Loss Data Analytics

An open text authored by the Actuarial Community

Contents

Preface

Date: 11 September 2018

Book Description

Loss Data Analytics is an interactive, online, freely available text.

- The online version contains many interactive objects (quizzes, computer demonstrations, interactive graphs, video, and the like) to promote deeper learning.
- A subset of the book is available for offline reading in pdf and EPUB formats.
- The online text will be available in multiple languages to promote access to a worldwide audience.

What will success look like?

The online text will be freely available to a worldwide audience. The online version will contain many interactive objects (quizzes, computer demonstrations, interactive graphs, video, and the like) to promote deeper learning. Moreover, a subset of the book will be available in pdf format for low-cost printing. The online text will be available in multiple languages to promote access to a worldwide audience.

How will the text be used?

This book will be useful in actuarial curricula worldwide. It will cover the loss data learning objectives of the major actuarial organizations. Thus, it will be suitable for classroom use at universities as well as for use by independent learners seeking to pass professional actuarial examinations. Moreover, the text will also be useful for the continuing professional development of actuaries and other professionals in insurance and related financial risk management industries.

Why is this good for the profession?

An online text is a type of open educational resource (OER). One important benefit of an OER is that it equalizes access to knowledge, thus permitting a broader community to learn about the actuarial profession. Moreover, it has the capacity to engage viewers through active learning that deepens the learning process, producing analysts more capable of solid actuarial work. Why is this good for students and teachers and others involved in the learning process?

Cost is often cited as an important factor for students and teachers in textbook selection (see a recent post on the $400 textbook). Students will also appreciate the ability to "carry the book around" on their mobile devices.

Why loss data analytics?

The intent is that this type of resource will eventually permeate throughout the actuarial curriculum. Given the dramatic changes in the way that actuaries treat data, loss data seems like a natural place to start. The idea behind the name loss data analytics is to integrate classical loss data models from applied probability with modern analytic tools. In particular, we recognize that big data (including social media and usage based insurance) are here to stay and that high speed computation is readily available.

Project Goal

The project goal is to have the actuarial community author our textbooks in a collaborative fashion.

To get involved, please visit our Loss Data Analytics Project Site.

Acknowledgements

Edward Frees acknowledges the John and Anne Oros Distinguished Chair for Inspired Learning in Business which provided seed money to support the project. Frees and his Wisconsin colleagues also acknowledge a Society of Actuaries Center of Excellence Grant that provided funding to support work in dependence modeling and health initiatives.

We acknowledge the Society of Actuaries for permission to use problems from their examinations.

We thank Rob Hyndman, Monash University, for allowing us to use his excellent style files to produce the online version of the book.

We thank Yihui Xie and his colleagues at Rstudio for the R bookdown package that allows us to produce this book.

We also wish to acknowledge the support and sponsorship of the International Association of Black Actuaries in our joint efforts to provide actuarial educational content to all.

IABA
INTERNATIONAL ASSOCIATION OF BLACK ACTUARIES

Contributors

The project goal is to have the actuarial community author our textbooks in a collaborative fashion. The following contributors have taken a leadership role in developing Loss Data Analytics.

- **Zeinab Amin** is the Director of the Actuarial Science Program and Associate Dean for Undergraduate Studies of the School of Sciences and Engineering at the American University in Cairo (AUC). Amin holds a PhD in Statistics and is an Associate of the Society of Actuaries. Amin is the recipient of the 2016 Excellence in Academic Service Award and the 2009 Excellence in Teaching Award from AUC. Amin has designed and taught a variety of statistics and actuarial science courses. Amin's current area of research includes quantitative risk assessment, reliability assessment, general statistical modelling, and Bayesian statistics.

- **Katrien Antonio**, KU Leuven
- **Jan Beirlant**, KU Leuven
- **Carolina Castro** - University of Buenos Aires
- **Curtis Gary Dean** is the Lincoln Financial Distinguished Professor of Actuarial Science at Ball State University. He is a Fellow of the Casualty Actuarial Society and a CFA charterholder. He has extensive practical experience as an actuary at American States Insurance, SAFECO, and Travelers. He has served the CAS and actuarial profession as chair of the Examination Committee, first editor-in-chief for Variance: Advancing the Science of Risk, and as a member of the Board of Directors and the Executive Council. He contributed a chapter to Predictive Modeling Applications in Actuarial Science published by Cambridge University Press.

- **Edward W. (Jed) Frees** is an emeritus professor, formerly the Hickman-Larson Chair of Actuarial Science at the University of Wisconsin-Madison. He is a Fellow of both the Society of Actuaries and the American Statistical Association. He has published extensively and is a four-time winner of the Halmstad and Prize for best paper published in the actuarial literature. He has written three books and is a co-editor of the two-volume series Predictive Modeling Applications in Actuarial Science published by Cambridge University Press.

- **Guojun Gan** is an assistant professor in the Department of Mathematics at the University of Connecticut, where he has been since August 2014. Prior to that, he worked at a large life insurance company in Toronto, Canada for six years. He received a BS degree from Jilin University, Changchun, China, in 2001 and MS and PhD degrees from York University, Toronto, Canada, in 2003 and 2007, respectively. His research interests include data mining and actuarial science. He has published several books and papers on a variety of topics, including data clustering, variable annuity, mathematical finance, applied statistics, and VBA programming.

- **Lisa Gao** is a doctoral student at the University of Wisconsin-Madison.

- **José Garrido**, Concordia University

- **Noriszura Ismail** is a Professor and Head of Actuarial Science Program, Universiti Kebangsaan Malaysia (UKM). She specializes in Risk Modelling and Applied Statistics. She obtained her BSc and MSc (Actuarial Science) in 1991 and 1993 from University of Iowa, and her PhD (Statistics) in 2007 from UKM. She also passed several papers from Society of Actuaries in 1994. She has received several research grants from Ministry of Higher Education Malaysia (MOHE) and UKM, totaling about MYR1.8 million. She has successfully supervised and co-supervised several PhD students (13 completed and 11 on-going). She currently has about 180 publications, consisting of 88 journal articles and 95 proceedings.

- **Joseph H.T. Kim**, Ph.D., FSA, CERA, is Associate Professor of Applied Statistics at Yonsei University, Seoul, Korea. He holds a Ph.D. degree in Actuarial Science from the University of Waterloo, at which he taught as Assistant Professor. He also worked in the life insurance industry. He has published papers in Insurance Mathematics and Economics, Journal of Risk and Insurance, Journal of Banking and Finance, ASTIN Bulletin, and North American Actuarial Journal, among others.

- **Shyamalkumar Nariankadu** - University of Iowa

- **Nii-Armah Okine** is a dissertator at the business school of University of Wisconsin-Madison with a major in actuarial science. He obtained his master's degree in Actuarial science from Illinois State University. His research interests includes micro-level reserving, joint longitudinal-survival modeling, dependence modelling, micro insurance and machine learning.

- **Margie Rosenberg** - University of Wisconsin

- **Emine Selin Sarıdaş** is a doctoral candidate in the Statistics department of Mimar Sinan University, Istanbul, Turkey. She holds a bachelor degree in Actuarial Science with a minor in Economics and a master degree in Actuarial Science from Hacettepe University. Her research interest includes dependence modeling, regression, loss models and life contingencies.

- **Peng Shi** - University of Wisconsin - Madison

- **Jianxi Su**, Purdue University

- **Tim Verdonck**, KU Leuven

- **Krupa Viswanathan** is an Associate Professor in the Risk, Insurance and Healthcare Management Department in the Fox School of Business, Temple University. She is an Associate of the Society of Actuaries. She teaches courses in Actuarial Science and Risk Management at the undergraduate and graduate levels. Her research interests include corporate governance of insurance companies, capital management, and sentiment analysis. She received her Ph.D. from The Wharton School of the University of Pennsylvania.

Reviewers

Our goal is to have the actuarial community author our textbooks in a collaborative fashion. Part of the writing process involves many reviewers who generously donated their time to help make this book better. They are:

- Chunsheng Ban, Ohio State University
- Vytaras Brazauskas, University of Wisconsin - Milwaukee
- Chun Yong Chew, Universiti Tunku Abdul Rahman (UTAR)
- Eren Dodd, University of Southampton
- Gordon Enderle, University of Wisconsin - Madison
- Rob Erhardt, Wake Forest University
- Liang (Jason) Hong, Robert Morris University
- Hirokazu (Iwahiro) Iwasawa
- Himchan Jeong, University of Connecticut
- Paul Herbert Johnson, University of Wisconsin - Madison
- Samuel Kolins, Lebonan Valley College
- Andrew Soon-Yong Kwon, Zurich Re
- Ambrose Lo, University of Iowa
- Mark Maxwell, University of Texas at Austin
- Tatjana Miljkovic, Miami University
- Bell Ouelega, American University in Cairo
- Zhiyu (Frank) Quan, University of Connecticut
- Jiandong Ren, Western University
- Rajesh V. Sahasrabuddhe, Oliver Wyman
- Ranee Thiagarajah, Illinois State University
- Ping Wang, Saint Johns University
- Chengguo Weng, University of Waterloo
- Toby White, Drake University
- Michelle Xia, Northern Illinois University
- Di (Cindy) Xu, University of Nebraska - Lincoln
- Lina Xu, Columbia University
- Jorge Yslas, University of Copenhagen
- Jeffrey Zheng, Temple University
- Hongjuan Zhou, Arizona State University

For our Readers

We hope that you find this book worthwhile and even enjoyable. For your convenience, at our Github Landing site (https://openacttexts.github.io/), you will find links to the book that you can (freely) download for offline reading, including a pdf version (for Adobe Acrobat) and an EPUB version suitable for mobile devices. Data for running our examples are available at the same site.

In developing this book, we are emphasizing the online version that has lots of great features such as a glossary, code and solutions to examples that you can be revealed interactively. For example, you will find that the statistical code is hidden and can only be seen by clicking on terms such as R Code for Frequency Table. We hide the code because we don't want to insist that you use the `R` statistical software (although we like it). Still, we encourage you to try the some statistical code as you read the book – we have opted to make it easy to learn `R` as you go. We have even set up a separate R Code for Loss Data Analytics site to explain more of the details of the code.

Freely available, interactive textbooks represent a new venture in actuarial education and we need your input. Although a lot of effort has gone into the development, we expect hiccoughs. Please let either your instructor know about opportunities for improvement or write us through the discussion features in the online text or contact chapter contributors directly.

Chapter 1

Introduction to Loss Data Analytics

Chapter Preview. This book introduces readers to methods of analyzing insurance data. Section 1.1 begins with a discussion of why the use of data is important in the insurance industry. Section 1.2 gives a general overview of the purposes of analyzing insurance data which is reinforced in the Section 1.3 case study. Naturally, there is a huge gap between the broad goals summarized in the overview and a case study application; this gap is covered through the methods and techniques of data analysis covered in the rest of the text.

1.1 Relevance of Analytics

In this section, you learn how to:

- Summarize the importance of insurance to consumers and the economy
- How to describe analytics
- To identify data generating events associated with the timeline of a typical insurance contract

This book introduces the process of using data to make decisions in an insurance context. It does not assume that readers are familiar with insurance but introduces insurance concepts as needed. Insurance may not be as entertaining as the sports industry (another industry that depends heavily on data) but it does affect the financial livelihoods of many. By almost any measure, insurance is a major economic activity. On a global level, insurance premiums comprised about 6.2% of the world gross domestic product (GDP) in 2014, (Insurance Information Institute, 2016). As examples, premiums accounted for 18.9% of GDP in Taiwan (the highest in the study) and represented 7.3% of GDP in the United States. On a personal level, almost everyone owning a home has insurance to protect themselves in the event of a fire, hailstorm, or some other calamitous event. Almost every country requires insurance for those driving a car. In sum, although not particulary entertaining, insurance plays an important role in the economies of nation and the lives of individuals.

Insurance is a data-driven industry. Like other major corporations, insurers use data when trying to decide how much to pay employees, how many employees to retain, how to market their services, how to forecast financial trends, and so on. These represent general areas of activities that are not specific to the insurance industry. Although each industry retains its own nuances, you will find that the methods and tools introduced in this text are relevant across industries.

When introducing data methods, we will focus on losses that arise from obligations in insurance contracts. This could be the amount of damage to one's apartment under a renter's insurance agreement, the amount needed to compensate someone that you hurt in a driving accident, and the like. We call these obligations insurance claims or loss amounts. With this focus, we will be able to introduce generally applicable statistical tools and techniques in real-life situations.

1.1.1 What is Analytics?

Insurance is a data-driven industry and analytics is a key to deriving information from data. But what is analytics? Making data-driven business decisions has been described as business analytics, business intelligence, and data science. These terms, among others, are sometimes used interchangeably and sometimes refer to distinct applications. Business intelligence may focus on processes of collecting data, often through databases and data warehouses, whereas business analytics utilizes tools and methods for statistical analyses of data. In contrast to these two terms that emphasize business applications, the term data science can encompass broader applications in many scientific domains. For our purposes, we use the term analytics to refer to the process of using data to make decisions. This process involves gathering data, understanding models of uncertainty, making general inferences, and communicating results.

1.1.2 Short and Long-term Insurance

This text will focus on short-term insurance contracts. By short-term, we mean contracts where the insurance coverage is typically provided for six months or a year. If you are new to insurance, then it is probably easiest to think about an insurance policy that covers the contents of an apartment or house that you are renting (known as renters insurance) or the contents and property of a building that is owned by you or a friend (known as homeowners insurance). Another easy example is automobile insurance. In the event of an accident, this policy may cover damage to your vehicle, damage to other vehicles in the accident, as well as medical expenses of those injured in the accident.

In the US, policies such as renters and homeowners are known as property insurance whereas a policy such as auto that covers medical damages to people is known as casualty insurance. In the rest of the world, these are both known as nonlife or general insurance, to distinguish them from life insurance.

Both life and nonlife insurances are important components of the world economy. The Insurance Information Institute (2016) estimates that direct insurance premiums in the world for 2014 was 2,654,549 for life and 2,123,699 for nonlife; these figures are in millions of US dollars. As noted earlier, the total represents 6.2% of the world GDP. Put another way, life accounts for 55.5% of insurance premiums and 3.4% of world GDP whereas nonlife accounts for 44.5% of insurance premiums and 2.8% of world GDP. Both life and nonlife represent important economic activities and are worthy of study in their own right.

Yet, life insurance considerations differ from nonlife. In life insurance, the default is to have a multi-year contract. For example, if a person 25 years old purchases a whole life policy that pays upon death of the insured and that person does not die until age 100, then the contract is in force for 75 years. We think of this as a long-term contract.

Further, in life insurance, the benefit amount is often stipulated in the contract provisions. In contrast, most short-term contracts provide for compensation of insured losses which are unknown before the accident. (There are usually limits placed on the compensation amounts.) In a life insurance contract that stretches over many years, the time value of money plays a prominent role. In contrast, in a short-term nonlife contract, the random amount of compensation takes priority.

In both life and nonlife insurances, the frequency of claims is very important. For many life insurance contracts, the insured event (such as death) happens only once. In contrast, for nonlife insurances such as automobile, it is common for individuals (especially young male drivers) to get into more than one accident during a year. So, our models need to reflect this observation; we will introduce different frequency models than you may also see when studying life insurance.

For short-term insurance, the framework of the probabilistic model is straightforward. We think of a one-period model (the period length, e.g., six months, will be specified in the situation).

- At the beginning of the period, the insured pays the insurer a known premium that is agreed upon by both parties to the contract.
- At the end of the period, the insurer reimburses the insured for a (possibly multivariate) random loss.

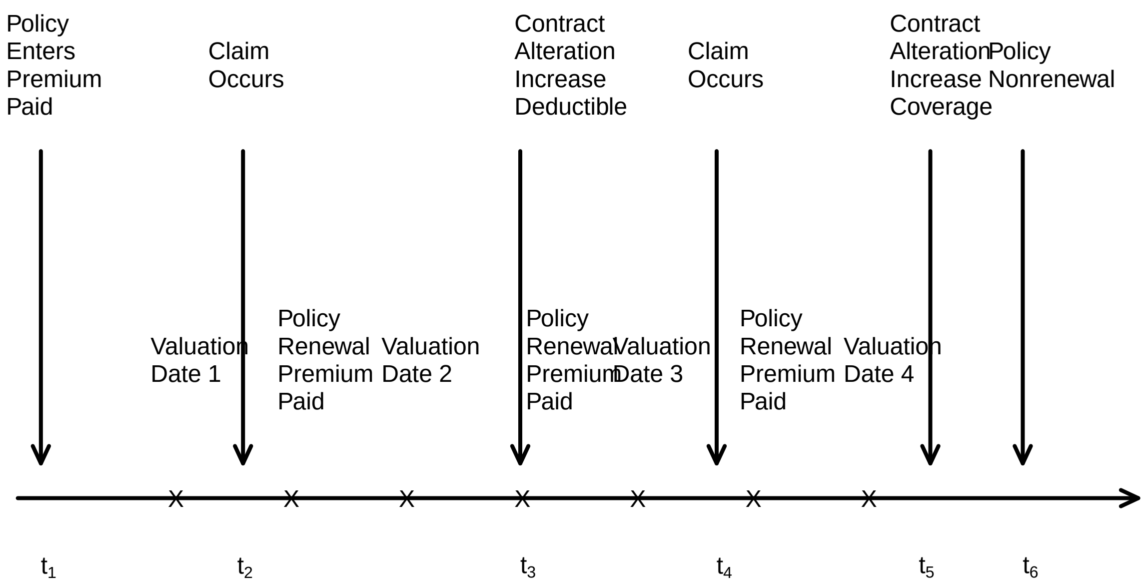

Figure 1.1: Timeline of a Typical Insurance Policy. Arrows mark the occurrences of random events. Each x marks the time of scheduled events that are typically non-random.

This framework will be developed as we proceed but we first focus on integrating this framework with concerns about how the data may arise.

1.1.3 Insurance Processes

One way to describe the data arising from operations of a company that sells insurance products is to use a granular approach. In this micro oriented view, we can think about what happens to a contract at various stages of its existence. Figure 1.1 traces a timeline of a typical insurance contract. Throughout the life of the contract, the company regularly processes events such as premium collection and valuation, described in Section 1.2; these are marked with an **x** on the timeline. Non-regular and unanticipated events also occur. To illustrate, t_2 and t_4 mark the event of an insurance claim (some contracts, such as life insurance, can have only a single claim). Times t_3 and t_5 mark events when a policyholder wishes to alter certain contract features, such as the choice of a deductible or the amount of coverage. Moreover, from a company perspective, one can even think about the contract initiation (arrival, time t_1) and contract termination (departure, time t_6) as uncertain events.

1.2 Insurance Company Operations

In this section, you learn how to:

- Describe five major operational areas of insurance companies.
- Identify the role of data and analytics opportunities within each operational area.

Armed with insurance data, the end goal is to use data to make decisions. We will learn more about methods of analyzing and extrapolating data in future chapters. To begin, let us think about why we want to do the analysis. Let us take the insurer's viewpoint (not a person) and introduce ways of bringing money in, paying it out, managing costs, and making sure that we have enough money to meet obligations.

Specifically, in many insurance companies, it is customary to aggregate detailed insurance processes into larger operational units; many companies use these functional areas to segregate employee activities and areas of responsibilities. Actuaries and other financial analysts work within these units and use data for the following activities:

1. **Initiating Insurance**. At this stage, the company makes a decision as to whether or not to take on a risk (the underwriting stage) and assign an appropriate premium (or rate). Insurance analytics has its actuarial roots in ratemaking, where analysts seek to determine the right price for the right risk.

2. **Renewing Insurance**. Many contracts, particularly in general insurance, have relatively short durations such as 6 months or a year. Although there is an implicit expectation that such contracts will be renewed, the insurer has the opportunity to decline coverage and to adjust the premium. Analytics is also used at this policy renewal stage where the goal is to retain profitable customers.

3. **Claims Management**. Analytics has long been used in (1) detecting and preventing claims fraud, (2) managing claim costs, including identifying the appropriate support for claims handling expenses, as well as (3) understanding excess layers for reinsurance and retention.

4. **Loss Reserving**. Analytic tools are used to provide management with an appropriate estimate of future obligations and to quantify the uncertainty of those estimates.

5. **Solvency and Capital Allocation**. Deciding on the requisite amount of capital and on ways of allocating capital among alternative investments are also important analytics activities. Companies must understand how much capital is needed so that they will have sufficient flow of cash available to meet their obligations. This is an important question that concerns not only company managers but also customers, company shareholders, regulatory authorities, as well as the public at large. Related to issues of how much capital is the question of how to allocate capital to differing financial projects, typically to maximize an investor's return. Although this question can arise at several levels, insurance companies are typically concerned with how to allocate capital to different lines of business within a firm and to different subsidiaries of a parent firm.

Although data represent a critical component of solvency and capital allocation, other components including an economic framework and the financial investments environment are also important. Because of the background needed to address these components, we will not address solvency and capital allocation issues further in this text.

Nonetheless, for all operating functions, we emphasize that analytics in the insurance industry is not an exercise that a small group of analysts can do by themselves. It requires an insurer to make significant investments in their information technology, marketing, underwriting, and actuarial functions. As these areas represent the primary end goals of the analysis of data, additional background on each operational unit is provided in the following subsections.

1.2.1 Initiating Insurance

Setting the price of an insurance product can be a perplexing problem. This is in contrast to other industries such as manufacturing where the cost of a product is (relatively) known and provides a benchmark for assessing a market demand price. Similarly, in other areas of financial services, market prices are available and provide the basis for a market-consistent pricing structure of products. However, for many lines of insurance, the cost of a product is uncertain and market prices are unavailable. Expectations of the random cost is a reasonable place to start for a price. (If you have studies finance, then you will recall that an expectation is the optimal price for a risk-neutral insurer.) It has been traditional in insurance pricing to begin with the expected cost. Insurers then add to this margins to account for the product's riskiness, expenses incurred in servicing the product, and an allowance for profit/surplus of the company.

Use of expected costs as a foundation for pricing is prevalent in some lines of the insurance business. These include automobile and homeowners insurance. For these lines, analytics has served to sharpen the market by making the calculation of the product's expected cost more precise. The increasing availability of the internet to consumers has also promoted transparency in pricing; in today's marketplace, consumers have ready access to competing quotes from a host of insurers. Insurers seek to increase their market share by refining their risk classification systems and employing skimming the cream underwriting strategies. Recent surveys (e.g., Earnix (2013)) indicate that pricing is the most common use of analytics among insurers.

Underwriting, the process of classifying risks into homogenous categories and assigning policyholders to these categories, lies at the core of ratemaking. Policyholders within a class have similar risk profiles and so are charged the same insurance price. This is the concept of an actuarially fair premium; it is fair to charge different rates to policyholders only if they can be separated by identifiable risk factors. An early article, Two Studies in Automobile Insurance Ratemaking (Bailey and LeRoy, 1960), provided a catalyst to the acceptance of analytic methods in the insurance industry. This paper addresses the problem of classification ratemaking. It describes an example of automobile insurance that has five use classes cross-classified with four merit rating classes. At that time, the contribution to premiums for use and merit rating classes were determined independently of each other. Thinking about the interacting effects of different classification variables is a more difficult problem.

1.2.2 Renewing Insurance

Insurance is a type of financial service and, like many service contracts, insurance coverage is often agreed upon for a limited time period, such as six months or a year, at which time commitments are complete. Particularly for general insurance, the need for coverage continues and so efforts are made to issue a new contract providing similar coverage. Renewal issues can also arise in life insurance, e.g., term (temporary) life insurance, although other contracts, such as life annuities, terminate upon the insured's death and so issues of renewability are irrelevant.

In absence of legal restrictions, at renewal the insurer has the opportunity to:

- accept or decline to underwrite the risk and
- determine a new premium, possibly in conjunction with a new classification of the risk.

Risk classification and rating at renewal is based on two types of information. First, as at the initial stage, the insurer has available many rating variables upon which decisions can be made. Many variables will not change, e.g., sex, whereas others are likely to have changed, e.g., age, and still others may or may not change, e.g., credit score. Second, unlike the initial stage, at renewal the insurer has available a history of policyholder's loss experience, and this history can provide insights into the policyholder that are not available from rating variables. Modifying premiums with claims history is known as experience rating, also sometimes referred to as merit rating.

Experience rating methods are either applied retrospectively or prospectively. With retrospective methods, a refund of a portion of the premium is provided to the policyholder in the event of favorable (to the insurer) experience. Retrospective premiums are common in life insurance arrangements (where policyholders earn dividends in the U.S. and bonuses in the U.K.). In general insurance, prospective methods are more common, where favorable insured experience is rewarded through a lower renewal premium.

Claims history can provide information about a policyholder's risk appetite. For example, in personal lines it is common to use a variable to indicate whether or not a claim has occurred in the last three years. As another example, in a commercial line such as worker's compensation, one may look to a policyholder's average claim over the last three years. Claims history can reveal information that is hidden (to the insurer) about the policyholder.

1.2.3 Claims and Product Management

In some of areas of insurance, the process of paying claims for insured events is relatively straightforward. For example, in life insurance, a simple death certificate is all that is needed as the benefit amount is provided in the contract terms. However, in non-life areas such as property and casualty insurance, the process is much more complex. Think about even a relatively simple insured event such as automobile accident. Here, it is often helpful to determine which party is at fault, one needs to assess damage to all of the vehicles and people involved in the incident, both insured and non-insured, the expenses incurred in assessing the damages, and so forth. The process of determining coverage, legal liability, and settling claims is known as claims adjustment.

Insurance managers sometimes use the phrase claims leakage to mean dollars lost through claims management inefficiencies. There are many ways in which analytics can help manage the claims process, Gorman and Swenson (2013). Historically, the most important has been fraud detection. The claim adjusting process involves reducing information asymmetry (the claimant knows exactly what happened; the company knows some of what happened). Mitigating fraud is an important part of claims management process.

Fraud detection is only one aspect of managing claims. More broadly, one can think about claims management as consisting of the following components:

- **Claims triaging**. Just as in the medical world, early identification and appropriate handling of high cost claims (patients, in the medical world), can lead to dramatic company savings. For example, in workers compensation, insurers look to achieve early identification of those claims that run the risk of high medical costs and a long payout period. Early intervention into those cases could give insurers more control over the handling of the claim, the medical treatment, and the overall costs with an earlier return-to-work.
- **Claims processing**. The goal is to use analytics to identify routine situations that are anticipated to have small payouts and more complex situations. More complex situations may require more experienced adjusters and legal assistance to appropriately handle claims with high potential payouts.
- **Adjustment decisions**. Once a complex claim has been identified and assigned to an adjuster, analytic driven routines can be established to aid subsequent decision-making processes. Such processes can also be helpful for adjusters in developing case reserves, an estimate of the insurer's future liability. This is an important input to the insurer's loss reserves, described in Section 1.2.4.

In addition to the insured's reimbursement for insured losses, the insurer also needs to be concerned with another source of revenue outflow, expenses. Loss adjustment expenses are part of an insurer's cost of managing claims. Analytics can be used to reduce expenses directly related to claims handling (allocated) as well as general staff time for overseeing the claims processes (unallocated). The insurance industry has high operating costs relative to other portions of the financial services sectors.

In addition to claims payments, there are many other ways in which insurers use to data to manage their products. We have already discussed the need for analytics in underwriting, that is, risk classification at the initial acquisition stage. Insurers are also interested in which policyholders elect to renew their contract and, as with other products, monitor customer loyalty.

Analytics can also be used to manage the portfolio, or collection, of risks that an insurer has acquired. When the risk is initially obtained, the insurer's risk can be managed by imposing contract parameters that modify contract payouts. Chapters **??** and **??** describe common modifications including coinsurance, deductibles, and policy upper limits.

After the contract has been agreed upon with an insured, the insurer may still modify its net obligation by entering into a reinsurance agreement. This type of agreement is with a reinsurer, an insurer of an insurer. It is common for insurance companies to purchase insurance on its portfolio of risks to gain protection from unusual events, just as people and other companies do.

1.2.4 Loss Reserving

An important feature that distinguishes insurance from other sectors of the economy is the timing of the exchange of considerations. In manufacturing, payments for goods are typically made at the time of a transaction. In contrast, for insurance, money received from a customer occurs in advance of benefits or services; these are rendered at a later date. This leads to the need to hold a reservoir of wealth to meet future obligations in respect to obligations made. The size of this reservoir of wealth, and the importance of ensuring its adequacy, is a major concern for the insurance industry.

Setting aside money for unpaid claims is known as loss reserving; in some jurisdictions, reserves are also known as technical provisions. We saw in Figure 1.1 several times at which a company summarizes its financial position; these times are known as valuation dates. Claims that arise prior to valuation dates have typically been or are about to paid; claims in the future of these valuation dates are unknown. A company must estimate these outstanding liabilities when determining its financial strength. Accurately determining loss reserves is important to insurers for many reasons.

1. Loss reserves represent an anticipated claim that the insurer owes its customers. Under-reserving may result in a failure to meet claim liabilities. Conversely, an insurer with excessive reserves may present a weaker financial position than it truly has.
2. Reserves provide an estimate for the unpaid cost of insurance that can be used for pricing contracts.
3. Loss reserving is required by laws and regulations. The public has a strong interest in the financial strength of insurers.
4. In addition to insurance company management and regulators, other stakeholders such as investors and customers make decisions that depend on company loss reserves.

Loss reserving is a topic where there are substantive differences between life and general (also known as property and casualty, or non-life), insurance. In life insurance, the severity (amount of loss) is often not a source of uncertainty as payouts are specified in the contract. The frequency, driven by mortality of the insured, is a concern. However, because of the length of time for settlement of life insurance contracts, the time value of money uncertainty as measured from issue to date of death can dominate frequency concerns. For example, for an insured who purchases a life contract at age 25, it would not be unusual for the contract to still be open in 75 years time. See, for example, Bowers et al. (1986) or Dickson et al. (2013) for introductions to reserving for life insurance.

1.3 Case Study: Wisconsin Property Fund

In this section, we use the Wisconsin Property Fund as a case study. You learn how to:

- Describe how data generating events can produce data of interest to insurance analysts.
- Identify each type of variable.
- Produce relevant summary statistics for each variable.
- Describe how these summary statistcs can be used in each of the major operational areas of an insurance company.

Let us illustrate the kind of data under consideration and the goals that we wish to achieve by examining the Local Government Property Insurance Fund (LGPIF), an insurance pool administered by the Wisconsin Office of the Insurance Commissioner. The LGPIF was established to provide property insurance for local government entities that include counties, cities, towns, villages, school districts, and library boards. The fund insures local government property such as government buildings, schools, libraries, and motor vehicles. The fund covers all property losses except those resulting from flood, earthquake, wear and tear, extremes in temperature, mold, war, nuclear reactions, and embezzlement or theft by an employee.

The property fund covers over a thousand local government entities who pay approximately $25 million in premiums each year and receive insurance coverage of about $75 billion. State government buildings are not covered; the LGPIF is for local government entities that have separate budgetary responsibilities and who need insurance to moderate the budget effects of uncertain insurable events. Coverage for local government property has been made available by the State of Wisconsin since 1911.

1.3.1 Fund Claims Variables: Frequency and Severity

At a fundamental level, insurance companies accept premiums in exchange for promises to compensate a policyholder upon the occurrence of an insured event. Indemnification is the compensation provided by the insurer for incurred hurt, loss, or damage that is covered by the policy. This compensation is also known as a claim. The extent of the payout, known as the severity, is a key financial expenditure for an insurer.

In terms of money outgo, an insurer is indifferent to having ten claims of 100 when compared to one claim of 1,000. Nonetheless, it is common for insurers to study how often claims arise, known as the frequency of claims. The frequency is important for expenses, but it also influences contractual parameters (such as deductibles and policy limits that are described later) that are written on a per occurrence basis, is routinely monitored by insurance regulators, and can be a key driver in the overall indemnification obligation of the insurer. We shall consider the frequency and severity as the two main outcome variables that we wish to understand, model, and manage.

To illustrate, in 2010 there were 1,110 policyholders in the property fund who experienced a total of 1,377 claims. Table 1.1 shows the distribution. Almost two-thirds (0.637) of the policyholders did not have any claims and an additional 18.8% had only one claim. The remaining 17.5% (=1 - 0.637 - 0.188) had more than one claim; the policyholder with the highest number recorded 239 claims. The average number of claims for this sample was 1.24 (=1377/1110).

Table 1.1: 2010 Claims Frequency Distribution

Type											
Number	0	1	2	3	4	5	6	7	8	9 or more	Sum
Count	707	209	86	40	18	12	9	4	6	19	1,110
Claims	0	209	172	120	72	60	54	28	48	617	1,377
Proportion making claim	s 0.637	0.188	0.077	0.036	0.016	0.011	0.008	0.004	0.005	0.017	1.000

R Code for Frequency Table

```
Insample <- read.csv("Insample.csv", header=T,  na.strings=c("."), stringsAsFactors=FALSE)
Insample2010 <- subset(Insample, Year==2010)
table(Insample2010$Freq)
```

For the severity distribution, one common approach is to examine the distribution of the sample of 1,377 claims. However, another common approach is to examine the distribution of the average claims of those policyholders with claims. In our 2010 sample, there were 403 (=1110-707) such policyholders. For 209 of these policyholders with one claim, the average claim equals the only claim they experienced. For the policyholder with highest frequency, the average claim is an average over 239 separately reported claim events. This average is also known as the pure premium or loss cost.

Table 1.2 summarizes the sample distribution of average severities from the 403 policyholders; it shows that

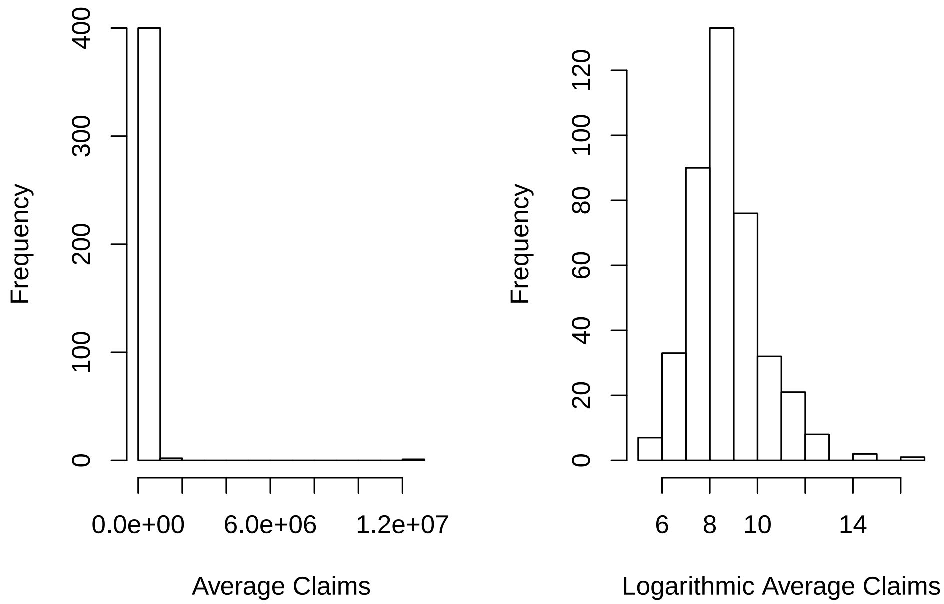

Figure 1.2: Distribution of Positive Average Severities

the average claim amount was 56,330 (all amounts are in US Dollars). However, the average gives only a limited look at the distribution. More information can be gleaned from the summary statistics which show a very large claim in the amount of 12,920,000. Figure 1.2 provides further information about the distribution of sample claims, showing a distribution that is dominated by this single large claim so that the histogram is not very helpful. Even when removing the large claim, you will find a distribution that is skewed to the right. A generally accepted technique is to work with claims in logarithmic units especially for graphical purposes; the corresponding figure in the right-hand panel is much easier to interpret.

Table 1.2: 2010 Average Severity Distribution

Minimum	First Quartile	Median	Mean	Third Quartile	Maximum
167	2,226	4,951	56,330	11,900	12,920,000

R Code for Severity Distribution Table and Figures

```
Insample <- read.csv("Data/PropertyFundInsample.csv", header=T, na.strings=c("."), stringsAsFactors=FALS
Insample2010 <- subset(Insample, Year==2010)
InsamplePos2010 <- subset(Insample2010, yAvg>0)
# Table
summary(InsamplePos2010$yAvg)
length(InsamplePos2010$yAvg)
# Figures
par(mfrow=c(1, 2))
hist(InsamplePos2010$yAvg, main="", xlab="Average Claims")
hist(log(InsamplePos2010$yAvg), main="", xlab="Logarithmic Average Claims")
```

1.3.2 Fund Rating Variables

Developing models to represent and manage the two outcome variables, frequency and severity, is the focus of the early chapters of this text. However, when actuaries and other financial analysts use those models, they do so in the context of external variables. In general statistical terminology, one might call these explanatory or predictor variables; there are many other names in statistics, economics, psychology, and other disciplines. Because of our insurance focus, we call them rating variables as they will be useful in setting insurance rates and premiums.

We earlier considered a sample of 1,110 observations which may seem like a lot. However, as we will see in our forthcoming applications, because of the preponderance of zeros and the skewed nature of claims, actuaries typically yearn for more data. One common approach that we adopt here is to examine outcomes from multiple years, thus increasing the sample size. We will discuss the strengths and limitations of this strategy later but, at this juncture, just want to show the reader how it works.

Specifically, Table 1.3 shows that we now consider policies over five years of data, years 2006, …, 2010, inclusive. The data begins in 2006 because there was a shift in claim coding in 2005 so that comparisons with earlier years are not helpful. To mitigate the effect of open claims, we consider policy years prior to 2011. An open claim means that all of the obligations are not known at the time of the analysis; for some claims, such an injury to a person in an auto accident or in the workplace, it can take years before costs are fully known.

Table 1.3 shows that the average claim varies over time, especially with the high 2010 value due to a single large claim [1]. The total number of policyholders is steadily declining and, conversely, the coverage is steadily increasing. The coverage variable is the amount of coverage of the property and contents. Roughly, you can think of it as the maximum possible payout of the insurer. For our immediate purposes, it is our first rating variable. Other things being equal, we would expect that policyholders with larger coverage will have larger claims. We will make this vague idea much more precise as we proceed.

Table 1.3: Building and Contents Claims Summary

Year	Average Frequency	Average Severity	Average Coverage	Number of Policyholders
2006	0.951	9,695	32,498,186	1,154
2007	1.167	6,544	35,275,949	1,138
2008	0.974	5,311	37,267,485	1,125
2009	1.219	4,572	40,355,382	1,112
2010	1.241	20,452	41,242,070	1,110

R Code for Building and Contents Claims Summary

```
Insample <- read.csv("Data/PropertyFundInsample.csv", header=T, na.strings=c("."), stringsAsFactors=FAL
library(doBy)
T1A <- summaryBy(Freq ~ Year, data = Insample,
   FUN = function(x) { c(m = mean(x), num=length(x)) } )
T1B <- summaryBy(yAvg    ~ Year, data = Insample,
   FUN = function(x) { c(m = mean(x), num=length(x)) } )
T1C <- summaryBy(BCcov    ~ Year, data = Insample,
   FUN = function(x) { c(m = mean(x), num=length(x)) } )
Table1In <- cbind(T1A[1],T1A[2],T1B[2],T1C[2],T1A[3])
names(Table1In) <- c("Year", "Average Frequency","Average Severity", "Average","Number of Policyholders
Table1In
```

[1] Note that the average severity in Table 1.3 differs from that reported in Table 1.2. This is because the former includes policyholders with zero claims where as the latter does not. This is an important distinction that we will address in later portions of the text.

For a different look at this data, Table 1.4 summarizes the distribution of our two outcomes, frequency and claims amount. In each case, the average exceeds the median, suggesting that the two distributions are right-skewed. In addition, the table summarizes our continuous rating variables, coverage and deductible amount. The table also suggests that these variables also have right-skewed distributions.

Table 1.4: Summary of Claim Frequency and Severity, Deductibles, and Coverages

	Minimum	Median	Average	Maximum
Claim Frequency	0	0	1.109	263
Claim Severity	0	0	9,292	12,922,218
Deductible	500	1,000	3,365	100,000
Coverage (000's)	8.937	11,354	37,281	2,444,797

R Code for Summary of Claim Frequency and Severity, Deductibles, and Coverages

```
Insample <- read.csv("Data/PropertyFundInsample.csv", header=T, na.strings=c("."), stringsAsFactors=FALS
t1<- summaryBy(Insample$Freq ~ 1, data = Insample,
   FUN = function(x) { c(ma=min(x), m1=median(x),m=mean(x),mb=max(x)) } )
names(t1) <- c("Minimum", "Median","Average", "Maximum")
t2 <- summaryBy(Insample$yAvg ~ 1, data = Insample,
   FUN = function(x) { c(ma=min(x), m1=median(x), m=mean(x),mb=max(x)) } )
names(t2) <- c("Minimum", "Median","Average", "Maximum")
t3 <- summaryBy(Deduct ~ 1, data = Insample,
   FUN = function(x) { c(ma=min(x), m1=median(x), m=mean(x),mb=max(x)) } )
names(t3) <- c("Minimum", "Median","Average", "Maximum")
t4 <- summaryBy(BCcov/1000 ~ 1, data = Insample,
   FUN = function(x) { c(ma=min(x), m1=median(x), m=mean(x),mb=max(x)) } )
names(t4) <- c("Minimum", "Median","Average", "Maximum")
Table2 <- rbind(t1,t2,t3,t4)
Table2a <- round(Table2,3)
Rowlable <- rbind("Claim Frequency","Claim Severity","Deductible","Coverage (000's)")
Table2aa <- cbind(Rowlable,as.matrix(Table2a))
Table2aa
```

The following display describes the rating variables considered in this chapter. To handle the skewness, we henceforth focus on logarithmic transformations of coverage and deductibles. To get a sense of the relationship between the non-continuous rating variables and claims, Table 1.5 relates the claims outcomes to these categorical variables. Table 1.5 suggests substantial variation in the claim frequency and average severity of the claims by entity type. It also demonstrates higher frequency and severity for the `Fire5` variable and the reverse for the `NoClaimCredit` variable. The relationship for the `Fire5` variable is counter-intuitive in that one would expect lower claim amounts for those policyholders in areas with better public protection (when the protection code is five or less). Naturally, there are other variables that influence this relationship. We will see that these background variables are accounted for in the subsequent multivariate regression analysis, which yields an intuitive, appealing (negative) sign for the `Fire5` variable.

Description of Rating Variables

Variable	*Description*
EntityType	Categorical variable that is one of six types: (Village, City, County, Misc, School, or Town)
LnCoverage	Total building and content coverage, in logarithmic millions of dollars
LnDeduct	Deductible, in logarithmic dollars
AlarmCredit	Categorical variable that is one of four types: (0, 5, 10, or 15) for automatic smoke alarms in main rooms
NoClaimCredit	Binary variable to indicate no claims in the past two years
Fire5	Binary variable to indicate the fire class is below 5 (The range of fire class is 0 to 10)

Table 1.5: Claims Summary by Entity Type, Fire Class, and No Claim Credit

Variable	Number of Policies	Claim Frequency	Average Severity
EntityType			
Village	1,341	0.452	10,645
City	793	1.941	16,924
County	328	4.899	15,453
Misc	609	0.186	43,036
School	1,597	1.434	64,346
Town	971	0.103	19,831
Fire			
Fire5=0	2,508	0.502	13,935
Fire5=1	3,131	1.596	41,421
No Claims Credit			
NoClaimCredit=0	3,786	1.501	31,365
NoClaimCredit=1	1,853	0.310	30,499
Total	5,639	1.109	31,206

R Code for Claims Summary by Entity Type, Fire Class, and No Claim Credit

```
ByVarSumm<-function(datasub){
  tempA <- summaryBy(Freq    ~ 1 , data = datasub,
     FUN = function(x) { c(m = mean(x), num=length(x)) } )
  datasub1 <-  subset(datasub, yAvg>0)
  tempB <- summaryBy(yAvg   ~ 1, data = datasub1,FUN = function(x) { c(m = mean(x)) } )
  tempC <- merge(tempA,tempB,all.x=T)[c(2,1,3)]
  tempC1 <- as.matrix(tempC)
  return(tempC1)
  }
datasub <-  subset(Insample, TypeVillage == 1);
t1 <- ByVarSumm(datasub)
datasub <-  subset(Insample, TypeCity == 1);
t2 <- ByVarSumm(datasub)
datasub <-  subset(Insample, TypeCounty == 1);
t3 <- ByVarSumm(datasub)
datasub <-  subset(Insample, TypeMisc == 1);
t4 <- ByVarSumm(datasub)
datasub <-  subset(Insample, TypeSchool == 1);
```

```
t5 <- ByVarSumm(datasub)
datasub <-  subset(Insample, TypeTown == 1);
t6 <- ByVarSumm(datasub)
datasub <-  subset(Insample, Fire5 == 0);
t7 <- ByVarSumm(datasub)
datasub <-  subset(Insample, Fire5 == 1);
t8 <- ByVarSumm(datasub)
datasub <-  subset(Insample, Insample$NoClaimCredit == 0);
t9 <- ByVarSumm(datasub)
datasub <-  subset(Insample, Insample$NoClaimCredit == 1);
t10 <- ByVarSumm(datasub)
t11 <- ByVarSumm(Insample)

Tablea <- rbind(t1,t2,t3,t4,t5,t6,t7,t8,t9,t10,t11)
Tableaa <- round(Tablea,3)
Rowlable <- rbind("Village","City","County","Misc","School",
          "Town","Fire5--No","Fire5--Yes","NoClaimCredit--No",
        "NoClaimCredit--Yes","Total")
Table4 <- cbind(Rowlable,as.matrix(Tableaa))
Table4
```

Table 1.6 shows the claims experience by alarm credit. It underscores the difficulty of examining variables individually. For example, when looking at the experience for all entities, we see that policyholders with no alarm credit have on average lower frequency and severity than policyholders with the highest (15%, with 24/7 monitoring by a fire station or security company) alarm credit. In particular, when we look at the entity type School, the frequency is 0.422 and the severity 25,523 for no alarm credit, whereas for the highest alarm level it is 2.008 and 85,140. This may simply imply that entities with more claims are the ones that are likely to have an alarm system. Summary tables do not examine multivariate effects; for example, Table 1.5 ignores the effect of size (as we measure through coverage amounts) that affect claims.

Table 1.6: Claims Summary by Entity Type and Alarm Credit (AC) Category

Entity Type	AC0 Claim Frequency	AC0 Avg. Severity	AC0 Num. Policies	AC5 Claim Frequency	AC5 Avg. Severity	AC5 Num. Policies
Village	0.326	11,078	829	0.278	8,086	54
City	0.893	7,576	244	2.077	4,150	13
County	2.140	16,013	50	-	-	1
Misc	0.117	15,122	386	0.278	13,064	18
School	0.422	25,523	294	0.410	14,575	122
Town	0.083	25,257	808	0.194	3,937	31
Total	0.318	15,118	2,611	0.431	10,762	239

Table 1.7: Claims Summary by Entity Type and Alarm Credit (AC) Category

Entity Type	AC10 Claim Frequency	AC10 Avg. Severity	AC10 Num. Policies	AC15 Claim Frequency	AC15 Avg. Severity	AC15 Num. Policies
Village	0.500	8,792	50	0.725	10,544	408
City	1.258	8,625	31	2.485	20,470	505
County	2.125	11,688	8	5.513	15,476	269
Misc	0.077	3,923	26	0.341	87,021	179

Entity Type	AC10 Claim Frequency	AC10 Avg. Severity	AC10 Num. Policies	AC15 Claim Frequency	AC15 Avg. Severity	AC15 Num. Policies
School	0.488	11,597	168	2.008	85,140	1,013
Town	0.091	2,338	44	0.261	9,490	88
Total	0.517	10,194	327	2.093	41,458	2,462

R Code for Claims Summary by Entity Type and Alarm Credit Category

```
#Claims Summary by Entity Type and Alarm Credit
ByVarSumm<-function(datasub){
  tempA <- summaryBy(Freq    ~ AC00 , data = datasub,
                     FUN = function(x) { c(m = mean(x), num=length(x)) } )
  datasub1 <-  subset(datasub, yAvg>0)
  if(nrow(datasub1)==0) { n<-nrow(datasub)
    return(c(0,0,n))
  } else
  {
    tempB <- summaryBy(yAvg   ~ AC00, data = datasub1,
                       FUN = function(x) { c(m = mean(x)) } )
    tempC <- merge(tempA,tempB,all.x=T)[c(2,4,3)]
    tempC1 <- as.matrix(tempC)
    return(tempC1)
  }
}
AlarmC <- 1*(Insample$AC00==1) + 2*(Insample$AC05==1)+ 3*(Insample$AC10==1)+ 4*(Insample$AC15==1)
ByVarCredit<-function(ACnum){
datasub <-  subset(Insample, TypeVillage == 1 & AlarmC == ACnum);
  t1 <- ByVarSumm(datasub)
datasub <-  subset(Insample, TypeCity == 1 & AlarmC == ACnum);
  t2 <- ByVarSumm(datasub)
datasub <-  subset(Insample, TypeCounty == 1 & AlarmC == ACnum);
  t3 <- ByVarSumm(datasub)
datasub <-  subset(Insample, TypeMisc == 1 & AlarmC == ACnum);
  t4 <- ByVarSumm(datasub)
datasub <-  subset(Insample, TypeSchool == 1 & AlarmC == ACnum);
  t5 <- ByVarSumm(datasub)
datasub <-  subset(Insample, TypeTown == 1 & AlarmC ==ACnum);
  t6 <- ByVarSumm(datasub)
datasub <-  subset(Insample, AlarmC == ACnum);
  t7 <- ByVarSumm(datasub)
Tablea <- rbind(t1,t2,t3,t4,t5,t6,t7)
Tableaa <- round(Tablea,3)
Rowlable <- rbind("Village","City","County","Misc","School",
                  "Town","Total")
Table4 <- cbind(Rowlable,as.matrix(Tableaa))
}
Table4a <- ByVarCredit(1)    #Claims Summary by Entity Type and Alarm Credit==00
Table4b <- ByVarCredit(2)    #Claims Summary by Entity Type and Alarm Credit==05
Table4c <- ByVarCredit(3)    #Claims Summary by Entity Type and Alarm Credit==10
Table4d <- ByVarCredit(4)    #Claims Summary by Entity Type and Alarm Credit==15
```

1.3.3 Fund Operations

We have now seen the Fund's two outcome variables, a count variable for the number of claims and a continuous variable for the claims amount. We have also introduced a continuous rating variable, coverage, discrete quantitative variable, (logarithmic) deductibles, two binary rating variables, no claims credit and fire class, as well as two categorical rating variables, entity type and alarm credit. Subsequent chapters will explain how to analyze and model the distribution of these variables and their relationships. Before getting into these technical details, let us first think about where we want to go. General insurance company functional areas are described in Section 1.2; let us now think about how these areas might apply in the context of the property fund.

Initiating Insurance

Because this is a government sponsored fund, we do not have to worry about selecting good or avoiding poor risks; the fund is not allowed to deny a coverage application from a qualified local government entity. If we do not have to underwrite, what about how much to charge?

We might look at the most recent experience in 2010, where the total fund claims were approximately 28.16 million USD (= 1377 claims $\times$ 20452 average severity). Dividing that among 1,110 policyholders, that suggests a rate of 24,370 ($\approx$ 28,160,000/1110). However, 2010 was a bad year; using the same method, our premium would be much lower based on 2009 data. This swing in premiums would defeat the primary purpose of the fund, to allow for a steady charge that local property managers could utilize in their budgets.

Having a single price for all policyholders is nice but hardly seems fair. For example, Table 1.5 suggests that Schools have much higher claims than other entities and so should pay more. However, simply doing the calculation on an entity by entity basis is not right either. For example, we saw in Table 1.6 that had we used this strategy, entities with a 15% alarm credit (for good behavior, having top alarm systems) would actually wind up paying more.

So, we have the data for thinking about the appropriate rates to charge but will need to dig deeper into the analysis. We will explore this topic further in Chapter **??** on premium calculation fundamentals. Selecting appropriate risks is introduced in Chapter **??** on risk classification.

Renewing Insurance

Although property insurance is typically a one-year contract, Table 1.3 suggests that policyholders tend to renew; this is typical of general insurance. For renewing policyholders, in addition to their rating variables we have their claims history and this claims history can be a good predictor of future claims. For example, Table 1.5 shows that policyholders without a claim in the last two years had much lower claim frequencies than those with at least one accident (0.310 compared to 1.501); a lower predicted frequency typically results in a lower premium. This is why it is common for insurers to use variables such as `NoClaimCredit` in their rating. We will explore this topic further in Chapter **??** on experience rating.

Claims Management

Of course, the main story line of 2010 experience was the large claim of over 12 million USD, nearly half the claims for that year. Are there ways that this could have been prevented or mitigated? Are their ways for the fund to purchase protection against such large unusual events? Another unusual feature of the 2010 experience noted earlier was the very large frequency of claims (239) for one policyholder. Given that there were only 1,377 claims that year, this means that a single policyholder had 17.4 % of the claims. This also suggests opportunities for managing claims, the subject of Chapter **??**.

Loss Reserving

In our case study, we look only at the one year outcomes of closed claims (the opposite of open). However, like many lines of insurance, obligations from insured events to buildings such as fire, hail, and the like, are not known immediately and may develop over time. Other lines of business, including those were there are injuries to people, take much longer to develop. Chapter **??** introduces this concern and loss reserving, the discipline of determining how much the insurance company should retain to meet its obligations.

1.4 Further Resources and Contributors

Contributor

- **Edward W. (Jed) Frees**, University of Wisconsin-Madison, is the principal author of the initial version of this chapter. Email: jfrees@bus.wisc.edu for chapter comments and suggested improvements.

This book introduces loss data analytic tools that are most relevant to actuaries and other financial risk analysts. Here are a few reference cited in the chapter.

Bibliography

Bailey, Robert A. and J. Simon LeRoy (1960). “Two studies in automobile ratemaking,” *Proceedings of the Casualty Actuarial Society Casualty Actuarial Society*, Vol. XLVII.

Bowers, Newton L., Hans U. Gerber, James C. Hickman, Donald A. Jones, and Cecil J. Nesbitt (1986). *Actuarial Mathematics*. Society of Actuaries Itasca, Ill.

Dickson, David C. M., Mary Hardy, and Howard R. Waters (2013). *Actuarial Mathematics for Life Contingent Risks*. Cambridge University Press.

Earnix (2013). “2013 Insurance Predictive Modeling Survey,” Earnix and Insurance Services Office, Inc. URL: http://earnix.com/2013-insurance-predictive-modeling-survey/3594/, [Retrieved on July 7, 2014].

Gorman, Mark and Stephen Swenson (2013). “Building believers: How to expand the use of predictive analytics in claims,” SAS, URL: http://www.sas.com/resources/whitepaper/wp_59831.pdf, [Retrieved on August 17, 2014].

Insurance Information Institute (2016). “International Insurance Fact Book,” Insurance Information Institute, URL: http://www.iii.org/sites/default/files/docs/pdf/international_insurance_factbook_2016.pdf, [Retrieved on Sept 9, 2018].