Introduction to Loss Data Analytics

Fall 2017

Outline

Relevance of Insurance

By almost any measure, insurance is a major economy activity

- On a global level, insurance premiums comprised about 6.3% of the world gross domestic product (GDP) in 2013 (Source: International Insurance Fact Book: 2015)
  - Premiums accounted for 11.2% of GDP in Japan
  - Represented 7.5% 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

Analytics and Loss Data

- Insurance is big business
- Because of the size, it is not surprising that these firms employ analytics in the same manner as other large corporations
- These areas include (i) sales and marketing, (ii) compensation analysis, (iii) productivity analysis, and (iv) financial forecasting. For example, in sales and marketing:
  - Predict customer behavior/needs (target appropriate customers)
  - Anticipate customer reactions to promotions/rate changes
  - Manage acquisition costs (online sales, agent compensation)
- One could introduce analytics from many perspectives; we focus on *loss data*, also known as *insurance claims* or *insurance amounts*

What is Analytics?

- Insurance is a data-driven industry – analytics is a key to deriving information from data
- But what is analytics?

What is Analytics?

- Insurance is a data-driven industry – analytics is a key to deriving information from data
- But what is analytics?
- Some alternative descriptors:
  - “business intelligence” may focus on processes of collecting data, often through databases and data warehouses
  - “business analytics” utilizes tools and methods for statistical analyses of data
  - “data science” can encompass broader applications in many scientific domains

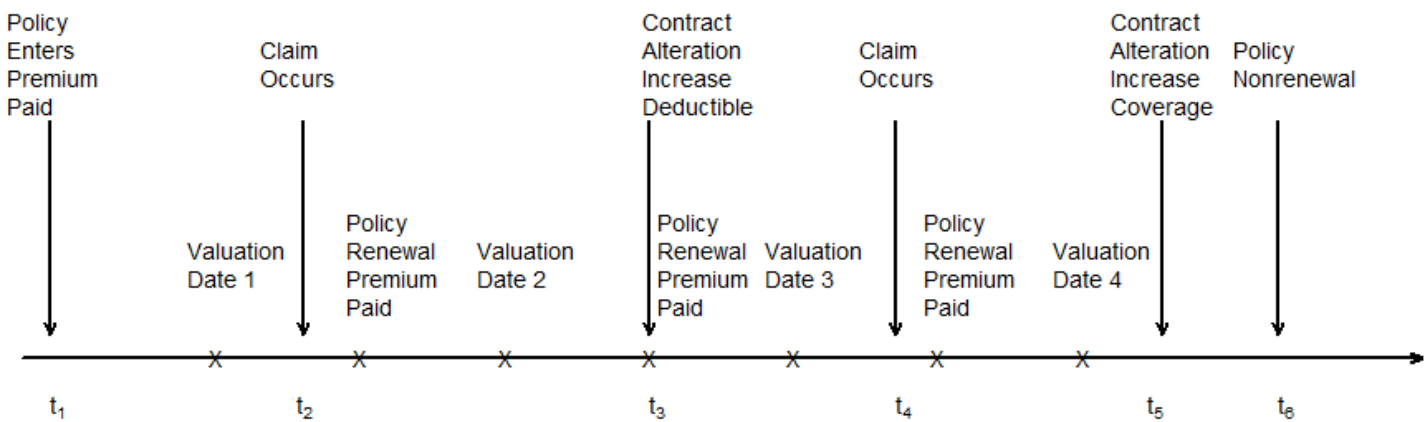
What is Analytics?

- Insurance is a data-driven industry – analytics is a key to deriving information from data
- But what is analytics?
- Some alternative descriptors:
  - "business intelligence" may focus on processes of collecting data, often through databases and data warehouses
  - "business analytics" utilizes tools and methods for statistical analyses of data
  - "data science" can encompass broader applications in many scientific domains
- Analytics – the process of using data to make decisions
  - This process involves gathering data, understanding models of uncertainty, making general inferences, and communicating results

Insurance Processes

- How does data arise from an insurer?
- In a “micro” oriented view, we can think specifically about what happens to a contract at various stages of its existence

Figure: Timeline of a Typical Insurance Policy. Arrows mark the occurrences of random events.



Variable Types

- As in many areas of disciplined thought:
  - Work with data (summarize)
  - Represent data using models
  - Use models calibrated with data to make decisions
- As a first step, it is helpful to classify a variable into one of several types
- This classification drives the summarization and modeling decisions that we will make

Variable Types

Variable Type	Example
Qualitative	
Binary	Sex
Categorical (Unordered, Nominal)	Territory (e.g., state/province) in which an insured resides
Ordered Category (Ordinal)	Claimant satisfaction (five point scale ranging from 1=dissatisfied to 5 =satisfied)
Quantitative	
Continuous	Policyholder's age, weight, income
Discrete	Amount of deductible
Count	Number of insurance claims
Combinations of Discrete and Continuous	Policy losses, mixture of 0's (for no loss) and positive claim amount
Interval Variable	Driver Age: 16-24 (young), 25-54 (intermediate), 55 and over (senior)
Circular Data	Time of day measures of customer arrival
Multivariate Variable	
High Dimensional Data	Characteristics of a firm purchasing worker's compensation insurance (location of plants, industry, number of employees, and so on)
Spatial Data	Longitude/latitude of the location an insurance hailstorm claim
Missing Data	Policyholder's age (continuous/interval) and "-99" for "not reported," that is, missing
Censored and Truncated Data	Amount of insurance claims in excess of a deductible
Aggregate Claims	Losses recorded for each claim in a motor vehicle policy.
Stochastic Process Realizations	The time and amount of each occurrence of an insured loss

Insurance Company Operations I

Insurer's Viewpoint:
✓Need ways of bringing money in, paying it out, managing costs, and making sure that we have enough money to meet obligations

✓Insurers aggregate detailed insurance processes into larger "operational" units

- Initiating Insurance
  - Offer right price for the right risk (underwriting)
  - Avoid adverse selection
- Renewing Insurance
  - Retain profitable customers longer
  - Update prices using experience (bonus-malus, credibility)

Insurance Company Operations II

- Claims and Product Management
  - Detect and manage claims fraud
  - Manage claims costs (triaging, processing, adjustment decisions)
  - Understand excess layers for reinsurance and retention
- Reserving
  - Predict future obligations: set an adequate reserve to fund losses that have been incurred but not developed
  - Quantify the uncertainty of the estimates
  - Match projections of obligations to income streams
  - Consider “lines of business;” for non-life, typical lines include personal auto, personal homeowners, commercial auto, etc.
- Capital Allocation and Solvency
  - Decide appropriate level of necessary capital
  - Manage external stakeholders’ expectations; regulators, rating agencies, reputation

Operations – Initiating Insurance

- Setting the price of an insurance good can be a perplexing problem.
  - In manufacturing, the cost of a good is (relatively) known
  - In other areas of financial services, market prices are available
  - In many lines of insurance, start with an expected cost, add “margins” to account for the product’s riskiness, expenses incurred in servicing the product, and a profit/surplus allowance for the insurance company.

Operations – Initiating Insurance

- Setting the price of an insurance good can be a perplexing problem.
  - In manufacturing, the cost of a good is (relatively) known
  - In other areas of financial services, market prices are available
  - In many lines of insurance, start with an expected cost, add "margins" to account for the product's riskiness, expenses incurred in servicing the product, and a profit/surplus allowance for the insurance company.
- For some lines of business, especially automobile and homeowners insurance, analytics has served to sharpen the market by making the calculation of the good's expectation more precise.
- Pricing strategies now routinely involve generalized linear model (GLM) techniques
- *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.

Big Data

- Traditionally, insurers use information reported by policyholders on application forms, combined with selected external sources.
  - E.g., police reports for automobile insurance or medical exam results for life insurance.
  - Many variables are categorical, making even limited info complex.
- Now, there is interest in collecting more information about policyholders
  - An early example was the use of credit scores by Progressive Insurance for automobile insurance.
  - Ethically permissible? - these debates are important.
  - From a statistician's viewpoint, these additional sources have proven to be significant from hypothesis testing, predictive, and economic viewpoints.

Big Data

- Traditionally, insurers use information reported by policyholders on application forms, combined with selected external sources.
  - E.g., police reports for automobile insurance or medical exam results for life insurance.
  - Many variables are categorical, making even limited info complex.
- Now, there is interest in collecting more information about policyholders
  - An early example was the use of credit scores by Progressive Insurance for automobile insurance.
  - Ethically permissible? - these debates are important.
  - From a statistician's viewpoint, these additional sources have proven to be significant from hypothesis testing, predictive, and economic viewpoints.
- Policyholders are also agreeing to let insurers gather more data about them for risk classification purposes.
  - The best example is the GPS and cameras that are mounting in cars for monitoring how much and how a policyholder drives.
  - There are many that feel that this will represent a tremendous benefit to society.

Operations – Renewing Insurance

- Renewal analytics are similar to initial underwriting but a little different. In addition to (contract initiation) rating variables, we also have claims history.
  - Rating variables (e.g., credit score) may evolve over time
  - Common to use history (e.g., whether a claim has occurred in the last three years)
- Ways to incorporate experience into prices
  - a “bonus-malus” system where prior claim frequency is used to prospectively adjust premiums. Bonus-malus methods of experience rating are used extensively in automobile insurance pricing in Europe and Asia,
  - “credibility” adjusted prices that are a blend of a manual rate with past experience - used since the 1920’s
  - bonus-malus and credibility are prospective systems. In life insurance, “dividends” are awarded retrospectively.

Operations – Claims Management

Insurance managers sometimes use the phrase "*claims leakage*" to mean dollars lost through claims management inefficiencies.

- Fraud detection. Mitigating fraud is an important part of claims management process.
- Management of claims severity
  - 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.
  - Claims processing. The goal is to use analytics to identify situations suitable for small claims handling processes and those for adjuster assignment to complex claims.
  - 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.
- 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.

Operations – Claims and Product Management

Analytics is used in:

- Fraud
- Claims Management
  - Fraud, Claims Triaging, Expense Allocation
- Product Management
  - Customer Loyalty, Price Optimization
- Portfolio Management
  - Portfolio risk distribution
- Reinsurance

Operations – Loss Reserving

- The primary goal of loss reserving is to set an adequate reserve to fund losses that have been incurred but not yet developed.
- Insurance companies are organized by “lines of business”
  - Top-level is life versus non-life
  - For non-life, typical lines include personal auto, personal homeowners, commercial auto, and so forth.
- In non-life insurance, losses are arranged in a triangular fashion as they develop over time and as different obligations are incurred from year to year.
- This triangular format emphasizes the longitudinal and censored nature of the data.

Loss Reserve Example

Table 1. Incremental Paid Losses for Personal Auto Line (in thousand of dollars)

		Development Lag									
Accident Year	Premiums	0	1	2	3	4	5	6	7	8	9
1988	4,711,333	1,376,384	1,211,168	535,883	313,790	168,142	79,972	39,235	15,030	10,865	4,086
1989	5,335,525	1,576,278	1,437,150	652,445	342,694	188,799	76,956	35,042	17,089	12,507	
1990	5,947,504	1,763,277	1,540,231	678,959	364,199	177,108	78,169	47,391	25,288		
1991	6,354,197	1,779,698	1,498,531	661,401	321,434	162,578	84,581	53,449			
1992	6,738,172	1,843,224	1,573,604	613,095	299,473	176,842	106,296				
1993	7,079,444	1,962,385	1,520,298	581,932	347,434	238,375					
1994	7,254,832	2,033,371	1,430,541	633,500	432,257						
1995	7,739,379	2,072,061	1,458,541	727,098							
1996	8,154,065	2,210,754	1,517,501								
1997	8,435,918	2,206,886									

Table 2. Incremental Paid Losses for Commercial Auto Line (in thousand of dollars)

		Development Lag									
Accident Year	Premiums	0	1	2	3	4	5	6	7	8	9
1988	267,666	33,810	45,318	46,549	35,206	23,360	12,502	6,602	3,373	2,373	778
1989	274,526	37,663	51,771	40,998	29,496	12,669	11,204	5,785	4,220	1,910	
1990	268,161	40,630	56,318	56,182	32,473	15,828	8,409	7,120	1,125		
1991	276,821	40,475	49,697	39,313	24,044	13,156	12,595	2,908			
1992	270,214	37,127	50,983	34,154	25,455	19,421	5,728				
1993	280,568	41,125	53,302	40,289	39,912	6,650					
1994	344,915	57,515	67,881	86,734	18,109						
1995	371,139	61,553	132,208	20,923							
1996	323,753	112,103	33,250								
1997	221,448	37,554									

Data from Frees and Shi (2011)

Loss Reserve Example

Table 1. Incremental Paid Losses for

Accident Year	Premiums	0	1	2
1988	4,711,333	1,376,384	1,211,168	535,883
1989	5,335,525	1,576,278	1,437,150	652,445
1990	5,947,504	1,763,277	1,540,231	678,959
1991	6,354,197	1,779,698	1,498,531	661,401
1992	6,738,172	1,843,224	1,573,604	613,095
1993	7,079,444	1,962,385	1,520,298	581,932
1994	7,254,832	2,033,371	1,430,541	633,500
1995	7,739,379	2,072,061	1,458,541	727,098
1996	8,154,065	2,210,754	1,517,501	
1997	8,435,918	2,206,886		

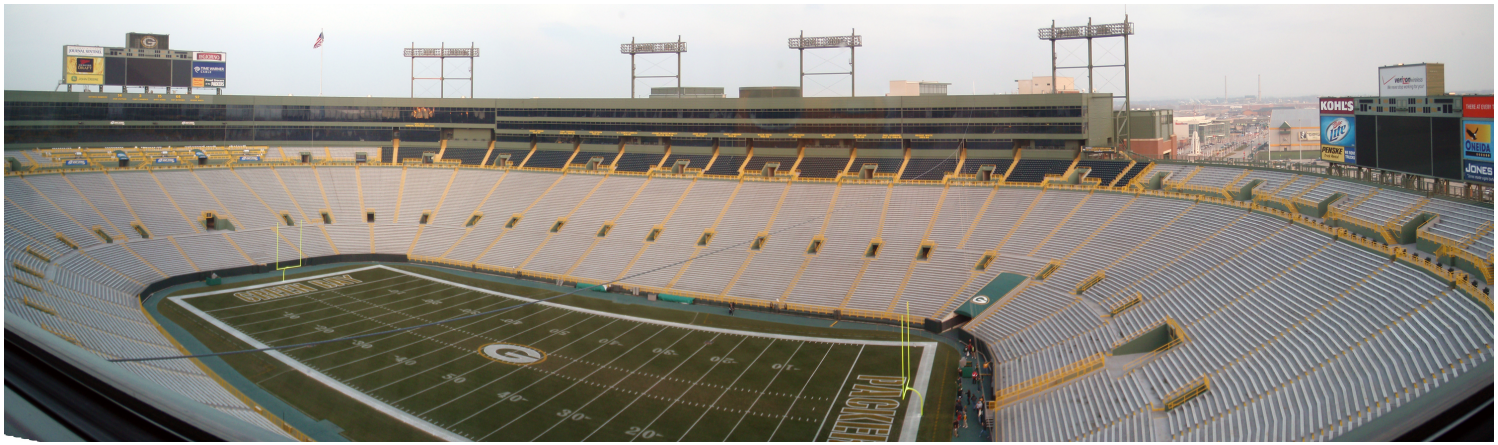
Wisconsin Property Fund

- The Wisconsin Office of the Insurance Commissioner administers the Local Government Property Insurance Fund (LGPIF)
- Property coverage has been available since 1911
- The fund insures property such as government buildings, schools, libraries, and motor vehicles
- “Local government” entities include counties, cities, towns, villages, school districts, and library boards
  - The fund has over 1,000 such entities

LGPIF Policyholder A

- Example – Madison Metropolitan School District
  - it has 98 buildings, 18 major pieces of equipment (mowers, etc.), and 630 properties in the open (benches, playsets, goals, etc.)
  - the property coverage alone is $640 million
  - this is Crestwood Elementary School, one of the 98 buildings

LGPIF Policyholder B



- The largest contract – City of Green Bay
  - contains 118 sites
  - one of which is Lambeau Field – a stadium in which a professional football team, the Green Bay Packers, plays
  - Property coverage is approximately $2.4 billion
  - LGPIF has a separate terrorism reinsurance coverage for this property

Property Fund

- The fund receives approximately $25 million in premiums each year and provides insurance coverage for about $75 billion
- The fund offers three major groups of insurance coverage: building and contents, construction equipment, and motor vehicles
- For building and contents, 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

Claims Frequency (2010)

- The table shows 1,110 policyholders who have 1,377 claims
- Almost two-thirds (0.637) of the policyholders did not have any claims, 18.8% had one claim and 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)

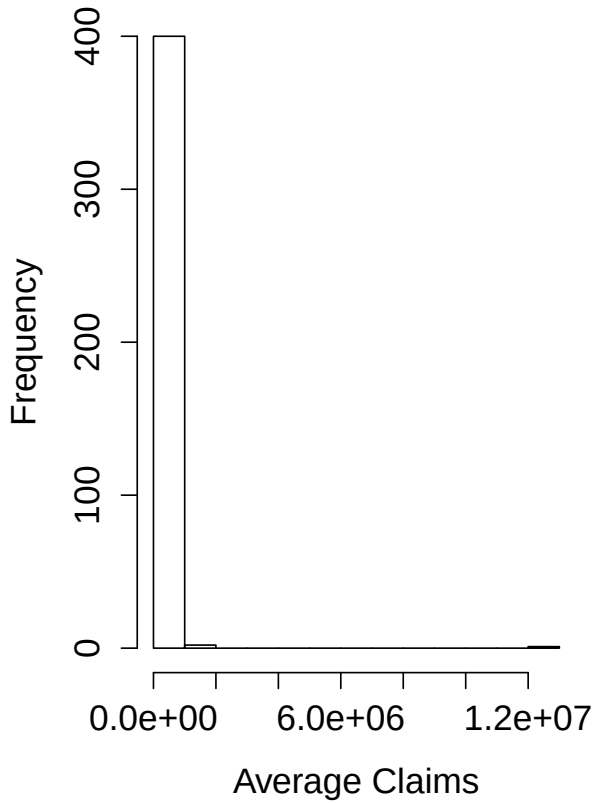
Type	Number of Claims					
Number	0	1	2	3	4	5
Count	707	209	86	40	18	12
Proportion	0.637	0.188	0.077	0.036	0.016	0.011
Number	6	7	8	9 or more	Sum	
Count	9	4	6	19	1,110	
Proportion	0.008	0.004	0.005	0.017	1.000	

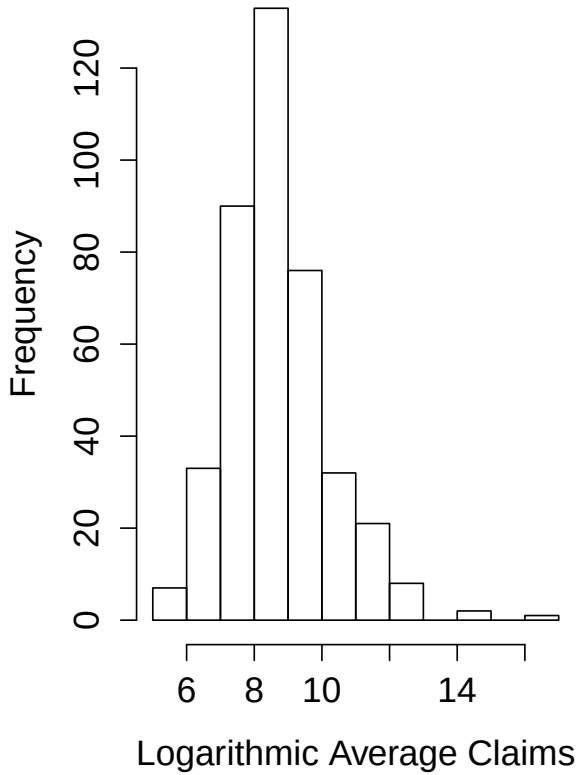
Claims Severity (2010)

- 403 (=1110-707) policyholders had at least one claim
- The following summarizes the distribution of the average claims of those policyholders with claims
  - To illustrate, 209 policyholders had only one claim. Here, the claim amount equals the average claim

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

Claims Severity Distribution (2010)





Claim Outcomes and Coverage by Year

- To increase the sample size, look to more years
- Average frequency is more stable than severity over years
- Coverage is stable and increasing
- Number of policyholders is stable but declining

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

Claim Frequency and Severity, Deductibles, and Coverages

- The two outcomes variables are frequency and severity. Each has many zeros (more than half)
- The two rating variables are deductible and coverages
- For each of the four distributions, Mean > Median, suggesting skewed distributions

	Minimum	Median	Mean	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

Description of Rating Variables

- Identify variables types: binary, categorical, quantitative (discrete), and continuous

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 $0 \sim 10$)

Claims by Entity Type, Fire Class, and No Claim Credit

- There is substantial variation in the frequency and severity by entity type
- As anticipated, lower frequency and severity when the policyholder had no claims in the past two years, (`NoClaimCredit=1`=1)
- Higher frequency and severity for the `Fire5` (=1) variable
  - Counter-intuitive: one would expect lower claim amounts for those policyholders in areas with better public protection (when the protection code is five or less)

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
`Fire5`=0	2,508	0.502	13,935
`Fire5`=1	3,131	1.596	41,421
`NoClaimCredit`=0	3,786	1.501	31,365
`NoClaimCredit`=1	1,853	0.310	30,499
Total	5,639	1.109	31,206

Claims Summary by Entity Type and Alarm Credit Category

- Counter-intuitive results for Alarm Credit. Would expect lower frequency/severity for 15% alarm credits

Entity Type	No Alarm Credit Claim Frequency	No Alarm Credit Avg. Severity	No Alarm Credit Num. Policies	Alarm Credit 5% Claim Frequency	Alarm Credit 5% Avg. Severity	Alarm Credit 5% 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

Entity Type	Alarm Credit 10% Claim Frequency	Alarm Credit 10% Avg. Severity	Alarm Credit 10% Num. Policies	Alarm Credit 15% Claim Frequency	Alarm Credit 15% Avg. Severity	Alarm Credit 15% 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
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

Initiating Insurance

- Because coverage cannot be denied, underwriting not a major issue
- How much to charge?
  - Based on 2010 data, might use 33,026

$$= \frac{\text{total fund claims}}{\text{number of policyholders}} = \frac{\text{36.66 million USD}}{1110}$$

- However, very different answer based on 2009 data (9,934)
- Single premium for all policyholders does not seem fair
  - Outcomes seem to vary by entity type
  - Charge more for those with greater amounts of coverage
  - What about alarm credits???

Renewing Insurance

- For renewing policyholders, in addition to their rating variables, we have their claims history
- Claims history can be a good predictor of future claims
  - For example, 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

Claims Management

- For the 2010 experience, a single large claim of over 12 million USD resulted in over a third of 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; 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

Conclusion

More information is available on *Loss Data Analytics* at

`https://sites.google.com/a/wisc.edu/loss-data-analytics/`