



---

AMERICAN ACADEMY *of* ACTUARIES

---

*Objective. Independent. Effective.™*

**Report on the Stop Loss Factors**

**American Academy of Actuaries  
Stop Loss Factors Work Group**

Presented to the National Association of Insurance Commissioners'  
Health Risk-Based Capital (E) Working Group

November 14, 2016

The American Academy of Actuaries is an 18,500+ member professional association whose mission is to serve the public and the U.S. actuarial profession. For more than 50 years, the Academy has assisted public policymakers on all levels by providing leadership, objective expertise, and actuarial advice on risk and financial security issues. The Academy also sets qualification, practice, and professionalism standards for actuaries in the United States.

## **Stop Loss Factors Work Group**

Chairperson: David Olsho, MAAA, FSA

Devin Dixon, MAAA, ASA

Thomas Doran, MAAA, FSA

Jeremy Freestone, MAAA, FSA

James Gutterman, MAAA, FSA

Annette James, MAAA, FSA, FCA, EA

John Mange, MAAA, FSA

Jack Oldenburg, MAAA, FSA

Shaun Peterson, MAAA, FSA

John Price, MAAA, FCA

Bernard Rabinowitz, MAAA, FSA, FIA,  
FCIA, CERA

David Stefanski, MAAA, FSA

In 2009, the National Association of Insurance Commissioners (NAIC) requested that the American Academy of Actuaries (Academy) review the Risk-Based Capital (RBC) formulas currently used for medical excess of loss business and recommend one formula for all carriers writing this business using data from the Society of Actuaries (SOA). As a result, the Academy's Stop Loss Work Group (work group) developed this report.

### Data Collection

In November 2009, the Academy and SOA sent out a letter requesting stop loss data for medical self-funded stop loss (specific and aggregate), medical portfolio excess, HMO reinsurance, and provider excess coverage. A copy of the data request letter and the data request survey (worksheet) are shown in Appendix 1. To encourage data contribution by keeping individual company data confidential, the process of data collection, aggregation, and analysis was handled by SOA staff and only the aggregated blinded results were given to the work group. Data were collected by the SOA for policy years 1998 to 2008.

The Academy work group provided the blinded data to Joseph Marker, an instructor at the University of Michigan, and worked with him to analyze the de-identified data while maintaining the data contributors' confidentiality.

### Data Analysis

Marker summarized the carriers' data and submitted preliminary summarized data to the work group in July 2010, which reviewed those data. Because the data were limited for medical portfolio excess, HMO reinsurance, and provider excess coverage data, the work group recommended that Marker include only the self-funded stop loss data for his final analysis.

Marker sent final summarized data in November 2010. These stop loss data were collected from 20 companies over 11 years. Of those 20 companies, 17 produced usable data, with 146 company years (CYs) of data. (The three excluded companies included one company that provided only one year of data.) These data are shown on the second page Appendix 2.

Marker, with one of his students, further analyzed the data and produced a final report dated April 29, 2015. That report is shown as Appendix 3.

### Marker Report Highlights

For the purpose of RBC analysis, it was determined that an adverse result was when the combined ratio exceeded 100%, that is, the stop loss block had an actual loss. (The calculation was Reported Claims + Target Expenses > Reported Premium.) Under this condition, there were 31 CYs out of 146 CYs of adverse results, 21.2% of total CYs, with losses ranging from 0.1% to 118.8% of reported premium. Of those 31 CYs, only 4 CYs (2.7% of total) had a loss exceeding 25%. This cutoff was selected because the current stop loss RBC factor is 25% of premium.

It should be noted that one 11-year data contributor had no more than \$36,090 premium in any one year and had three of the four CYs with losses exceeding 25%. Excluding that one company would

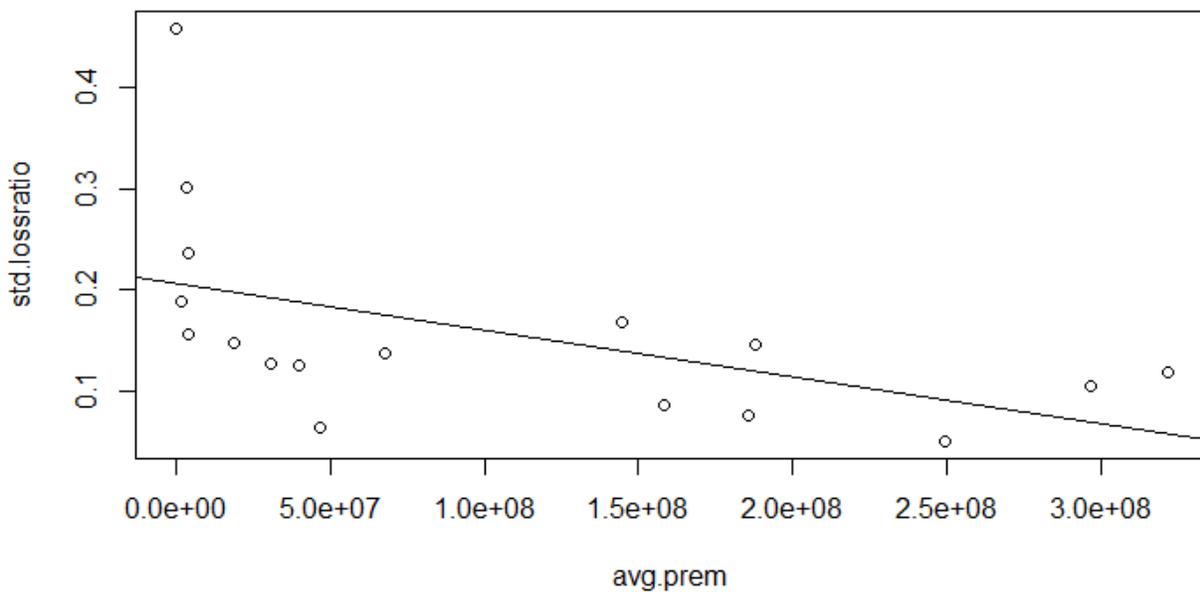
change the results in the above paragraph to 26 CYs of adverse results out of 135 total CYs, 19.3% of total. There was one CY (0.7%) with about \$66 million in premium and a 27.4% loss. The table below summarizes the above results.

Self-Funded Stop Loss Adverse Results				
	Adverse Results Definition	Companies with Adverse Results	Total Companies	Adverse Results Percentage
All Companies	Loss > 0%	31	146	21.2%
	Loss > 25%	4	146	2.7%
Companies with > \$1,000,000 Premium	Loss > 0%	26	135	19.3%
	Loss > 25%	1	135	0.7%

An alternate definition was also considered, defined as a CY when the company made less than target profit (or had a loss). This produced 62 CYs of adverse results (42.5%). This alternate definition was not used because RBC is meant to protect the companies' insureds, not the shareholders.

An additional highlight of Marker's report is that the smaller the block, the more variable the block, as shown below:

Linear Regression Standard Deviation of Loss Ratio as a Function of Premium



This result suggests that it may be reasonable to have a tiered factor, similar to the comprehensive medical, Medicare Supplement, and dental lines of business.

The work group also has some comments on portions of Marker's report, as follows:

- On page 10, section 6.2, which is a discussion of Model 2, Revised, the report states that the dependent variable is the standard deviation of the loss ratio. The work group believes that the dependent variable is something related to an adverse result rather than the standard deviation of the loss ratio.
- On page 14, there are a couple of references to the "mediums" of the loss ratios. We believe it should be the "medians" of the loss ratios.
- On page 15, there are several references to "turning point." As discussed below, we generally refer to this as a "break point."

### Possible RBC Factors

As stated above, the current RBC factor is 25% of premium. This is higher than all but 4 CY out of 146 CY losses (and all but 1 CY of 135 CYs with premium greater than \$1 million). Also produced was an Excel spreadsheet that allowed the work group to model various tiered RBC factors and determine the number of adverse results. A portion of the results section of that spreadsheet is shown below. The column definitions and example of table lines are:

Turning Point (or break point): The premium amount up to which the x% possible RBC factor is applied, and above which the y% RBC factor is applied.

x%, y%: possible RBC factors, as a percentage of premium, below and above the turning (break) point.

Number of Adverse Results in Excess of RBC Factor Calculation: The number of years the adverse results are higher than the calculation using the turning point, x%, and y% shown.

For example, the line reading \$10,000,000/30/10/5 means that there would be 5 adverse CY results with a factor of 30% of the first \$10 million and 10% of the excess of annualized premium.

### Sample RBC Factor Results Table

			Number of Adverse Results in Excess of				Number of Adverse Results in Excess of				
Turning Point	x%	y%	RBC Factor Calculation	Turning Point	x%	y%	RBC Factor Calculation	Turning Point	x%	y%	RBC Factor Calculation
\$10,000,000	20	0	19	\$50,000,000	20	0	10	\$100,000,000	20	0	8
\$10,000,000	20	5	11	\$50,000,000	20	5	8	\$100,000,000	20	5	4
\$10,000,000	20	10	9	\$50,000,000	20	10	6	\$100,000,000	20	10	4
\$10,000,000	25	0	16	\$50,000,000	25	0	10	\$100,000,000	25	0	4
\$10,000,000	25	5	11	\$50,000,000	25	5	7	\$100,000,000	25	5	4
\$10,000,000	25	10	6	\$50,000,000	25	10	5	\$100,000,000	25	10	4
\$10,000,000	30	0	15	\$50,000,000	30	0	7	\$100,000,000	30	0	2
\$10,000,000	30	5	10	\$50,000,000	30	5	5	\$100,000,000	30	5	2
\$10,000,000	30	10	5	\$50,000,000	30	10	4	\$100,000,000	30	10	2
\$100,000,000	20	0	8	\$150,000,000	20	0	4	\$250,000,000	20	0	2
\$100,000,000	20	5	5	\$150,000,000	20	5	4	\$250,000,000	20	5	2
\$100,000,000	20	10	5	\$150,000,000	20	10	4	\$250,000,000	20	10	2
\$100,000,000	25	0	4	\$150,000,000	25	0	4	\$250,000,000	25	0	2
\$100,000,000	25	5	4	\$150,000,000	25	5	4	\$250,000,000	25	5	2
\$100,000,000	25	10	4	\$150,000,000	25	10	4	\$250,000,000	25	10	2
\$100,000,000	30	0	2	\$150,000,000	30	0	2	\$250,000,000	30	0	2
\$100,000,000	30	5	2	\$150,000,000	30	5	2	\$250,000,000	30	5	2
\$100,000,000	30	10	2	\$150,000,000	30	10	2	\$250,000,000	30	10	2

There is also an interactive portion of the spreadsheet, which can produce alternative results. For example, using:

- \$25,000,000/30%/15%, you get 4 adverse results
- \$25,000,000/30%/20%, you get 3 adverse results
- \$25,000,000/30%/30%, you get 2 adverse results

Note in all cases, 2 of the adverse results in excess are CYs with premiums under \$11,000.

Limitations of the Work Group Report

We believe there are several issues the NAIC should consider when reviewing the Work Group report.

The graph below shows data contributors (blinded), including three companies not included in the results discussed above. There are more contributors in the later years than in the earlier years. This could be the result of companies that wrote stop loss at the beginning of the study period experiencing higher-than-expected loss ratios and dropping out of the market (i.e., “survivorship bias”). If that is the case, the contributing companies may have better experience, on average, than other companies in the stop loss market. It is also possible that companies with below average results did not contribute data. We included all usable data from the contributing companies and did not exclude data from companies that exited the stop loss market before the end of the data collection period, which may mitigate some of the possible survivorship bias in the data.



Another issue to consider is the effect of the Affordable Care Act (ACA). All data from this study are from when \$1 million lifetime maximums were predominant. The ACA has eliminated lifetime and annual maximums, so very high claims are not included in this study. This may cause more variable results.

Also, possibly as the result of the ACA, there has been a relatively large growth in small group stop loss. Therefore, current companies may have a larger proportion of smaller groups' premium and claims, relative to the contributed data. While smaller groups individually have more variable results, when combined with other groups in a company's block of stop loss, we believe that total premium volume remains predictive of the variability of a company's overall results.

### Recommended RBC Factors

The members of the work group considered options such as Monte Carlo analyses and Risk of Ruin calculations. We believe that we did not have enough data to build an appropriate models, given the limited number of contributors and CYs of data. We were concerned that any model built would imply a higher level of credibility than was warranted by the data. We therefore chose to select a break point and percentage of premium factors using the Excel spreadsheet provided with the Marker report that allowed the work group to model various tiered RBC factors and break points and examine the results.

Before the consideration of the ACA, the consensus of the work group was that factors of 30%/20% with a break point of \$25,000,000 would be reasonable. At that level, there are 3 CYs of adverse results, two of which are for CYs less than \$11,000 and one of which is for a CY of \$66.3 million. These factors would result in

- A 5% increase in the Stop Loss RBC requirement for companies with up to \$25,000,000 in annual premium,
- A smaller increase for companies with up to \$50,000,000 in annual premium, and
- A decrease for companies with over \$50,000,000 in premium (e.g., the weighted factor at \$100,000,000 would be 22.5%).

To account for the added volatility due to the ACA, we believe an increase of both factors by 5% would be appropriate. Therefore, the work group's final recommended factors are 35%/25%, with a break point of \$25,000,000. These factors would increase the Stop Loss RBC requirement by 10% of the first \$25,000,000 in annual premium.



AMERICAN ACADEMY *of* ACTUARIES



SOCIETY OF ACTUARIES

November xx, 2009

#### Stop-Loss Data Contributing Companies:

The National Association of Insurance Commissioners (NAIC) has requested that the American Academy of Actuaries<sup>1</sup> review the Risk-Based Capital (RBC) formulas currently used for medical excess of loss business and recommend one formula for all carriers (HMOs, Blues plans, A&H insurers, P&C insurers) writing this business. As a result, the Academy established the Stop-Loss Work Group (Work Group), and is partnering with the Society of Actuaries<sup>2</sup> (SOA) in response to this request.

To accomplish its charge, the Work Group is conducting a study of financial data from insurance and reinsurance companies offering stop loss insurance, including self-funded stop loss (specific and aggregate), medical (portfolio) excess, HMO reinsurance and provider excess coverage. The Work Group is working with the SOA to contact companies and collect applicable data for this study. The process of collection, analysis, and aggregation of the data will be handled by SOA staff. All data submissions will be kept confidential. Society of Actuaries staff will process the responses and release only aggregated results. No individual company data will be published. Where comparative results are presented, no companies will be identified. The Work Group will then use this aggregated data to develop a proposed formula for consideration by the NAIC. Keys to the success of this effort will include obtaining a critical mass of experience from insurance companies and reinsurers for each of the stop loss product lines.

The goal of the Work Group regarding the proposed Stop-Loss RBC formula is that the formula be:

- Reasonable, relative to other products;
- Theoretically sound;
- Relatively simple;
- Consistent for Life/Health carriers, Health Organizations, and Property/Casualty carriers; and
- Applicable to a number of products (specific and aggregate stop loss, HMO reinsurance, provider excess, carrier medical excess reinsurance).

---

<sup>1</sup> The American Academy of Actuaries is a 16,000-member professional association whose mission is to serve the public on behalf of the U.S. actuarial profession. The Academy assists public policymakers on all levels by providing leadership, objective expertise, and actuarial advice on risk and financial security issues. The Academy also sets qualification, practice, and professionalism standards for actuaries in the United States.

<sup>2</sup> The Society of Actuaries is a 20,000-member non-profit educational, research, and professional organization dedicated to serving the public and the actuarial profession.

To support this effort, we will need the following experience information on your medical excess of loss business from 1998 to 2008:

- Product Type
- Calendar Year
- Total Gross Premium
- Total Incurred Claims
- Gross Loss Ratio
- Expected Loss Ratio
- Target Profit Margin

In order to normalize the data among companies, we are also requesting your expected loss ratio and target profit margins. If you feel these measures have changed significantly for your company over the data request period, please note that in your response. If not, you do not need to include figures for all years.

An electronic version of this request will be emailed to you and include a formatted Excel workbook as an attachment. Please submit your company's experience by completing the blue cells in the workbook and sending it to Barbara Scott at [bscott@soa.org](mailto:bscott@soa.org). We ask that you provide your experience data no later than November 20, 2009. If for any reason you do not receive an electronic copy of this data request, please contact Ms. Scott, at 847-706-3592, as soon as possible.

As stated above, the Work Group will be looking at aggregated data compiled by the SOA. Please do not send any information which would be of a business proprietary or confidential financial data to the Academy of Actuaries or to the Work Group members..

We thank you in advance for your help with this important project.

Sincerely,

Eric Smithback  
Chair, Stop-Loss Work Group  
American Academy of Actuaries

Steven C. Siegel  
Research Actuary  
Society of Actuaries

# Experience Data Request Instructions

## **Product Type**

Please submit experience for your company for all of the following products. We request that you submit the experience separately on the tabs provided for each product. Please use the following product

Specific SL = Specific stop loss (including aggregating specific)

Aggregate SL = Aggregate stop loss

HMO Re = specific reinsurance of an HMO's commercial, Medicare, Medicaid or Point of Service

PXS = specific excess written on Providers including IPAs, hospitals, clinics

MXS = specific reinsurance of an insurance company's medical business (first dollar or self-insured)

**Please do not include quota share or excess reinsurance written on Stop Loss business.**

## **Calendar Year**

Please submit experience information for each calendar year from 1998 to 2008. We request that you submit the experience separately for each calendar year.

## **Total Gross Premium**

This is the gross premium revenue, before ceded reinsurance and including commissions.

## **Total Incurred Claims:**

This is the gross incurred claims result, before ceded reinsurance. It is the sum of paid claims, change in reported claims reserves and change in IBNR (incurred but not reported reserves) for each Product Type

Please report the data as reported in your annual statement and do not adjust for any anomalies in the

## **Gross Loss Ratio**

This is equal to Incurred Claims / Gross Premium. This will calculate automatically.

## **Expected Loss Ratio**

This is equal to your current pricing target loss ratio for each Product Type listed. This will be used to normalize your experience results with other carriers' results. You only need to provide this for the most recent year unless your target loss ratio has changed significantly over the experience period.

## **Target Profit Margin**

This is equal to your current target profit margin for each Product Type listed, and will also be used to normalize your results with those from other carriers. We are defining Profit Margin to be: 100% less Expected Loss Ratio less Operating Expense Ratio. Again, you only need to provide this for the most recent year unless it has changed significantly over the course of the experience period.

Specific SL

Due Date to SOA- 11/20/09

Company Name

**Product Type**      **Specific Stop Loss (including aggregating specific)**

Calendar Year	1998	1999	#N/A								
Total Gross Premium	X	X	X	X	X	X	X	X	X	X	X
Total Incurred Claims	X	X	X	X	X	X	X	X	X	X	X
Gross Loss Ratio	#VALUE!	#N/A									
Expected Loss Ratio	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	X
Target Profit Margin	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	#N/A	X

Note -

- 1) Enter the Expected Loss Ratio and Target Profit Margin in the 2008 year. They will be carried into earlier years.
- 2) If the Expected Loss Ratio or Target Profit Margin changed significantly and one value is not appropriate for the entire study period, enter the updated Expected Loss Ratio and Target Profit Margin in the column for the year prior to the change. The updated values will be carried into earlier years.

Aggregate SL

**Product Type**      **Aggregate Stop Loss**

Calendar Year	1998	1999	#N/A									
Total Gross Premium	X	X	X	X	X	X	X	X	X	X	X	X
Total Incurred Claims	X	X	X	X	X	X	X	X	X	X	X	X
Gross Loss Ratio	#VALUE!	#N/A										

Expected Loss Ratio	#N/A	X	X									
Target Profit Margin	#N/A	X										

Note -

- 1) Enter the Expected Loss Ratio and Target Profit Margin in the 2008 year. They will be carried into earlier years.
- 2) If the Expected Loss Ratio or Target Profit Margin changed significantly and one value is not appropriate for the entire study period, enter the updated Expected Loss Ratio and Target Profit Margin in the column for the year prior to the change. The updated values will be carried into earlier years.

**Product Type** HMO Reinsurance - Specific reinsurance of an HMO's commercial, Medicare, Medicaid, or Point of Service products

Calendar Year	1998	1999	#N/A									
Total Gross Premium	X	X	X	X	X	X	X	X	X	X	X	X
Total Incurred Claims	X	X	X	X	X	X	X	X	X	X	X	X
Gross Loss Ratio	#VALUE!	#N/A										

Expected Loss Ratio	#N/A	X	X									
Target Profit Margin	#N/A	X										

Note -

- 1) Enter the Expected Loss Ratio and Target Profit Margin in the 2008 year. They will be carried into earlier years.
- 2) If the Expected Loss Ratio or Target Profit Margin changed significantly and one value is not appropriate for the entire study period, enter the updated Expected Loss Ratio and Target Profit Margin in the column for the year prior to the change. The updated values will be carried into earlier years.

**Product Type**      **Provider Excess Reinsurance - specific excess written on providers, including IPAs, hospitals, and clinics**

Calendar Year	1998	1999	#N/A									
Total Gross Premium	X	X	X	X	X	X	X	X	X	X	X	X
Total Incurred Claims	X	X	X	X	X	X	X	X	X	X	X	X
Gross Loss Ratio	#VALUE!	#N/A										

Expected Loss Ratio	#N/A	X	X									
Target Profit Margin	#N/A	X										

Note -

- 1) Enter the Expected Loss Ratio and Target Profit Margin in the 2008 year. They will be carried into earlier years.
- 2) If the Expected Loss Ratio or Target Profit Margin changed significantly and one value is not appropriate for the entire study period, enter the updated Expected Loss Ratio and Target Profit Margin in the column for the year prior to the change. The updated values will be carried into earlier years.

**Product Type**      **Medical Excess Reinsurance - specific reinsurance of an insurance company's medical business (first dollar or self-insured)**  
*Please do not include quota share or excess reinsurance written on Stop Loss business*

Calendar Year	1998	1999	#N/A									
Total Gross Premium	X	X	X	X	X	X	X	X	X	X	X	X
Total Incurred Claims	X	X	X	X	X	X	X	X	X	X	X	X
Gross Loss Ratio	#VALUE!	#N/A										

Expected Loss Ratio	#N/A	X	X									
Target Profit Margin	#N/A	X										

Note -

- 1) Enter the Expected Loss Ratio and Target Profit Margin in the 2008 year. They will be carried into earlier years.
- 2) If the Expected Loss Ratio or Target Profit Margin changed significantly and one value is not appropriate for the entire study period, enter the updated Expected Loss Ratio and Target Profit Margin in the column for the year prior to the change. The updated values will be carried into earlier years.

Stop Loss  
Aggregate - All Data  
November 2010

Year	Number of Respondents	Ranges of Submissions														Sum of Premium	Sum of Claims	Sum of Expected Claims			
		Gross Premium		Actual Loss Ratio		Weighted Avg. Actual LR		Expected Loss Ratio		Weighted Avg. Expected LR		Target Profit		Actual to Expected Loss Ratio							
1998	8	\$ 1,373,344.94	-	\$ 133,367,699.00	68.9%	-	104.4%	85.7%	65%	-	84%	68.5%	4%	-	19%	92%	-	154%	\$ 252,069,495.45	\$ 215,946,018.62	\$ 172,546,456.40
1999	11	\$ 1,435,780.42	-	\$ 142,936,602.00	40.4%	-	133.6%	87.5%	60%	-	84%	69.0%	4%	-	19%	61%	-	199%	\$ 408,254,985.07	\$ 357,389,037.52	\$ 281,780,349.08
2000	11	\$ 2,198,505.40	-	\$ 250,083,413.00	26.7%	-	193.3%	76.0%	60%	-	84%	68.8%	4%	-	19%	44%	-	289%	\$ 621,714,726.37	\$ 472,511,493.80	\$ 428,047,843.93
2001	13	\$ 882,806.70	-	\$ 280,520,248.00	25.3%	-	85.5%	68.6%	60%	-	84%	68.8%	4%	-	19%	42%	-	121%	\$ 1,145,595,405.85	\$ 786,277,109.74	\$ 787,761,856.51
2002	15	\$ 2,354,875.31	-	\$ 385,658,894.00	24.0%	-	89.4%	66.4%	60%	-	84%	68.2%	4%	-	19%	40%	-	133%	\$ 1,767,440,632.84	\$ 1,174,036,632.84	\$ 1,205,286,987.15
2003	16	\$ 262,800.00	-	\$ 448,826,803.00	0.0%	-	88.4%	58.2%	60%	-	84%	68.0%	4%	-	18%	0%	-	133%	\$ 1,840,936,809.65	\$ 1,072,341,547.55	\$ 1,252,021,528.42
2004	16	\$ 863,900.10	-	\$ 402,777,027.00	0.0%	-	82.9%	60.9%	60%	-	84%	68.7%	4%	-	18%	0%	-	112%	\$ 1,896,076,228.96	\$ 1,154,157,619.02	\$ 1,302,736,121.64
2005	16	\$ 2,416,208.00	-	\$ 377,687,833.00	50.6%	-	81.8%	65.4%	60%	-	84%	69.1%	4%	-	18%	84%	-	115%	\$ 1,983,044,433.02	\$ 1,296,945,525.31	\$ 1,371,104,664.47
2006	17	\$ 2,021,904.00	-	\$ 346,820,482.00	40.9%	-	93.5%	68.7%	60%	-	84%	69.3%	4%	-	18%	55%	-	134%	\$ 2,116,086,607.56	\$ 1,454,682,614.01	\$ 1,466,829,963.08
2007	17	\$ 1,999,935.00	-	\$ 368,337,553.26	20.3%	-	101.3%	68.7%	60%	-	84%	69.3%	4%	-	18%	30%	-	127%	\$ 2,238,448,955.91	\$ 1,537,201,032.42	\$ 1,551,725,038.75
2008	18	\$ 2,297,949.00	-	\$ 412,150,919.00	35.9%	-	84.5%	68.2%	65%	-	84%	69.6%	3%	-	18%	48%	-	113%	\$ 2,419,160,817.07	\$ 1,650,487,138.82	\$ 1,684,767,723.52

**Note:** All companies, on average, have more than \$1 million in annual premium.

Stop Loss  
Aggregate - Specific and Aggregate Only  
November 2010

Year	Number of Respondents	Ranges of Submissions														Sum of Premium	Sum of Claims	Sum of Expected Claims
		Gross Premium		Actual Loss Ratio		Weighted Avg. Actual LR		Expected Loss Ratio		Weighted Avg. Expected LR		Target Profit		Actual to Expected Loss Ratio				
1998	7	\$ 2,031,484.37	- \$ 133,367,699.00	71.7%	- 104.4%	84.9%	65%	- 84%	67.9%	4%	- 19%	102%	- 154%	\$ 233,468,063.56	\$ 198,204,706.21	\$ 158,632,507.49		
1999	10	\$ 4,577,794.59	- \$ 142,936,602.00	40.4%	- 133.6%	86.5%	60%	- 84%	68.8%	4%	- 19%	61%	- 199%	\$ 388,694,573.82	\$ 336,325,680.81	\$ 267,299,664.88		
2000	10	\$ 7,179,718.34	- \$ 250,083,413.00	26.7%	- 193.3%	76.3%	60%	- 84%	68.8%	4%	- 19%	44%	- 289%	\$ 608,728,669.84	\$ 464,284,419.32	\$ 418,640,173.67		
2001	12	\$ 2,538,773.08	- \$ 280,520,248.00	13.2%	- 85.6%	68.8%	60%	- 84%	68.6%	4%	- 19%	18%	- 122%	\$ 1,113,042,426.35	\$ 766,017,790.59	\$ 763,285,369.52		
2002	14	\$ 2,418,701.08	- \$ 385,658,894.00	24.0%	- 89.4%	66.3%	60%	- 84%	68.1%	4%	- 19%	40%	- 133%	\$ 1,731,934,691.02	\$ 1,147,422,934.38	\$ 1,178,911,724.79		
2003	15	\$ 1,207,720.91	- \$ 448,826,803.00	32.1%	- 88.4%	58.3%	60%	- 84%	67.9%	4%	- 18%	47%	- 133%	\$ 1,801,507,635.42	\$ 1,050,255,533.83	\$ 1,222,507,986.96		
2004	14	\$ 3,877,124.87	- \$ 402,777,027.00	54.2%	- 99.4%	61.0%	60%	- 84%	68.6%	4%	- 18%	75%	- 133%	\$ 1,854,042,991.69	\$ 1,130,897,293.77	\$ 1,271,294,885.69		
2005	15	\$ 2,416,208.00	- \$ 377,687,833.00	50.6%	- 81.8%	65.2%	60%	- 84%	69.0%	4%	- 18%	82%	- 115%	\$ 1,936,554,069.80	\$ 1,263,074,038.88	\$ 1,336,365,274.50		
2006	16	\$ 2,021,904.00	- \$ 346,820,482.00	40.9%	- 93.5%	68.6%	60%	- 84%	69.2%	4%	- 18%	55%	- 134%	\$ 2,058,172,700.56	\$ 1,412,646,874.73	\$ 1,423,564,799.64		
2007	16	\$ 1,999,935.00	- \$ 368,337,553.26	20.3%	- 101.3%	68.1%	60%	- 84%	69.1%	4%	- 18%	30%	- 127%	\$ 2,157,370,632.46	\$ 1,468,645,478.54	\$ 1,491,120,698.11		
2008	17	\$ 2,297,949.00	- \$ 412,150,919.00	59.7%	- 81.9%	67.8%	65%	- 84%	69.5%	3%	- 18%	83%	- 107%	\$ 2,333,244,448.58	\$ 1,581,810,116.66	\$ 1,620,539,484.13		

**Note:** All companies have more than \$1 million in annual premium.

Stop Loss  
Aggregate - Over One Million  
November 2010

Year	Number of Respondents	Ranges of Submissions														Sum of Premium	Sum of Claims	Sum of Expected Claims			
		Gross Premium		Actual Loss Ratio		Weighted Avg. Actual LR		Expected Loss Ratio		Weighted Avg. Expected LR		Target Profit		Actual to Expected Loss Ratio							
1998	8	\$ 1,373,344.94	-	\$ 133,367,699.00	68.9%	-	104.4%	85.7%	65%	-	84%	68.5%	4%	-	19%	92%	-	154%	\$ 252,069,495.45	\$ 215,946,018.62	\$ 172,546,456.40
1999	11	\$ 1,435,780.42	-	\$ 142,936,602.00	40.4%	-	133.6%	87.5%	60%	-	84%	69.0%	4%	-	19%	61%	-	199%	\$ 408,254,985.07	\$ 357,389,037.52	\$ 281,780,349.08
2000	11	\$ 2,198,505.40	-	\$ 250,083,413.00	26.7%	-	193.3%	76.0%	60%	-	84%	68.8%	4%	-	19%	44%	-	289%	\$ 621,714,726.37	\$ 472,511,493.80	\$ 428,047,843.93
2001	12	\$ 9,259,000.00	-	\$ 280,520,248.00	25.3%	-	85.5%	68.6%	60%	-	84%	68.8%	4%	-	19%	42%	-	121%	\$ 1,144,712,599.15	\$ 785,711,841.36	\$ 787,761,856.51
2002	15	\$ 2,354,875.31	-	\$ 385,658,894.00	24.0%	-	89.4%	66.4%	60%	-	84%	68.2%	4%	-	19%	40%	-	133%	\$ 1,767,440,632.84	\$ 1,174,036,632.84	\$ 1,205,286,987.15
2003	15	\$ 4,993,792.40	-	\$ 448,826,803.00	32.1%	-	88.4%	58.3%	60%	-	84%	68.0%	4%	-	18%	49%	-	133%	\$ 1,840,674,009.65	\$ 1,072,341,547.55	\$ 1,252,021,528.42
2004	15	\$ 3,903,878.05	-	\$ 402,777,027.00	54.2%	-	82.9%	60.9%	60%	-	84%	68.7%	4%	-	18%	75%	-	112%	\$ 1,895,212,328.86	\$ 1,154,157,619.02	\$ 1,302,736,121.64
2005	16	\$ 2,416,208.00	-	\$ 377,687,833.00	50.6%	-	81.8%	65.4%	60%	-	84%	69.1%	4%	-	18%	84%	-	115%	\$ 1,983,044,433.02	\$ 1,296,945,525.31	\$ 1,371,104,664.47
2006	17	\$ 2,021,904.00	-	\$ 346,820,482.00	40.9%	-	93.5%	68.7%	60%	-	84%	69.3%	4%	-	18%	55%	-	134%	\$ 2,116,086,607.56	\$ 1,454,682,614.01	\$ 1,466,829,963.08
2007	17	\$ 1,999,935.00	-	\$ 368,337,553.26	20.3%	-	101.3%	68.7%	60%	-	84%	69.3%	4%	-	18%	30%	-	127%	\$ 2,238,448,955.91	\$ 1,537,201,032.42	\$ 1,551,725,038.75
2008	18	\$ 2,297,949.00	-	\$ 412,150,919.00	35.9%	-	84.5%	68.2%	65%	-	84%	69.6%	3%	-	18%	48%	-	113%	\$ 2,419,160,817.07	\$ 1,650,487,138.82	\$ 1,684,767,723.52

Note: All companies, on average, have more than \$1 million in annual premium.

Loss Ratios By Year

1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
68.9%	40.4%	26.7%	25.3%	24.0%	0.0%	0.0%	50.6%	40.9%	20.3%	35.9%
71.7%	41.5%	59.4%	33.5%	36.8%	32.1%	54.2%	58.5%	52.4%	55.2%	59.7%
78.8%	70.7%	61.0%	54.2%	49.6%	39.4%	56.0%	60.6%	62.1%	56.5%	62.6%
80.4%	76.8%	62.6%	57.9%	54.0%	43.8%	57.2%	61.7%	64.0%	57.9%	62.6%
86.6%	78.4%	65.8%	64.0%	57.5%	46.1%	57.3%	63.1%	64.8%	60.1%	63.0%
96.6%	81.3%	67.3%	65.1%	61.8%	47.5%	58.9%	63.2%	65.0%	67.7%	65.6%
103.5%	90.4%	71.8%	66.5%	62.0%	52.6%	59.4%	64.8%	67.6%	71.0%	67.3%
104.4%	90.9%	77.0%	67.4%	65.7%	53.3%	59.8%	65.0%	68.0%	73.4%	67.4%
	101.7%	93.2%	69.9%	68.2%	55.0%	62.2%	66.0%	69.4%	74.9%	69.4%
	104.9%	117.9%	72.0%	68.7%	55.6%	63.0%	66.7%	73.3%	76.9%	70.0%
	133.6%	193.3%	72.5%	68.8%	61.5%	64.2%	69.4%	73.5%	78.0%	71.3%
			75.8%	74.8%	61.6%	64.4%	70.2%	74.5%	78.1%	72.4%
			85.5%	79.8%	65.5%	74.3%	71.2%	80.5%	79.2%	74.5%
				81.6%	69.0%	74.9%	74.9%	83.4%	79.8%	77.2%
				89.4%	78.9%	79.9%	80.8%	83.7%	82.1%	79.9%
					88.4%	82.9%	81.8%	85.1%	90.6%	81.4%
								93.5%	101.3%	81.9%
										84.5%

low LR
wtd avg LR +/- 10%
high LR

85.7%    87.5%    76.0%    68.6%    66.4%    58.2%    60.9%    65.4%    68.7%    68.7%    68.2% wtd avg LR

2	4	5	4	4	5	1	1	2	4	1	33	21.0% low LR
3	4	3	8	8	8	11	13	10	8	13	89	56.7% wtd avg LR +/- 10%
2	3	3	1	3	3	4	2	5	5	4	35	22.3% high LR
7	11	11	13	15	16	16	16	17	17	18	157	100.0%

**American Academy of Actuaries**  
**Stop Loss Project**

**By**

**Joseph Marker, FCAS, MAAA, Instructor**

**University of Michigan**

# American Academy Stop-Loss Project Progress

## 1. Data cleaning and data issue

1.1 Manually checked Company 1- Company 20: **(4)** specific + aggregate **is the total**.

(Note: Mostly, (4) = (1) + (2) and (3) is blank. If (3) is not blank, (4)=(1)+(2)+(3) or (4)=(3)=(1)+(2))

1.2 Filtered the original data set using category “Product Type” – specific + aggregate

1.3 Created a new excel file called “stop loss data” after 1.2 filter

1.4 Read excel file “stop loss data” to the statistical language "R" for further analysis

## 2. Model 1

2.1 Hypothesis: the variation in loss ratios across the year is related to the company size (based on premium). If the premium is higher, the variation in loss ratios is smaller, and vice versa.

### 2.2 Data preparation

2.2.1 Recode “Company 1” to “1” in the Company column

2.2.2 Calculate averages of premiums and standard deviations of loss ratios across year for each company. Below is the summary

	V1	avg.prem	std.lossratio
1	1	4204013.81	0.23644866
2	2	157992750.69	0.08511069
3	3	249542857.14	0.05000821
4	4	144399180.43	0.16854331
5	5	25533.27	0.45832895
6	6	3968098.00	0.15527348
7	7	19148100.00	0.14760078
8	8	30978155.82	0.12690435
9	9	296531458.30	0.10432633
10	10	3492973.68	0.30036596
11	11	25962152.00	NA
12	12	188028571.43	0.14477714
13	13	2106596.00	0.18897267
14	14	185567717.10	0.07644536
15	15	NaN	NA
16	16	67628940.77	0.13751750
17	17	321273424.27	0.11781408
18	18	46808127.73	0.06318744
19	19	NaN	NA
20	20	40079056.81	0.12453758

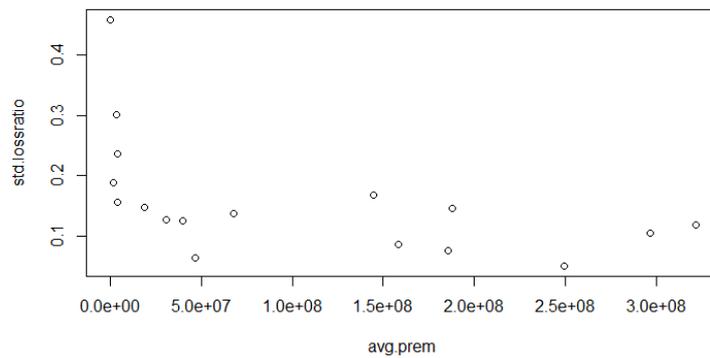
Note: (1) Company 11 has product only in 2008 so it’s impossible to calculate the variation of loss ratios (2) Company 15 and Company 19 don’t have data for product “specific + aggregate”.

**2.2.3 Delete Company 11,15 and 19. In sum, we have 17 observations.**

For each observation, the response is variation in loss ratios and the predictor is average premium. Below is the new summary.

	V1	avg.prem	std.lossratio
1	1	4204013.81	0.23644866
2	2	157992750.69	0.08511069
3	3	249542857.14	0.05000821
4	4	144399180.43	0.16854331
5	5	25533.27	0.45832895
6	6	3968098.00	0.15527348
7	7	19148100.00	0.14760078
8	8	30978155.82	0.12690435
9	9	296531458.30	0.10432633
10	10	3492973.68	0.30036596
11	12	188028571.43	0.14477714
12	13	2106596.00	0.18897267
13	14	185567717.10	0.07644536
14	16	67628940.77	0.13751750
15	17	321273424.27	0.11781408
16	18	46808127.73	0.06318744
17	20	40079056.81	0.12453758

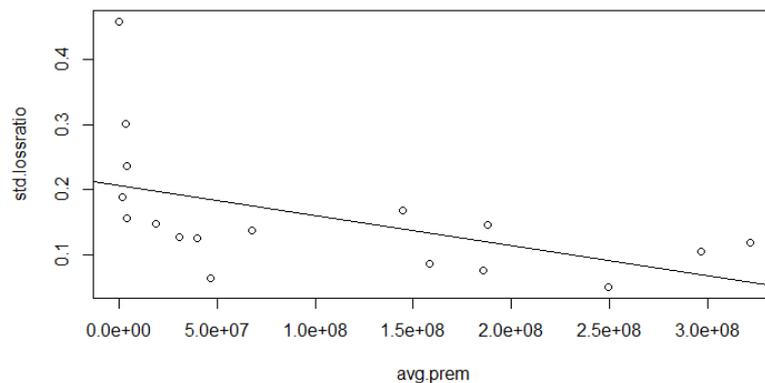
2.3 Scatterplot of variations in loss ratios and premium average for 17 observations.



There is a trend that when company size is small, the variation in loss ratios is high.

2.4 simple linear regression

2.4.1 Consider average premium as a numeric predictor. Below is the scatterplot and line of fitted value.



---

Below is output of the linear model

```
Call:
lm(formula = std.lossratio ~ avg.prem, data = model1)

Residuals:
    Min       1Q   Median       3Q      Max
-0.12114 -0.04888 -0.03716  0.03240  0.25234

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.060e-01  2.958e-02   6.965 4.54e-06 ***
avg.prem    -4.630e-10  1.981e-10  -2.337  0.0337 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.08776 on 15 degrees of freedom
Multiple R-squared: 0.2669, Adjusted R-squared: 0.218
F-statistic: 5.461 on 1 and 15 DF, p-value: 0.03372
```

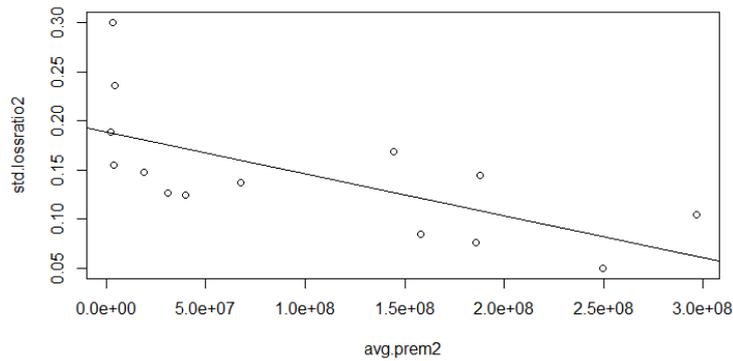
Analysis:

- p value is 0.0337, which means reject the null hypothesis. Thus, **average premium is significant** regarding the variation in loss ratios.
- R Squared means the coefficient of determination. It provides a measure of how well observed outcomes are fit by the model, as the proportion of total variation of outcomes explained by the model. The larger, the better. However, the multiple R-squared is 0.2669, which is not good enough.
- I used the standard deviation of loss ratios instead of standard deviation of incurred loss as the independent variable. The reason is the ratio represents better the variation in the loss. The Incurred loss will depend more on the premium itself.

2.4.2 If we delete the five smallest premiums (Company 1, 5, 6, 10 and 12), the p value of the regression result will become 0.391, which indicates that the relationship between variations in loss ratios and premiums is not statistically significant. Therefore, it might not be a good idea to simply delete the five smallest premiums.

2.4.3 However, there is a concept in statistics called Influential Observation. An Influential point is one whose removal from the dataset would cause a large change in the fit. Cook's distance is a popular measure for Influential Point. From Wikipedia, Cook's distance measures the effect of deleting a given observation. Data points with large residuals (outliers) and/or high leverage may distort the outcome and accuracy of a regression. Points with a large Cook's distance are considered to merit closer examination in the analysis. Company 5, 17 and 18 have the three largest cook's distances. After deleting data of company 5, 17 and 18, we ran regression analysis on the rest 14 observations. The p value of the result became extremely small (0.00963), which indicated statistically significant between variations in loss ratios and company sizes. Also, the multiple R-Squared is 0.4407(0.2669), which is larger than the R-Squared of regression result of original 17 observations. R-Squared measures how well the regression model fits the data. Thus, deleting company 5, 17 and 18 might be reasonable for the simple linear regression. Below is the screenshot of fitted line of

regression analysis of 14 observations.



3. Model 1 Revised: consider average premium above as a categorical predictor

3.1 data preparation

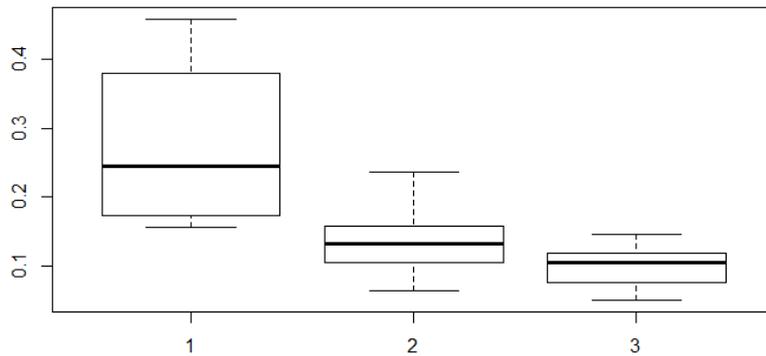
0%                      25%                      50%                      75%                      100%  
 25533.27    4204013.81    46808127.73    185567717.10    321273424.27

I consider companies whose premium average is less than 25% quantile (4204013.81) small companies, companies whose premium average are less than 75% quantile (185567717.10) but bigger than 25% quantile (4204013.81) medium companies, companies whose premium average are bigger than 75% quantile (185567717.10) big companies. I then ranked small companies 1, medium companies 2, and big companies 3. Below is the new summary.

	V1	avg.prem	rank.avg.prem	std.lossratio
1	1	4204013.81	2	0.23644866
2	2	157992750.69	2	0.08511869
3	3	249542857.14	3	0.05000821
4	4	144399180.43	2	0.16854331
5	5	25533.27	1	0.45832895
6	6	3968098.00	1	0.15527348
7	7	19148100.00	2	0.14760078
8	8	30978155.82	2	0.12690435
9	9	296531458.30	3	0.10432633
10	10	3492973.68	1	0.30036596
11	12	188028571.43	3	0.14477714
12	13	2106596.00	1	0.18897267
13	14	185567717.10	3	0.07644536
14	16	67628940.77	2	0.13751750
15	17	321273424.27	3	0.11781408
16	18	46808127.73	2	0.06318744
17	20	40079056.81	2	0.12453758

Note: rank 1, 2 and 3 are treated as explanatory variables instead of numeric.

Below is the boxplot of these three ranked kinds of companies and their corresponding variations in loss ratios.



It is obvious that rank 1 company (small) have high variations in loss ratios.

### 3.2 regression

Treated company ranking 1, 2 and 3 as categorical predictors and standard deviations of loss ratios as response. Below is the output of the new model.

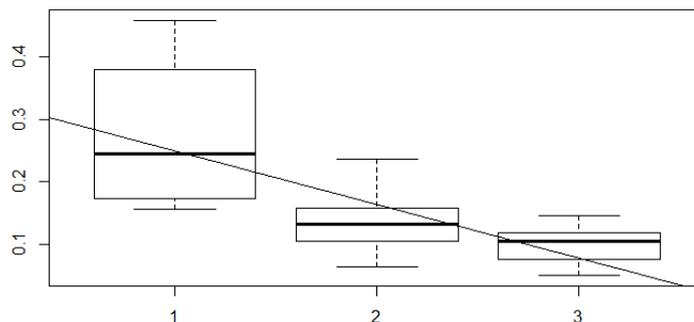
```
Call:
lm(formula = std.lossratio ~ rank.avg.prem, data = model1.new)

Residuals:
    Min       1Q   Median       3Q      Max
-0.09987 -0.03852 -0.01546  0.04060  0.20942

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.33475    0.05708   5.865 3.11e-05 ***
rank.avg.prem -0.08585    0.02615  -3.283  0.00503 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07819 on 15 degrees of freedom
Multiple R-squared:  0.4181, Adjusted R-squared:  0.3793
F-statistic: 10.78 on 1 and 15 DF, p-value: 0.005031
```

P-value is extremely small so that the ranking is significant. Also the R-squared is 0.4181, which is larger than the R-squared in Section 2 (where we treated company premium as a numeric variable). Thus, model with ranking 1, 2 and 3 fitted better.



For further analysis, we can change the standard of ranking.

## 4. Model 2

4.1 Hypothesis: the frequency of an "adverse result" is related to the company size. If the premium is higher, the frequency of an adverse result is smaller, and vice versa.

4.2 Adverse result:

Definition: An adverse result is one where

$$\text{premium} - \text{expected expenses} - \text{actual losses} < 0 \quad (*)$$

We can say that Target margin = Premium - expected loss - expected expense. This implies that Premium - Expected expense = expected loss + target margin

Plugging this into (\*), we get an alternate criterion for an adverse result:

$$\text{Actual Loss} > \text{Expected loss} + \text{Target margin}$$

4.3 Data preparation:

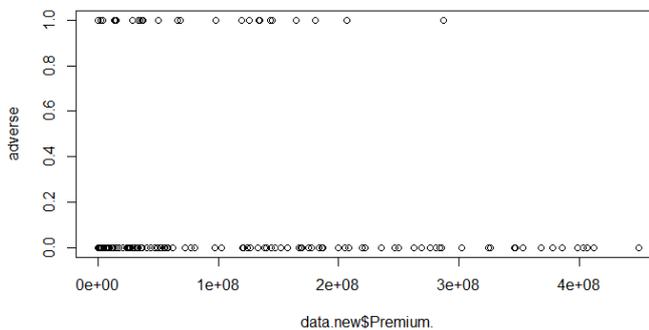
We treat each company each year as an observation. **There are 146 observations in total.**

The response variable the indicator for an adverse result. Its value is 1 for an adverse result and 0 otherwise. The predictor is the premium for that company that year.

Below is a screen shot of part of the new data summary. First column is the company code. Second is year. Third is premium. Fourth is adverse result (1 represents an adverse result).

	V1	V2	V3	adverse
1	1	2003	1207721	0
2	1	2004	3877125	1
3	1	2005	3004604	0
4	1	2006	4099387	1
5	1	2007	5061281	0
6	1	2008	7973966	0
7	2	2001	169715000	0
8	2	2002	185742001	0
9	2	2003	169017001	0
10	2	2004	176983001	0
11	2	2005	157607001	0
12	2	2006	139831001	0
13	2	2007	144695001	1
14	2	2008	120352000	0
15	3	2002	126200000	0

Below is the scatter plot of adverse result versus premium



**There are 31 adverse results.** We can see that if the premium is large enough (above 3 e+08), there is no adverse result.

#### 4.4 logistic regression treating premium as a numeric variable

Below is the output of logistic regression.

```
Call:
glm(formula = adverse ~ data.new$Premium., family = "binomial",
     data = model2)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.8277 -0.7873 -0.6242 -0.3732  2.1443

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -8.950e-01  2.654e-01  -3.373  0.000745 ***
data.new$Premium. -4.521e-09  2.199e-09  -2.056  0.039821 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 150.97  on 145  degrees of freedom
Residual deviance: 145.86  on 144  degrees of freedom
AIC: 149.86

Number of Fisher Scoring iterations: 4
```

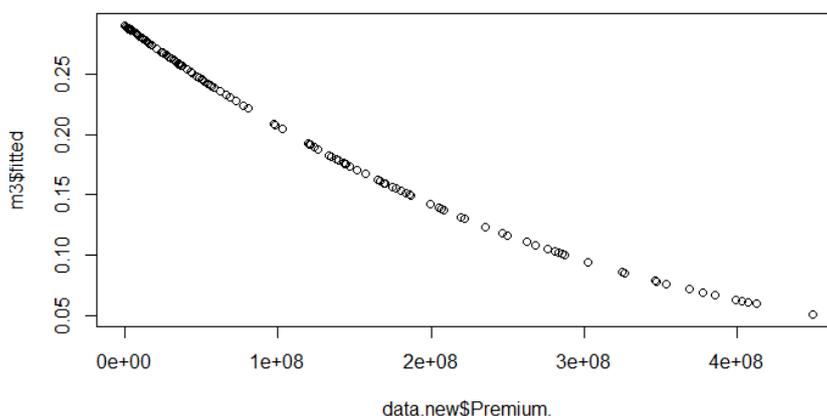
P-value is 0.0398 < 0.05. So Premium is considered significant in predicting adverse result. The F test yields the same result.

#### 4.5 logistic regression analysis

Logistic regression can produce predicted probabilities of adverse results for different premiums. Below is the formula of predicted probabilities of adverse results for different premiums.

$$\Pr(\text{adverse result}) = \frac{e^{-0.895 - 4.5205 \times 10^{-9} \times \text{Premium}}}{1 + e^{-0.895 - 4.5205 \times 10^{-9} \times \text{Premium}}}$$

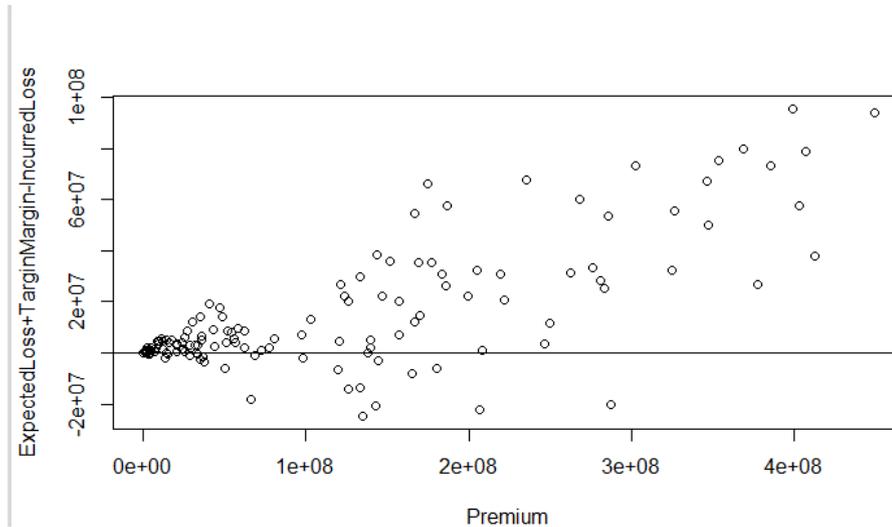
Below is the screenshot of predicted probabilities of adverse results versus different premiums. From logistics regression, the probability of adverse result and Premium are negatively correlated and the relationship is statistically significant.



#### 5. Adverse Results versus Premium

Section 4 defined adverse result as an indicator for **Actual Loss > Expected loss + Target margin**. In this section, we use the numeric value of **(Expected Loss + Target Margin) – Actual Loss**. Note that when the difference is less than 0, there is an adverse result. We

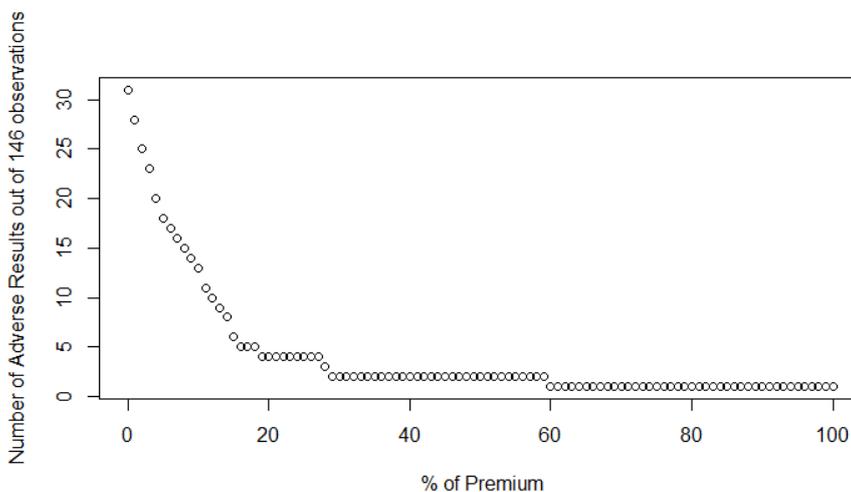
expected to see a positive correlation between (Expected Loss + Target Margin) – Actual Loss and Premium. Below is the screen shot of the scatter plot of the difference versus premium. If the dot is below  $y=0$ , it indicates an adverse result. We can see that when premium is larger, the difference of Expected loss + Target Margin and Actual Loss is larger.



Our goal we were given is to determine the percentage of premium we need to add to guarantee that there is no "adverse result". Here an adverse result means:

$$(\text{Expected Loss} + \text{Target Margin} + \text{Premium} * x \%) - \text{Actual Loss} > 0.$$

Below is the screen shot of Number of Adverse Results versus Percentage of Premium added.



Originally, there are 31 adverse results. We can see from 0% to 20%, the number of adverse results decreases quickly. For example, if we add 10% of premium, the number of adverse results out of 146 observations will become 13. Since there are few adverse results when Premium is larger than  $2E+08$ , we need 10% of Premium only if the premium is relatively small (say less than  $2E+08$ ).

## 6. Model 2 revised: consider premium above as a categorical predictor

### 6.1 data preparation

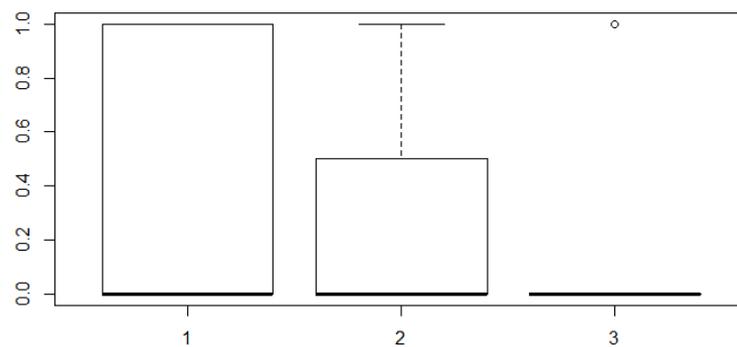
0%	25%	50%	75%	100%
4534	13962730	55361366	169540500	448826803

I consider companies (each year) whose premium are less than 25% quantile (13952730) small companies, companies whose premium average are less than 75% quantile (169540500) but bigger than 25% quantile (13962730) medium companies, companies whose premium average are bigger than 75% quantile (169540500) big companies. I then ranked small companies 1, medium companies 2, big companies 3. Below is a screen shot of part of the new summary.

	V1	V2	rank.prem	adverse
1	1	1207721	1	0
2	2	3877125	1	1
3	3	3004604	1	0
4	4	4099387	1	1
5	5	5061281	1	0
6	6	7973966	1	0
7	7	169715000	3	0
8	8	185742001	3	0
9	9	169017001	2	0
10	10	176983001	3	0
11	11	157607001	2	0
12	12	139831001	2	0
13	13	144695001	2	1
14	14	120352000	2	0
15	15	126200000	2	0

Note: rank 1,2 and 3 are treated as categorical variables instead of numeric.

Below is the boxplot of these three ranked kinds of companies and their corresponding adverse result.



We can see the differences between these three rankings.

## 6.2 regression

Treated company ranking 1, 2 and 3 as categorical predictors and standard deviations of loss ratios as response. Below is the output of the new model.

```

Call:
glm(formula = adverse ~ rank.prem, family = "binomial", data = model2.new)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.8694  -0.6761  -0.6761  -0.5180   2.0372

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -0.1967    0.5810  -0.338  0.7350
rank.prem    -0.5814    0.2962  -1.963  0.0497 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 150.97  on 145  degrees of freedom
Residual deviance: 146.96  on 144  degrees of freedom
AIC: 150.96

Number of Fisher Scoring iterations: 4

```

The rankings are significantly important for the adverse result.

## 7. Model 3

Note: This section describes another way to model "adverse result". Here we define an adverse result as one where the target margin is not achieved. In Model 2, we defined an adverse result as one where the company lost money. After consulting with the American Academy, we agreed that Model 2 is more relevant than Model 3. We include the Model 3 results for completeness.

7.1 Hypothesis: In model 2, we define adverse as premium – expected expenses – actual losses < 0, which means we actually lose money. Here we define adverse as **actual loss ratio > expected loss ratio**. The hypothesis is still that the frequency of an adverse result is related to the company size. If the premium is higher, the probability of an adverse result is smaller, and vice versa. The only difference with Model 2 is the definition of adverse result.

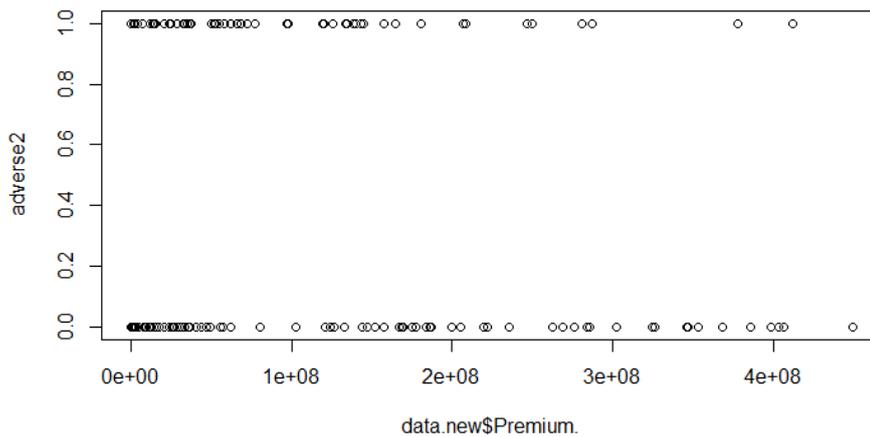
### 7.2 Data preparation:

We treat each company each year as an observation. **There are 146 observations in total**. The response is whether it got an adverse result. (**Adverse result coded 1**; otherwise coded 0). The predictor is the premium for that company that year. Below is a screen shot of part of the new data summary. First column is the company code. Second is year. Third is premium. Fourth is adverse result (1 represents there is adverse result).

	V1	V2	V3	adverse2
1	1	2003	1207721	0
2	1	2004	3877125	1
3	1	2005	3004604	0
4	1	2006	4099387	1
5	1	2007	5061281	0
6	1	2008	7973966	0
7	2	2001	169715000	0
8	2	2002	185742001	0
9	2	2003	169017001	0
10	2	2004	176983001	0
11	2	2005	157607001	0
12	2	2006	139831001	1
13	2	2007	144695001	1
14	2	2008	120352000	1
15	3	2002	126200000	0

There are totally 62 adverse results.

Below is the scatterplot of premium and adverse result



When the premium is high, the probability of adverse result is low.

### 7.3 Logistic regression

```
Call:
glm(formula = adverse2 ~ data.new$Premium., family = "binomial",
     data = mode14)
```

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.1900 -1.1315 -0.8024  1.2101  1.7480
```

```
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.959e-02  2.291e-01  0.129  0.8972
data.new$Premium. -3.185e-09  1.549e-09 -2.056  0.0398 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Dispersion parameter for binomial family taken to be 1)
```

```
Null deviance: 199.07 on 145 degrees of freedom
Residual deviance: 194.53 on 144 degrees of freedom
AIC: 198.53
```

```
Number of Fisher Scoring iterations: 4
```

The p value is  $0.0398 < 0.05$ . The premium (company size) can be considered significantly important to this definition of adverse result.

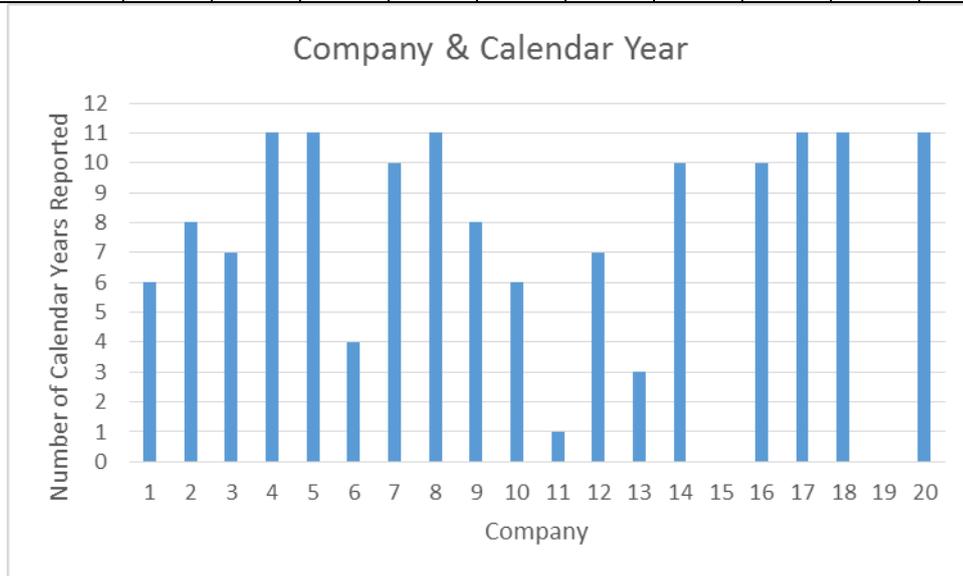
(1) Compare and conclusion

The result is very similar to model 2 (another definition of adverse result). Premium is a good indicator of adverse result.

8. Calendar Year

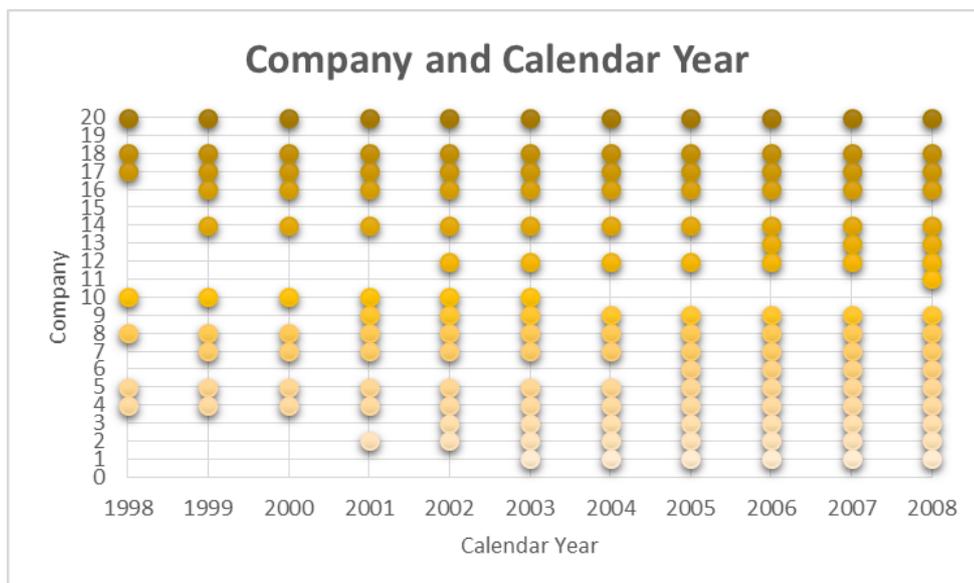
8.1 Number of years reported for each company

Company	1	2	3	4	5	6	7	8	9	10
Number of Years	6	8	7	11	11	4	10	11	8	6
Company	11	12	13	14	15	16	17	18	19	20
Number of Years	1	7	3	10	0	10	11	11	0	11



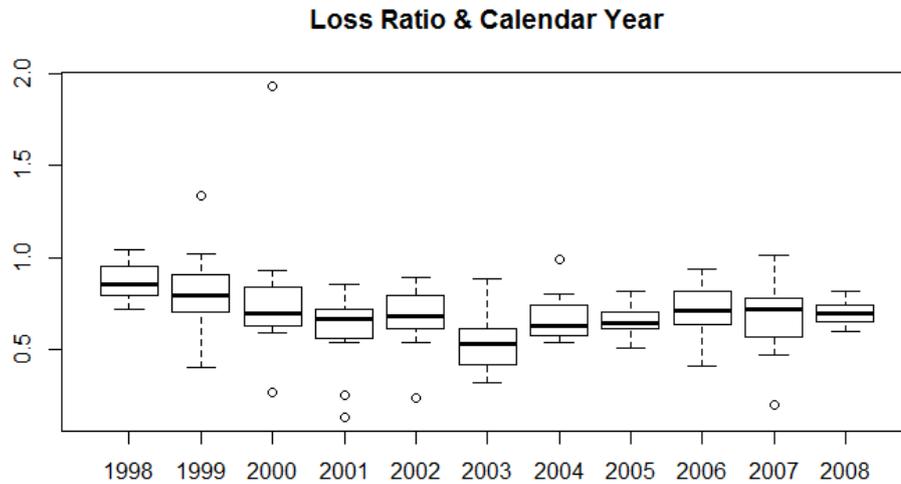
As the chart and graph show above, the number of calendar years reported for each company is quite different, from 0 to 11. Company 11 has only one year of data, while companies 15 and 19 have no data in the product line. Maybe consider deleting company 6 and 13.

8.2 Calendar Years reported for each company

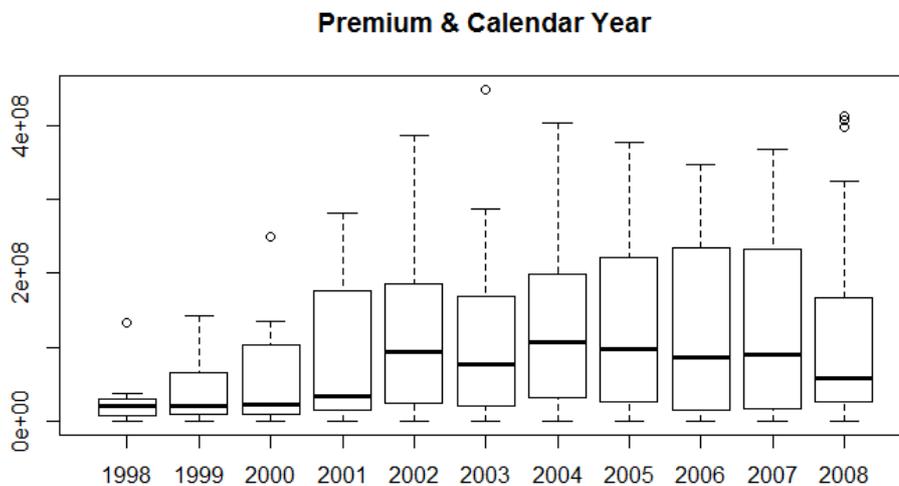


If one company has data in one year, there is a dot. It can be seen that **for the most recent years, more companies report data**. In 1998, only 7 companies out of 20 reported data.

### 8.3 Loss Ratio and Calendar Year & Premium and Calendar Year



The bold lines represent the mediums of loss ratios for each calendar year. **In 2003, the medium is the lowest.** And the box is apparently lower than other years'. 2003 is a special year.

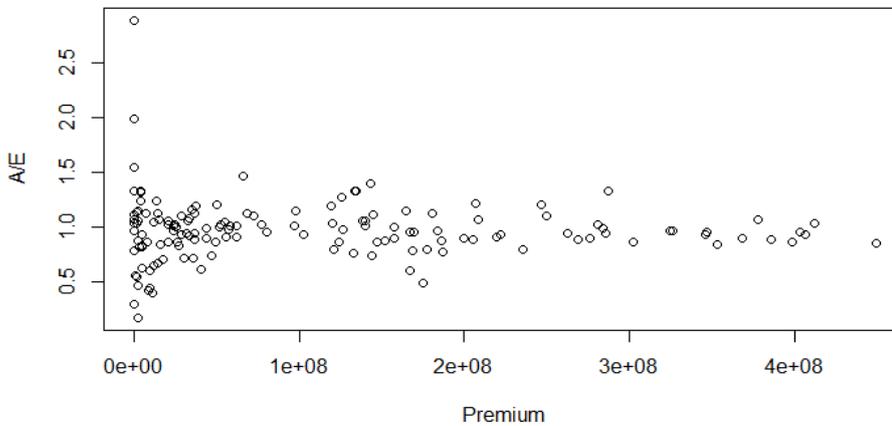


The premium remains constant except first 3 years. The companies grew up and remain the same level. I do not see that 2003 is a special year when it comes to premium.

However, we should keep in mind if we want to take calendar year into consideration, the first 3 years (1998, 1999 and 2000) have a less data. The less the data, the lower the data credibility.

9 Model 4

We are interested in the association between Actual Loss Ratio/Expected Loss Ratio (A/E) and company size (Premium).



Among 146 observations, there is no linear trend or other between Premium and A/E.

10. Additional Margin – Working sheet (“sliding”)

In section 5, we showed how requiring a company to hold an additional amount of its premium could reduce the probability of an "adverse result". In that section, we only considered additions that were a fixed per cent of premium. In this section, we allow this additional premium to be one percentage of the premium below a "turning point" and a different percentage for the portion of the premium above the "turning point".

10.1 Formula:

AdditionalMargin=

$$\begin{cases} \text{premium} * x\%, & \text{if premium} < \text{turning point} \\ \text{turning point} * x\% + (\text{premium} - \text{turning point}) * y\%, & \text{if premium} > \text{turning point} \end{cases}$$

Note:

- (1) We would like to make the probability of adverse result be less than 0.05. The total number of observations is 146. Thus we would like to have less than  $146 * 0.05 = 7.3$  adverse results, which means less or equal than 7 adverse results.
- (2)  $x > y$ . I chose x to be 20,25,30 and y to be 0,5,10
- (3) The median of premium is around 100,000,000. So I picked some turning points around 100,000,000, which were 10,000,000, 50,000,000, 100,000,000, 150,000,000 and 200,000,000.

10.2 The results appear in working sheet (“Sliding”)

A1:E147 is the data processed.

J2:M3 is the part we can play around. If we change Turning Point, x% and y%, the number of adverse results will change.

Turning Point	x%	y%	Number of Adverse Results
10000000	30.00	10.00	5

J7:U25 is the sample result of J2:M3.

I chose several combinations of turning point, x% and y%.

Turning Point	x%	y%	Number of Adverse Results
10,000,000	20	-	19
10,000,000	20	5	11
10,000,000	20	10	9
10,000,000	25	-	16
10,000,000	25	5	11
10,000,000	25	10	6
10,000,000	30	-	15
10,000,000	30	5	10
10,000,000	30	10	5
50,000,000	20	-	10
50,000,000	20	5	8
50,000,000	20	10	6
50,000,000	25	-	10
50,000,000	25	5	7
50,000,000	25	10	5
50,000,000	30	-	7
50,000,000	30	5	5
50,000,000	30	10	4
100,000,000	20	-	8
100,000,000	20	5	5
100,000,000	20	10	5
100,000,000	25	-	4
100,000,000	25	5	4
100,000,000	25	10	4
100,000,000	30	-	2
100,000,000	30	5	2
100,000,000	30	10	2
150,000,000	20	-	4
150,000,000	20	5	4
150,000,000	20	10	4
150,000,000	25	-	4
150,000,000	25	5	4
150,000,000	25	10	4
150,000,000	30	-	2
150,000,000	30	5	2
150,000,000	30	10	2
200,000,000	20	-	4
200,000,000	20	5	4
200,000,000	20	10	4
200,000,000	25	-	4
200,000,000	25	5	4
200,000,000	25	10	4
200,000,000	30	-	2
200,000,000	30	5	2
200,000,000	30	10	2

---

11. Raw data set of adverse result in working sheet ("Adverse Result")

There are total 31 adverse results.

Difference=expected margin + expected loss – Incurred loss

R code:

```
#Read data
library(xlsx)
getwd()
setwd("C:/Users/Daisy/Desktop")
data=read.xlsx("stop loss data.xlsx",sheetName="data")
data.new=data[!is.na(data$Premium.),]
data.new$Expected.loss=as.numeric(as.character(data.new$Expected.loss))
data.new$Incurred.Loss=as.numeric(as.character(data.new$Incurred.Loss))
data.new$Expected.margin=as.numeric(as.character(data.new$Expected.margin))
colnames(data)

#model 1 data preparation
avg.prem= sapply(1:20,function(x) mean(data$Premium.[data$Company==x],na.rm=T))
std.lossratio= sapply(1:20,function(x) sd(data$Loss.Ratio.[data$Company==x],na.rm=T))
model1=cbind(1:20,avg.prem,std.lossratio)
std.lossratio=std.lossratio[-c(11,15,19)]
avg.prem=avg.prem[-c(11,15,19)]
model1=cbind(c(1,2,3,4,5,6,7,8,9,10,12,13,14,16,17,18,20),avg.prem,std.lossratio)
model1=data.frame(model1)

#model1
plot(avg.prem,std.lossratio)
m1=lm(std.lossratio~avg.prem,data=model1)
summary(m1)
```

```

abline(m1)

#model1 new
quantile(avg.prem)
rank.avg.prem=integer(length(avg.prem))
for (i in 1:length(avg.prem))
{
  if (avg.prem[i]<quantile(avg.prem,0.25))
  {
    rank.avg.prem[i]=1
  }
  else if (avg.prem[i]<quantile(avg.prem,0.75))
  {
    rank.avg.prem[i]=2
  }
  else
    rank.avg.prem[i]=3
}
rank.avg.prem=as.factor(rank.avg.prem)
model1.new=cbind(c(1,2,3,4,5,6,7,8,9,10,12,13,14,16,17,18,20),avg.prem,rank.avg.prem,
std.lossratio)
model1.new=data.frame(model1.new)
model1.new[,3]=as.factor(model1.new[,3])
plot(rank.avg.prem,std.lossratio)
m2=lm(std.lossratio~rank.avg.prem,data=model1.new)
summary(m2)
abline(m2)

#model2 data preparation
data.new=data[!is.na(data$Loss.Ratio),]
View(data.new)
adverse=integer(nrow(data.new))
for (i in 1:nrow(data.new))
{
  if
(data.new$Incurred.Loss[i]>data.new$Expected.loss[i]+data.new$Expected.margin[i])
  adverse[i]=1
}
model2=cbind(data.new$Company,data.new$Year.,data.new$Premium.,adverse)
model2=data.frame(model2)
plot(data.new$Premium.,adverse)

#model 2

```

```

m3 = glm(adverse~ data.new$Premium., data = model2,family= "binomial")
summary(m3)
abline(m3)

#model 2 new
quantile(data.new$Premium.)
rank.prem=integer(length(data.new$Premium.))
for (i in 1:length(data.new$Premium.))
{
  if (data.new$Premium.[i]<quantile(data.new$Premium.,0.25))
  {
    rank.prem[i]=1
  }
  else if (data.new$Premium.[i]<quantile(data.new$Premium.,0.75))
  {
    rank.prem[i]=2
  }
  else
    rank.prem[i]=3
}
rank.prem=as.factor(rank.prem)
model2.new=cbind(1:length(data.new$Premium.),data.new$Premium.,rank.prem,adverse)
model2.new=data.frame(model2.new)
rank.prem=as.factor(rank.prem)
plot(rank.prem,adverse)
m4=glm(adverse~rank.prem,data=model2.new,family='binomial')
summary(m4)
abline(m4)

#calendar years
nyears=sapply(1:20,function(x) nrow(data.new[data.new$Company==x,]))
for (i in 1998:2008)
{plot(data.new[data.new$Year==i,1],data.new[data.new$Year==i,5],type='l',xlab="",ylab="")
,xlim=c(1,20),ylim=c(0,1.2),col=10+i-1997)
  par(new=T)
}
boxplot(data.new$Loss.Ratio.~data.new$Year.,main='Loss Ratio & Calendar Year')
boxplot(data.new$Premium.~data.new$Year.,main='Premium & Calendar Year')

#model 3 data preparation
data.new=data[!is.na(data$Loss.Ratio.),]

```

```
View(data.new)
adverse2=integer(nrow(data.new))
for (i in 1:nrow(data.new))
{
  if (data.new$Loss.Ratio.[i]>data.new$Expected.Loss.Ratio[i])
    adverse2[i]=1
}
Model3=cbind(data.new$Company,data.new$Year.,data.new$Premium.,adverse2)
Model3=data.frame(model4)
plot(data.new$Premium.,adverse2)

#model 3
m6= glm(adverse2~ data.new$Premium., data = model3,family= "binomial")
summary(m6)

#model 4 data preparation
data2=read.xlsx("stop loss data.xlsx",sheetName="data2")
data2.new=data2[!is.na(data2$Premium.),]
data2.new$AtoE=as.numeric(as.character(data2.new$AtoE))
model4=cbind(data2.new$Premium,data2.new$AtoE)
plot(data2.new$Premium,data2.new$AtoE,xlab='Premium',ylab='A/E')
```