCH PROVIDED

The students in the AS-LC were encouraged to take part in faculty-mentored research. At the beginning of each academic year, we emailed the research opportunities to the students in the AS-LC. Meanwhile, faculty also actively reached out to individual students and encouraged them to apply for the research opportunities. The students needed to submit their CV, transcript, and a statement highlighting their research motivation to apply for the research positions. We received about 10-20 applications each year. After an initial round of screening, we shortlisted 6 students for interviews. During the interviews, the students were asked questions such as why they are interested in research, how they resolve team conflicts, how they address challenges in class or at work. We selected 3-4 students for each research project. The selection criteria were mainly based on the motivation, persistence, and teamwork skill of the students.

The research project we provided to the students involved a mortality and lapse study using data provided by a large life insurance company. Specifically, the data set contained the death dates and termination dates of more than 250,000 deidentified policyholders of the company's whole life insurance and term life insurance products. Each individual record also contained covariate information such as gender, issue age, medical examination rating, face amount, and reinsurance amount. The purpose of the research was to apply state-of-the-art statistical methods to classify the mortality risk associated with policyholders having different covariate information, and predict the annual death counts, annual termination counts and total annual insurance payments in the next few years.

More technical details of the research project are provided in Appendix C.

One challenge of the research was the computation burdens involved in the calculation of prediction intervals. Specifically, bootstrap sampling technique was needed to estimate the empirical distributions of the estimators of death counts and termination counts. Due to the complexity of the simulation procedure and the large sample size, constructing the bootstrap samples was computationally onerous. The students learned how to use parallel computing on a grid to decompose a large scale simulation into smaller tasks. This allowed the students to have first-hand experience running jobs on a large, parallel grid. To implement the parallel computing, Purdue has a handful of world-class computational clusters, free to any Purdue user including students.

The students met with the lead Actuarial faculty once a week to provide updates and seek feedback or technical support. At least four times per semester, the students would meet with the actuaries from the industry partner to present their findings. In so doing, the students naturally learned the communication needed in data analysis: understanding the context; visualization; disseminating results; focusing on the outcomes and impact. An important part of writing about data and visualizing data is to explain the motivation and ideas. The students learned to treat communication as an iterative process, following the paradigm of not settling on the first attempt to summarize the results visually or in writing. The students had opportunities to interact with practicing actuaries, which may turn out to open up future employment opportunities. The company also benefited from having access to the most self-motivated students for future recruitment.

The students in the research worked as a team to apply the data analytics knowledge they learned in class to solve business problems. The students learned to assign and take on distinct roles, address intra-team conflict, manage individual and team accountability, provide constructive feedback to team members, and manage their time.

### 3.5 FACULTY-STUDENT INTERACTIONS

There were plenty of activities planned to enhance the interactions between faculty and students. Actuarial faculty visited the residential hall once a week and held office hours there. Faculty answered questions not limited to the course works, but other more general questions related to student life, job search, professional development, etc. Making faculty more accessible encouraged students to reach out when they face any challenges in life and study.

The Faculty Fellows program allows Actuarial faculty to have meals together with the students in the residential dining hall on a fixed day every week. External visitors of the Actuarial Science program (e.g., alumni, Actuarial professors from other universities, company's recruiting teams, etc.) were also invited to attend the meals so that the students have opportunities to interact with them.

Some faculty involved in the AS-LC hosted the students for dinner with their families in their homes. These dinners showed the students that they are valued and respected. The dinners showed students how faculty can balance being a researcher and teacher, with being a mentor and parent.

## Section 4: Impacts of COVID pandemic: From disruption to recovery

The outbreak of the COVID pandemic upended classrooms and campuses across the country. Consequently, profound challenges were brought into the implementation of Data Mine and AS-LC. At the start of the pandemic back in March, 2020, Purdue University moved most of teaching online. Normal operation of the university did not resume until the fall semester of 2021. However, there were still a considerable number of students concerned about living in dormitories due to the relatively crowded, shared living spaces. For this reason, we changed the residential requirement to allow the students to choose to live in dormitories or off-campus. However, the pandemic still impacted the enrollment of AS-LC. We saw a substantially smaller number of enrolled students in 2020—2021 AY and 2021—2022 AY.

The online teaching and social distancing requirements limited the interactions between students and faculty. Most social events and presentations were cancelled or limited to online. As a remedy, we gave class participation and event attendance bonus credits to incentivize students' engagement in learning. We also required students to meet with faculty either virtually, or in person when the university policy allowed, to follow up with the students' learning progress.

As we approach the end of the pandemic (hopefully), we expect the in-person components of the Data Mine and AS-LC will be fully resumed starting from the fall semester of 2022. Aside from the recovery from the pandemic, it is more important for us to use the experience to better prepare for future crises that are not limited to another pandemic but a boarder range of events that may cause education disruptions, such as natural or man-made catastrophes. The experience helped us learn how to switch between face-to-face and remote learning if needed in the future. The experience also forced faculty to stay updated with the available IT technologies to facilitate the remote teaching more efficiently.

The pandemic has encouraged us to consider a more flexible curriculum for the AS-LC to make the education model more resilient to future catastrophic shocks occurring to our education system. Another ongoing effort is how to make use to the online learning materials developed during the COVID pandemic to enlarge the scope of AS-LC in order to benefit more learners at all levels.

## Section 5: Lessons learned

This section aims to summarize the lessons we have learned during the implementation of the proposed AS-LC, so that programs interested in starting their own learning communities can learn from what we have learned.

### 5.1 FIRST YEAR VERSUS SECOND YEAR STUDENTS

A lesson learned was that we should have limited the class to just second year students and not allowed the first year students in the class.

The difference between the expectations of the first year students and the second year students was immense. The first year students thought that the class should be at the level of a high school class. They were not prepared or expecting the level of work that we expected from the students. Many of the first year students wanted to drop the classes after a couple of weeks because they did not want to work as hard in the classes as we expected them to do. On the other hand, the second year students felt that our expectations were reasonable and in line with other college classes.

Of course, the above statement was not true universally for either first year or second year students but in general it was true. This resulted in the first year students generally not having a positive experience in the classes.

The second yet perhaps more important reason that we should not have allowed first year students in the AS-LC was that these students were then not interested in participating the AS-LC during their second year. We also believe that because some of them did not have a positive experience in the classes, they discouraged other classmates from joining the AS-LC. As a result, we had a much smaller group of students in the AS-LC during academic year 2020-2021 with only 16 students.

### 5.2 WORKING IN GROUPS

There are many benefits for students working in groups. Our experience with students working in groups was with juniors and senior Actuarial Science students. The model had always worked very well.

Here it did not work as well. The weekly assignments were intended to require about 2 to 3 hours of work for each student outside the classroom. However, most students did not want to take the time to learn to code in R. Therefore, they basically relied on the "expert" in their group to do all the work in R. For some assignments, it would take up to 10 hours to complete and test the R codes. When this falls on one student in a five person group week after week, it is clearly unfair to the expert student. More importantly, the other students were not learning the material.

Therefore, a major lesson learned was that a larger portion of each students grade needed to be determined based on their individual work as opposed to the work of the group. Secondly, peer grading and pressure needed to be exerted within the groups to incent the less active students to take a more active role.

### 5.3 USE OF R

Our initial expectation was that students would be able to code sufficiently well in R based on their exposure to R in other classes and in the Data Mine (R focused) course as part of AS-LC. We did not intend to spend much time on teaching coding in R. This assumption turned out to be incorrect. Most of what the students did in other classes was that they were given R code and needed to make a few changes to the R code to complete the project that they had been assigned. Initially, we did not provide any R code with the expectation that students would write R code. However, we quickly realized that this would not work for the majority of students. We adjusted by having the TA's write basic R code which the students could then modify to accomplish the given tasks.

We should point out that this problem was exacerbated by the inclusion of first year students in the class as they had very little exposure to programing in any form.

### 5.4 DATA CLEANING

In the Data Validation for Actuarial Science course, we used real data provided by a large insurance company. The data that we received from the company had already gone through their data cleansing process so we did not find any major errors in the data. In subsequent years, we ask the company to provide a data set prior to cleaning and one after cleaning. It is interesting to see how many of the errors in the dirty data that the students found versus how many they missed.

### 5.5 FLIPPED CLASSROOM

The "flipped classroom" has some drawbacks in the setting of AS-LC. The material (both the Actuarial content and the Data Science content) was too technical for the students to fully grasp it from recorded lectures and they needed to be able to ask questions as the topics were being discussed, not later when the class was being held. Additionally, the difference between first year and second year students was also noticeable with regard to the flipped classroom.

### 5.6 RESEARCH

The students involved in the faculty-mentored research were committed to a full academic year experience. The students were expected to spend 3-5 hours per week on the research. The expected workload and research objectives must be clearly explained to the students at the beginning of the project. Since the students in the AS-LC still took 4 or 5 other regular courses per semester, the amount of research work assigned to the students must be appropriate such that the students can maintain desirable performances in their coursework. Faculty need to be flexible and understandable. During the exam periods, it is acceptable to put a pause on the research for a couple of weeks.

Because the students may do internships in summer, they may not be available to continue the research once the spring semester ended. The research scope and timeline should be planned for about nine months only. The research details including objectives, approaches utilized, all the attempts independent of failures or successes, findings, etc., should be properly documented so that students in the future AS-LC can benefit from the past experience. This also allowed the research team to consider larger scale projects that consist of multiple years of efforts.

We also found that domestic students were less enthusiastic about the research opportunities than international students. This is probably because of the domestic students' future career plans working as practicing actuaries, whereas the job opportunities (including internships and permanent jobs) are substantially less for international students due to the visa issue. A significant portion of international students at Purdue Actuarial Science end up going to graduate schools in Statistics, Financial Mathematics, and Computer Science, and they have been very successful in these other fields.

Nevertheless, we still strongly believe that academic research experiences are valuable for students choosing to become practicing actuaries after graduation. Namely, the mathematical theories students learned in classrooms are unlikely to work perfectly in practice. Through academic research, students naturally see when and how theories fail in practice, what are the limitations of mathematical tools, how to address the problems, etc. Academic research not only provides a hands-on experience for students to gain immersive experience in Actuarial Science and Data Science, but also triggers students' imaginations and encourages students' critical thinking. All these skillsets will essentially benefit the students' future careers. To encourage students to participate in research, we found that faculty actively reaching out to potential students was particularly useful. It was also helpful for the program or industry partners to provide a small scholarship or honorarium to recognize the contributions from the students.

### 5.6 CORPORATE PARTNERS

The Data Mine program at Purdue University also featured a corporate partnership function, through which academic-industrial collaborations were promoted. The Data Mine had 60 corporate partners projects with 47 partners[2]. The partnerships were in many domains, including digital agriculture, manufacturing, aerospace, health care, supply chain, consumer science, computational drug discovery, sports analytics and social media analytics. The corporate partners

[2] An exhausted list of corporate partners projects can be found in https://projects.the-examples-book.com/projects/

program enables the students to work directly with employees of companies or national laboratories on data-driven projects.

Unfortunately, the AS-LC---even though it was one of the largest cohorts in the Data Mine---had not been successful in developing corporate partners with insurance companies. We have tried to reach out to some insurance companies, but only one corporate partnership was found. Concerns that we often heard during the discussions with insurance companies include

- Data confidentiality. Insurance data may contain sensitive information, thus insurance companies are reluctant to share the data with universities for research purpose.
- Project timeline. The timeline of a research project in AS-LC is typically nine-month long so that the students can balance their time spending on research and coursework. However, some insurance companies may want to see more timely results.
- Motivation. Actuaries have the expertise in analyzing the data, and the motivations for them to work with universities are unclear.
- Expectations on the research outcomes. Insurance companies expect the research outcomes to be a work product that can be directly used in their business, while the outcomes from the student research may not meet their expectations.

Nevertheless, we do believe that insurance companies should be benefitted from the partnership via accesses to faculty and student intelligence, university's computing facilities, and knowledge diversity within a university setting. Although it is not yet clear what is the best approach for addressing the concerns outlined above, as we continue to implement the AS-LC model at Purdue, we are enthusiastic about exploring different solutions for expanding the partnership with insurance companies. Hopefully, once we have a core set of projects completed in the next future, they will lead us to others.

On the other hand, Actuarial Science students can be still benefit from partnerships with other entities such as consulting firms, government agencies, or companies from other disciplines such as agriculture, health, engineering, where students can apply their Actuarial and Data Science knowledges to deal with real world problems. These opportunities are available through the Data Mine and some students have taken advantage of these partnerships.

## Section 7: Acknowledgments

The authors' deepest gratitude goes to those without whose efforts this project could not have come to fruition: the volunteers who generously shared their wisdom, insights, advice, guidance, and arm's-length review of this study prior to publication. Any opinions expressed may not reflect their opinions nor those of their employers. Any errors belong to the authors alone.

Project Oversight Group members:

Russell Jay Hendel, Ph.D., ASA

Warren Luckner, FSA, MAAA, CFA

At the Society of Actuaries Research Institute:

R. Dale Hall, FSA, MAAA, CFA, CERA, Managing Director of Research

## Appendix A: Learning Objectives for STAT 490, Data Validation for Actuarial Science

- Understand the application of Data Science to an Actuarial Science Case Study
- Understand data quality issues and be able to identify common data errors
- Apply techniques to data to identify and fix data errors
- Understand data visualization techniques and apply those techniques to the data
- Use R, Excel, and other tools and techniques to analyze the data
- Use various data analytic techniques to the data and identify previously unidentified relationships
- Learn the uses of mortality and lapse studies
- Apply appropriate techniques to measure decrements (death or lapse) and the associated exposure for each cell being evaluated
- Understand how to and be able to write a professional report
- Develop presentation skills for a professional setting
- Understand the Actuarial Standards of Practice applicable to this work

## Appendix B: Learning Objectives for STAT 490, Predictive Modeling in Actuarial Science

- Understand the differences between parametric models and non-parametric models
- Understand various numerical approaches for finding the roots of equations and solving unconstrained/constrained optimization problems
- Apply numerical optimization techniques for solving maximum likelihood estimation problems
- Learn the uses of generalized linear models in insurance applications
- Model validations
- Learn the uses of kernel-density estimation
- Learn the use of tree-based models

# Appendix C: Technical details of the mortality modeling project

The mortality data we obtained from a life insurance company contain various information about each policy regarding the policy number, company, product, series, issue source, issue date, issue age, gender, smoker class, underwriting class, substandard rating, premium mode, specified amount, status, termination entry date, termination effective date, and cause of death.

## C.1 DATA CLEANING

Upon receiving the dataset, we reviewed and cleaned the data. Issues identified were:

- 7 policies had a specified amount of 0 and many policies had an unusually small specified amount (<10).
- There were termination entry dates long before termination effective dates (>30 days) and many missing causes of death.
- Multiple people had causes of death, but their statuses were listed as "Active."
- For a majority of the older records, the issue dates were the first of the month, and some of these older records were still listed as "Active" despite being issued nearly a century ago.
- There were multiple abnormal substandard ratings.
- Not all causes of death matched codes listed in the data notes provided by the partner company.
- Numerous rows existed with termination effective dates, but no termination entry dates.
- There were some entries with a premium mode not defined in the Data Notes spreadsheet.
- Several terminations occurred outside the range of dates for when the study took place.
- We also found some seemingly duplicate policies that only differ in policy number while all other values are the same.
- Two questionable data values arose during the review. First, there were two minors classified as smokers, however they were both 17 years old, so we do not anticipate this being an issue in our analysis.
- Second, there were a few large outliers in addition to the aforementioned specified amounts, with one in particular having a specified amount of $25 million.

Based on the discussion with the actuaries from the company provided the data, the following methods were used to clean the dataset before proceeding with the analysis of the data:

- For the policies with an unusually small specified amount, if this happens to a "modern" policy, it is because it went "reduced paid-up" very early on and was surrendered due to the small amount. For older policies, the data provider thinks those numbers are correct. The only policies that were fixed were the ones that had a specified amount of 0, which were replaced by the correct amounts provided by the company.
- For the policies with termination entry dates long before termination effective dates, the earlier of the two dates was used for both termination dates. Also, policies with termination entry dates during or before 2006 have been removed while the policies with termination entry dates in 2019 have been kept.

- For the policies with missing cause of deaths and/or deaths that contradicted the data codes, we did not correct them since cause of death will not be considered in this mortality study.

- For the policies where there was a cause of death, yet the status was still "Active", we deleted the cause of death and will consider these policies as "Active" since the data provider could not find these policies in their death records.

- Policies with an attained age of greater than 100 have been terminated except for two that are suspended on the company's system. The older policies with issue dates on the first of the month have been unaltered, as this is likely due to the fact that back then, records and calculations were done by hand, so this made the process easier.

- The policies with abnormal substandard ratings have been unaltered. Terminated policies with an empty termination entry date seem to be deaths very late in the study, therefore, the termination effective dates of those policies have been removed and the status has been changed to "Active".

- The two policies that had premium modes of 0 were fixed with the correct premium.

- The policies that seem to be duplicates were determined by the company to not be duplicates but are two policies issued on the same insured on the same date.

- For the two policies where the policy holder was 17 years old at the time of issue and classified as "T" (Tobacco User), the policy holders were reclassified as "C" (Juvenile/Unknown).

- For now, outliers will be kept in the study. Later on, we can perform the study with the outliers removed by replacing the specified amount with the company's retention limit, if necessary.

## C.2 THE COX PORPORTIONAL HAZARD MODEL

The partner company was interested in developing a holistic, transparent model for

- Understanding the impact of different predictor variables (e.g., gender, smoker class, issue age, substandard rating, etc.) has on the survival probability distribution of an individual.

- Estimating the survival probability distributions of different individuals.

- Constructing the confidence intervals for the death count estimates.

To this end, we chose to adopt the Cox proportional hazards (CPH) model to analyze the data. Briefly, the CPH model is a classical statistics technique that can be used to study the association between predictor variables and survival outcomes (e.g., death time, company default time, catastrophe occurrence time). Before presenting the CPH model, let us first set up some standard notations. Let $T$ denote the lifetime random variable. The survival probability function of $T$ is defined as

$$S_T(t) = P(T > t), \text{ for } t > 0,$$

and the probability density function of $T$ can be computed as

$$f_T(t) = \lim_{k \to 0} \frac{P(t \leq T \leq t + k)}{k} = -\frac{d}{dt} S_T(t), \text{ for } t > 0.$$

The hazard function is given by

$$h(t) = \lim_{k \to 0} \frac{P(t \le T \le t + k \mid T > t)}{k} = \frac{f_T(t)}{S_T(t)}, \text{ for } t > 0.$$

Note that the following equation holds for any lifetime random variable $T$ :

$$S_T(t) = exp\{-\int_0^t h(s)\, ds\}, \text{ for } t > 0.$$

The equation above reveals that the study of the probability distribution of the lifetime random variable $T$ is equivalent to that of the hazard function.

For the purpose of mortality risk stratification, it is of central importance for us to incorporate extra predictor variables (e.g., gender, smoker class, issue age) into the analysis. To this end, the CPH model is natural to evoke. Specifically, the CPH model caters the mortality heterogeneity by imposing a regression structure on the hazard function:

$$h(t|x) = h_0(t)\, g(x), \text{ for t} > 0,$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_d)$ contains the $d$ predictor variables, $h_0$ is the baseline mortality function, and

$$g(x) = exp\,(\beta_1 x_1 + \ldots + \beta_d x_d)$$

is the risk ratio that indicates the relative mortality risk compared to the baseline mortality $h_0$. The function $g(x)$ contains the effect parameters that describe how mortality changes with respect to the covariates of interest. The mortality responds exponentially and each unit change in a covariate has a multiplicative effect on mortality. A risk ratio greater than 1 indicates an increase in mortality risk, while a ratio less than 1 indicates a decrease in mortality risk.

## C.3 MODEL ESTIMATION AND INTERPRETATION OF THE RESULTS

In order to model the mortality data and thus construct the associated mortality risk stratification, the statistical task is to estimate the baseline hazard rate function $h_0$ and the regression coefficients $\beta_1, .., \beta_d$. We used the **survival** package in R to handle the estimation problem via treating the active policyholders and lapses/surrenders as censored data points.

The goodness-of-fit of the CPH model was validated based on the following three methods: concordance index, Somers' delta, and comparisons between the fitted survival curves and Kaplan-Meier curves. The fitted CPH model yields the concordance index at about 70% and Somers' delta at about 50%. Meanwhile, the fitted survival curves closely track the Kaplan-Meier curves estimated from the mortality data.

The estimated baseline hazard rate function $\hat{h}_0$ and regression coefficients $\hat{\beta}_1, \ldots, \hat{\beta}_d$ can be immediately used to construct survival probability estimates. Specifically, for a policyholder with covariate vector $\mathbf{x}$, the probability that the policyholder age $y > 0$ will survive to age $t + y$ can be estimated via

$${}_t\hat{p}_y(\mathbf{x}) = \hat{P}(T > t + y \mid T > y, \mathbf{x}) = exp(-\int_y^{t+y} \hat{h}_0(s) ds \times e^{\hat{\beta}_1 x_1 + \cdots + \hat{\beta}_d x_d}).$$

Moreover, the coefficient estimates $\hat{\beta}_1, \ldots, \hat{\beta}_d$ can be used to explain the impacts of different predictor variables have on the survival probability distribution. Based on the estimation results, we found:

- Positive smoking status and worse substandard rating correspond to higher mortality rates.

- Female, later issue age, and less frequent premium mode correspond to lower mortality rates.
- Smoking status has the most significant impact on the mortality of an individual.
- Premium mode is a statistically significant predictor of mortality for policyholders in a universal life product, but not for the policyholders in whole life and term life products.

### C.4 CONFIDENCE INTERVALS FOR THE DEATH COUNT ESTIMATES

Suppose that there are $m$ active policyholders in the data each have current age $y_j$ and covariate vector $\mathbf{x}_j = (x_{1,j}, \ldots, x_{d,j})$. Then the expected total death count during year $t$ among the active policyholders can be estimated via

$$\hat{N}_t = \sum_{j=1}^{m} {}_t\hat{p}_{y_j}(\mathbf{x}_j) \times [1 - {}_1\hat{p}_{t+y_j}(\mathbf{x}_j)].$$

It is much more challenging to construct a confidence interval for $\hat{N}_t$. To the best of our knowledge, there is no known results about the (asymptotic) distribution of $\hat{N}_t$. To this end, we resort to a parametric bootstrap method to obtain the empirical distribution of $\hat{N}_t$. Some additional notations are needed herein. Assume that there are $n$ policyholders in the mortality data each have issue age $d_i$, observed time $t_i = \min(r_i, c_i)$ where $r_i$ denotes the remaining lifetime since policy issuance and $c_i$ denotes the lapse/surrender time, status indicator $\delta_i = I_{(r_i \le c_i)}$ with $0$ means right censored and $1$ means event at time, and covariate vector $\mathbf{x}_i = (x_{1,i}, \ldots, x_{d,i})$.

The bootstrap procedure is described below:

1. Fit the CPH model to $\{(t_i + d_i, \delta_i, \mathbf{x}_i)\}_{i=1,\ldots,n}$ and obtain the survival function estimate for each individual ${}_t\hat{p}_{d_i}(\mathbf{x}_i)$.
2. Simulate $r_i^*$ according to ${}_t\hat{p}_{d_i}(\mathbf{x}_i)$.
3. Fit the CPH model to $\{(t_i, 1-\delta_i, \mathbf{x}_i)\}_{i=1,\ldots,n}$ and obtain the lapse/surrender time survival function estimate for each individual $\hat{G}(t \mid \mathbf{x}_i)$.
4. Simulate $c_i^*$ according to $\hat{G}(t \mid \mathbf{x}_i)$.
5. Form $t_i^* = \min(r_i^*, c_i^*)$, and $\delta_i^* = I_{(r_i^* \le c_i^*)}$.
6. Fit the CPH model to the resampled data $\{(t_i^* + d_i, \delta_i^*, \mathbf{x}_i)\}_{i=1,\ldots,n}$, and calculate the death count estimate $\hat{N}_t^*$.

Repeat the aforementioned bootstrap procedure $k$ times, then we obtain $k$ death count estimates $\hat{N}^*_{t,i}$, $i=1,\ldots,k.$ The quantile function can be estimated via $Q_p(\hat{N}_t)=\hat{F}^{-1}(p)$, where $p\in[0,1)$ and $\hat{F}$ is the empirical distribution function of $\{\hat{N}^*_{t,i}\}_{i=1,\ldots,k}$. The $p\times 100\%$ confidence interval of $\hat{N}_t$ is then computed as

$$[Q_{p/2}(\hat{N}_t),\ Q_{1-p/2}(\hat{N}_t)].$$

We must remark that a strong assumption involved in the above bootstrap procedure is that the survival time process and the lapse/surrender time process are conditionally independent given the covariate vector. This assumption may be violated if there is a strong dependence presented between the mortality and lapse/surrender behaviors due to the anti-selection issue. Namely, the termination behavior of a policyholder may depend on her/his health status. Such a dependence is particularly strong at the end of the level term period of term insurance products. Overlooking the dependence may significantly impact the prediction accuracy. In a follow-up research, the research team is interested in implementing a copula approach to capture the dependence between the survival time process and termination time process. This modeling approach is innovative from the industrial and academic standpoints.

## References


1. Hill, W. and Woodward, L. S. (2013). Examining the impact learning communities have on college of education students on an urban campus. Journal of College Student Development, 54(6):643–648.

2. Pomerantz, S. B. and Norleen (2000). Impact of learning communities on retention at a metropolitan university. Journal of College Student Retention: Research, Theory and Practice, 2(2):115–126.

3. Ward, M. D. (2015). Learning communities and the undergraduate statistics curriculum: A response to "Mere renovation is too little too late". The American Statistician, Online Discussion, 69(4).

4. Ward, M. D. (2017). Building bridges: The role of an undergraduate mentor. American Statistician, 71(1):30–33.

5. Zhao, C.-M. and Kuh, G. D. (2004). Adding value: Learning communities and student engagement. Research in Higher Education, 45(2):115–138.

## About The Society of Actuaries Research Institute

Serving as the research arm of the Society of Actuaries (SOA), the SOA Research Institute provides objective, data-driven research bringing together tried and true practices and future-focused approaches to address societal challenges and your business needs. The Institute provides trusted knowledge, extensive experience and new technologies to help effectively identify, predict and manage risks.

Representing the thousands of actuaries who help conduct critical research, the SOA Research Institute provides clarity and solutions on risks and societal challenges. The Institute connects actuaries, academics, employers, the insurance industry, regulators, research partners, foundations and research institutions, sponsors and non-governmental organizations, building an effective network which provides support, knowledge and expertise regarding the management of risk to benefit the industry and the public.

Managed by experienced actuaries and research experts from a broad range of industries, the SOA Research Institute creates, funds, develops and distributes research to elevate actuaries as leaders in measuring and managing risk. These efforts include studies, essay collections, webcasts, research papers, survey reports, and original research on topics impacting society.

Harnessing its peer-reviewed research, leading-edge technologies, new data tools and innovative practices, the Institute seeks to understand the underlying causes of risk and the possible outcomes. The Institute develops objective research spanning a variety of topics with its strategic research programs: aging and retirement; actuarial innovation and technology; mortality and longevity; diversity, equity and inclusion; health care cost trends; and catastrophe and climate risk. The Institute has a large volume of topical research available, including an expanding collection of international and market-specific research, experience studies, models and timely research.

Society of Actuaries Research Institute
475 N. Martingale Road, Suite 600
Schaumburg, Illinois 60173
www.SOA.org