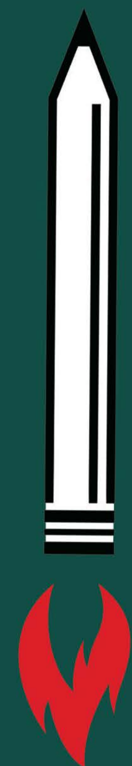
EXAM 8
COOKBOOK



Step-by-Step Recipes to Solve CAS Calculation Problems

2022

RisingFellow

Stephen Roll, FCAS

Exam 8 Cookbook

2022 Sitting

Stephen Roll, FCAS

RisingFellow

Published By:

Rising Fellow
United States, NY, 14610
www.RisingFellow.com

Contact: info@RisingFellow.com

Published in United States

Table of Contents

Introduction

The purpose of the Exam 8 Cookbook is to prepare you to confidently answer calculation-based problems on exam day without wasting time trying to "think through" a problem-solving approach before typing the solution.

Since the 2016 sitting, hundreds of actuaries have used the Exam 7, Exam 8, and Exam 9 Cookbooks to help them pass and get one step closer to their FCAS.

Our goal with Rising Fellow is to help you prepare for the exam with less frustration so that you have your best exam sitting yet!

The Structure

The Exam 8 Cookbook goes through the different calculation-based problem-types that I believe are reasonably testable based on the syllabus. By exam day, you should know how to solve each one.

Inside, you'll find a separate section for each testable problem-type. Each section has the following structure:

Original Practice Problem

Each section has an original practice problem that demonstrates the problem-type. I wrote these based off of the syllabus papers to have a similar difficulty-level and style to what you might see on an exam.

Solution Recipe

The solution recipe solves the practice problem from start to finish and shows the step-by-step approach you should take to answer a similar problem. For each step, you'll see:

- The description for what to do in the step
- The formula(s) necessary for the step
- The formula(s) translated from symbolic notation to plain-English
- Calculations for the step to solve the example problem

Discussion

Each section includes discussion to add clarity and more context. The discussion also covers underlying concepts that might come up on a part b or part c essay question.

For many problems, I point out potential "twists" that could show up on the exam that would make an exam problem more difficult. Since you've taken actuarial exams up to this point, you know that straightforward exam problems are more the exception than the rule.

CBT Spreadsheet Tips

This new section provides Excel formulas and tips for how to solve a problem more efficiently in the PearsonVue spreadsheet environment. There are many types of problems where setting up your solution intelligently and taking advantage of the spreadsheet capabilities such as `SUMIF( )`, `COUNTIF( )`, and array formulas, will save you valuable time on the exam.

Source

Each section references the pages in the syllabus reading that you can cross-reference for more information and details. Make sure to check the syllabus section for more context if you get stuck on a problem or to see how the author discusses the concepts.

More Practice

Here, you'll see references to past CAS problems. You'll find this helpful especially closer to the exam if there are particular types of problems that you are struggling with. This section includes references to CAS problems from the 2011-2019 exams, the years for the current syllabus structure.

How to Best Use the Exam 8 Cookbook

Below is a suggested guide for how you can incorporate the Exam 8 Cookbook in your own study schedule along with the syllabus material and a typical study manual. This is the general approach that I used when I took my fellowship exams.

For each of those exams I had a main study manual as well as the Exam Cookbook, which I built out while I studied for the exam (but you don't need to waste time doing that part!)

First pass through the syllabus

While you're reading a particular paper in the syllabus and your main study manual to learn the material, use the Exam 8 Cookbook to clearly identify what problem-types you need to know from the paper. Study the steps in the solution recipe to learn how to solve the problem-types. Make sure to do some practice problems as you go through the syllabus. This will help you learn faster.

Second pass through the syllabus

Review the steps for the problem-types and make sure you have an intuitive understanding of how to solve the problems. Start working the past CAS problems.

The first level of understanding is to be able to follow the recipe and understand the steps and calculations.

The next level of understanding is to be able to recall and apply the steps to solve a problem without relying on study material. During your second pass, focus on building this deeper level of understanding.

Review and Practice Problems (around 6 weeks to 2 weeks before the exam)

At this point you should have a good understanding of the syllabus and how to use the recipe steps to systematically solve the different calculation problems. During this period, you should be doing lots of problems across the syllabus and targeting problem-types that you are finding particularly challenging.
By the end of this phase, you might not have all the formulas memorized, but you should know all the steps and how to apply them to solve problems without needing to think too much before beginning to write the solution.

During this phase, make sure to focus on the types of problems and concepts that you're weak at. This may require some struggle, but struggling with some of the challenging problems will help you master these concepts.

You also should continue building your understanding of the concepts and preparing for essay and more complicated integrative questions. I found it helpful to create flashcards from the papers as well as to re-read sections of the syllabus papers that appear to be likely sources of essay problems.

Final Weeks

In the final weeks, focus on taking practice exams to see problems from the entire syllabus. When taking practice exams, work on your exam strategy to make sure you're able to finish the exam and maximize your points.

Prepare for essay problems in the final weeks by using flashcards to make sure that you know all the details necessary. An approach I found helpful is to say flashcards out loud and to explain the flashcard response in my own words as if I were teaching someone. It sounds weird, but it is a much more efficient way to learn and memorize than simply scanning the front and back of the flashcard.

Prepare for calculation problems by reviewing the recipes in the Exam 8 Cookbook in a similar fashion to how you use flashcards for essay problems. Using this approach on my fellowship exams, I was able to rapidly review the steps and formulas for how to solve each problem-type that might show up on the exam. This was a huge benefit and gave me a lot of confidence going into the exam.

Exam Day

I used the original Exam Cookbooks together with a traditional study manual using the approach above to take my fellowship exams. On exam day, for almost every calculation problem I was able to start writing the solution without wasting time trying to think through how to solve the problem. I had an intuitive understanding of how to solve each of the problems following the step-by-step recipes.

If you follow this approach, you should be able to develop a similar level of understanding and confidence going into the exam room.

Excel Version for Computer-Based Testing Preparation

For each recipe, there is an accompanying Excel version. Make sure to review those so that you know how to solve problems in the spreadsheet format. The CBT Spreadsheet Tips sections and the Excel version showing the formulas and setup for the spreadsheet solution will help you understand how to solve exam problems in the PearsonVue spreadsheet environment.

Errata

I always hated seeing errors in study manuals when I studied for exams, so I make every effort to ensure the study materials are accurate. Nevertheless, there may still be some errors in the final version, so I keep an updated errata. Please make sure to check it regularly for any fixes. The link is below:

https://risingfellow.com/errata

If you find any errors, please send me a message using the contact form on the Errata page so that I can make a correction.

Feedback

I am always working to improve the Exam 8 Cookbook and the rest of the Rising Fellow study material. Please send me an email to exam8@RisingFellow.com if you have feedback about any of the following:

- Recipes or sections that are confusing or could be improved
- New recipes I should include in future versions
- Better ways you've found to solve a problem-type in a spreadsheet
- Any comments or other feedback you have

Reviews

If you find the Exam 8 Cookbook helpful this sitting, please leave us a review and let us know how it helped you prepare for the exam. Other actuaries look at reviews to help decide what study material to buy and it's helpful for us to hear feedback from actuaries like you so that we can better understand what's working and what can be improved.

You can leave us a review by sending us an email to info@RisingFellow.com. Thank you!

Good luck as you start studying and I hope this will be your best sitting yet.

Experience Credibility of a Single Exposure

Bailey & Simon

Problem

Given the following information for Class 1 of private passenger auto liability:

Merit Rating Group	Number of Accident-Free Years	Earned Prem. at Present B Rates	Number of Claims Incurred	Earned Car Years
A	3+	1,070,000	1,310	12,800
X	2	70,000	100	810
Y	1	85,000	130	980
B	0	150,000	260	1,720
Total		1,375,000	1,800	16,310

a. Calculate the credibility of one exposure with two or more years of accident-free experience.

b. Calculate the credibility of one exposure with zero years of accident-free experience.

Solution Recipe

Part a – Credibility for an exposure with n+ accident-free years

1) **Calculate the group modification (Mod) as the relative claim frequency between the merit rating group and the class total. This is the merit rating factor for the group. Use the claim frequency per *Earned Premium at Present B Rates* instead of Earned Car Year.**

$$\boxed{Mod = \frac{\left(\dfrac{\#\,Claims_{Group}}{Earned\ Prem_{Group}}\right)}{\left(\dfrac{\#\,Claims_{Class}}{Earned\ Prem_{Class}}\right)}}$$

$$Mod = \frac{\left(\dfrac{1,310+100}{1,070,000+70,000}\right)}{\left(\dfrac{1,800}{1,375,000}\right)} = .945$$

Use groups A and X to get the 2+ AY-free group.

2) **Calculate the credibility of a single exposure in the group.**

$$\boxed{Z = 1 - Mod}$$

$$Z = 1 - .945$$
$$= \boxed{5.5\%}$$

Note:
The formula for Z above is derived from the full Mod credibility formula. Since R is zero for accident-free risks (R = the ratio of actual losses-to-expected losses), the credibility formula simplifies to the form above:

$$Mod = ZR + (1 - Z)$$

$$R = 0 \rightarrow Mod = 1 - Z$$

Part b – Credibility for an exposure with 0 accident-free years

1) Calculate the group modification, same as above.

$$Mod = \frac{\left(\dfrac{\#\,Claims_{Group}}{Earned\ Prem_{Group}}\right)}{\left(\dfrac{\#\,Claims_{Class}}{Earned\ Prem_{Class}}\right)}$$

$$Mod = \frac{\left(\dfrac{260}{150{,}000}\right)}{\left(\dfrac{1{,}800}{1{,}375{,}000}\right)} = 1.324$$

2) Calculate λ, the overall claims frequency per Earned Car Year for the entire class, *not* group B.

$$\boxed{\lambda = \frac{\#\,Claims_{Class}}{ECY_{Class}}}$$

$$\lambda = \frac{1{,}800}{16{,}310} = .110$$

3) Calculate *R*, the ratio of actual losses to expected losses for group B (with 0 accident-free years) based on the Poisson distribution.

$$\boxed{R = \frac{1}{1-\Pr(N=0)}}$$

Poisson Distribution:

$$\boxed{R = \frac{1}{1-e^{-\lambda}}}$$

$$R = \frac{1}{1-e^{-.110}} = 9.57$$

4) Solve for the credibility using the modification credibility formula.

$$\boxed{Mod = Z \cdot R + (1-Z)}$$

$$1.324 = Z \times 9.57 + (1-Z)$$

$$\rightarrow \boxed{Z = 3.8\%}$$

Discussion

Our goal is to use the experience of an individual risk to see how much the individual risk differs from the average risk in a rating class. This paper shows that the loss experience of a single exposure is credible and we can use the experience to segment risks *within* a rating class.

We calculate the Mod as the relative claim frequency between the rating group and the class as a whole using earned premium as the basis. Usually, we prefer to use earned premium in order to avoid the maldistribution of having higher claim frequency territories with more X/Y/B risks and also higher territorial premiums.

Make sure the earned premium is adjusted to the present rates at the same group level. Bailey & Simon adjust to group B rates, but it just matters that the earned premium is at the present rates of the same merit group level.

Using premium only corrects for maldistribution if:

- High frequency territories are high premium territories
- Territorial differentials are proper

Variation of Individual Risks within a Class

If we calculate the ratio of the 3+ year credibility to the annual claim frequency for each class, we can get an idea of how much variation there is of risks within each class. **Classes with higher ratios have a higher variation of individual hazards. This indicates that a class is more broadly defined than other classes.**

If the variation within a class was the same for each class, the credibility (Z) would vary proportionally to the average claim frequency meaning that the Z_{3+}-to-Claim $\text{Freq}_{\text{Class}}$ ratio would be the same for each class.

From the problem above, we can calculate some of the values for Class 1 (shaded below). The table below summarizes the other classes as well.

In the example below, Class 2 has a higher credibility for experience relative to the overall class claim frequency. This means the class is less narrowly defined and there's more variance in the likelihood of a claim for the insureds within Class 2.

Class	$Z_{3+\text{ AY-Free}}$ (1)	Claims (2)	Earned Car Years (3)	Class Claim Frequency (λ) (4) = (2)/(3)	Ratio: Z_{3+}-to-Claim $\text{Freq}_{\text{Class}}$ (5) = (1)/(4)
1	.065	1,800	16,310	0.110	0.587
2	.081	1,760	19,560	0.090	***0.900***
3	.030	520	11,200	0.046	0.646
4	.049	1,100	13,400	0.082	0.597

$$Mod_{3+AY-Free,\,Class\,1} = \frac{\left(\frac{1{,}310}{1{,}070{,}000}\right)}{\left(\frac{1{,}800}{1{,}375{,}000}\right)} = .935 \qquad Z_{3+AY-Free,Class\,1} = 1 - .935 = .065$$

A key point here is that greater variation of individual hazards within a class results in greater credibility for experience rating.

Derivation of R for an exposure with 0 accident-free years

$$R = \frac{Actual\ Losses_{0\ AY-Free}}{Expected\ Losses_{0\ AY-Free}} = \frac{\frac{\lambda}{(1-\Pr(N=0))}}{\lambda} = \frac{1}{1-\Pr\,(N=0)}$$

Source

Bailey & Simon – pg. 159-160 and pg. 164

More Practice

CAS 2019 – 3
CAS 2018 – 3
CAS 2015 – 1
CAS 2014 – 5

Relative Credibility: Empirical

Bailey & Simon

Problem

Given the following information for a class of private passenger auto liability:

Merit Rating Group	Number of Accident-Free Years	Earned Prem. at Present B Rates	Number of Claims Incurred
A	3+	3,700,000	4,600
X	2	430,000	600
Y	1	500,000	700
B	0	1,320,000	2,260
Total		5,950,000	8,160

Calculate the relative credibility for groups with 1+, 2+, and 3+ years of accident-free experience and interpret the results.

Solution Recipe

1) Calculate the relative claim frequency (Mod) for each group relative to the class total.

$$\boxed{Mod = \frac{\left(\dfrac{\#\,Claims_{Group}}{Earned\ Prem_{Group}}\right)}{\left(\dfrac{\#\,Claims_{Class}}{Earned\ Prem_{Class}}\right)}}$$

$$Mod_{1+} = \frac{\left(\dfrac{700+600+4,600}{500+430+3,700}\right)}{\left(\dfrac{8,160}{5,950}\right)} = .929$$

$$Mod_{2+} = .918$$

$$Mod_{3+} = .907$$

2) Calculate the credibility for each group.

$$\boxed{Z = 1 - Mod}$$

$$Z_{1+} = 1 - .929 = 7.1\%$$

$$Z_{2+} = 8.2\%$$

$$Z_{3+} = 9.3\%$$

3) Calculate the relative credibility compared to the 1+ years accident-free group.

$$Relative\,Credibility = \frac{Z_{Group}}{Z_{1+}}$$

	1+	2+	3+
Relative Credibility	1.00	1.16	1.32

If the class was stable with no risks leaving or entering and individual insured's chances of an accident constant over time, we'd expect credibility to increase roughly proportionally to the number of years. Since this is not the case (relative credibility of 3+ to 1+ is 1.32, much lower than 3), we conclude that the risks in a group are changing over time.

Discussion

In theory, we would expect the relative credibility to be roughly 1:2:3 for the groups above. This is because we'd expect credibility to vary roughly in proportion to the number of years.

Relative credibilities for the 2+ and 3+ groups may be significantly lower than 2 and 3 for the following reasons:

- Risks may be entering or leaving the class
- An individual insured's probability of an accident changes from one year to the next

The discussion by Hazam also points out that for larger credibilities, we wouldn't expect the relative credibilities to be as close to 1:2:3 as for smaller credibilities because of the credibility formula:

$$Z = \frac{p}{p + K}$$

The number of claims for the group, *p,* will increase proportionally as the number of years increase, but K is constant. Therefore, the true ratio of Z_{1+}:Z_{2+}:Z_{3+} will be less than 1:2:3.

Source

Bailey & Simon – pg. 160 and 163
Hazam – pg. 151

More Practice

CAS 2017 – 3
CAS 2011 – 1

Shifting Parameters: Chi-Squared Test

Mahler

Problem

Given the following claim frequency information for a book of business:

Year	Earned Car Years	Average Frequency
2005	12,000	8.3%
2006	13,000	9.4%
2007	14,500	9.2%
2008	14,000	9.7%
2009	13,650	9.5%
2010	13,800	8.8%
2011	14,500	9.3%
2012	15,000	9.4%

Year	Earned Car Years	Average Frequency
2013	15,750	9.7%
2014	16,000	10.2%
2015	16,000	10.1%
2016	16,750	9.9%
2017	17,500	10.3%
2018	18,000	9.6%
2019	18,000	10.5%

Chi-Squared Distribution

Degrees of Freedom	.05	.025
2	5.99	7.38
3	7.81	9.35
4	9.49	11.14
5	11.07	12.83

- Assume claim frequency follows a Poisson distribution

The claim frequency parameter is calculated for three-year intervals over the 15-year historical time period.

Evaluate whether the claim frequency is shifting over time using the Chi-squared test at the 5% significance level.

Solution Recipe

1) Calculate the overall parameter (frequency) for all years.

$$\lambda = \frac{\sum Exposure \cdot Avg\ Freq}{\sum Exposure}$$

$$\lambda = \frac{12,000 \times .083 + 13,000 \times .094 + \cdots}{12,000 + 13,000 + \cdots}$$
$$= 9.65\%$$

2) Calculate Earned Car Years and the weighted average claim frequency for each three-year interval.

$$\lambda = \frac{\sum Exposure \cdot Avg\ Freq}{\sum Exposure}$$

Years	Earned Car Years	Average Frequency
2005-2007	39,500	9.0%
2008-2010	41,450	9.3%
2011-2013	45,250	9.5%
2014-2016	48,750	10.1%
2017-2019	53,500	10.1%

3) **Calculate the χ^2 test statistic between actual and expected claim counts, based on exposures.**

$$\boxed{\chi^2 = \sum \frac{(actual - expected)^2}{expected}}$$

Years	Actual (A)	Expected (E)	$= \frac{(A-E)^2}{E}$
2005-2007	3,552	3,810	17.44
2008-2010	3,869	3,998	4.15
2011-2013	4,286	4,364	1.40
2014-2016	4,906	4,702	8.88
2017-2019	5,421	5,160	13.14
Total	22,034	22,034	45.00

$Actual_{2005-2007} = 39{,}500 \times 9.0\% = 3{,}552$

$Expected_{2005-2007} = 39{,}500 \times 9.65\% = 3{,}810$

4) **Compare the χ^2 test statistic to the critical value for the relevant χ^2 distribution (with n-1 degrees of freedom) and interpret the test.**

$critical\ value = \chi^2_{n-1,\alpha\%}$

$\mathrm{H_0}$: Frequency parameter is not shifting over time

${\chi^2}_{5-1,5\%} = 9.49$

45.00 > 9.49, so we reject the null hypothesis and conclude that frequency is shifting over time.

Discussion

The goal of the Chi-squared test as well as the lagged correlations test is to see if the risk parameters (e.g. average frequency) are shifting over time.

A key point from Mahler is that **when parameters are shifting over time, we should place less credibility on older data and higher credibility on more recent data.** Also, it will be more important to minimize the delays in data that are used for predictions and ratemaking.

CBT Spreadsheet Tips

As a shortcut, you can also calculate χ^2 directly with an array formula without first calculating the helper values of (Actual - Expected)^2 / Expected. See the Excel version for how this works.

χ^2 = SUM((Actual Values – Expected Values)^2 / Expected Values)

Important: In Pearson Vue, make sure to press CTRL + SHIFT + ENTER to calculate array formulas.

Source

Mahler – pg. 235-236

More Practice

CAS 2018 – 1
CAS 2015 – 4
CAS 2012 – 3

Shifting Parameters: Lagged Correlations

Mahler

Problem

Given the following loss ratios between 2010 and 2020 for a line of business:

Year	2020	2019	2018	2017	2016	2015	2014	2013	2012	2011	2010
Loss Ratio	.605	.689	.606	.727	.579	.744	1.066	.977	.944	.906	.764

Using correlations between loss ratios at lags of 1-5 years, test if the loss ratio for this line of business is shifting over time.

Solution Recipe

1) Calculate the correlation of the parameter at increasing lag periods (1-year, 2-year, …).

$$\boxed{r = \frac{\overline{XY} - \overline{X} \cdot \overline{Y}}{\sigma_X \cdot \sigma_Y}}$$

For the first pair (1-year lag):
X = Loss Ratio_t
X = Loss Ratio_{t-1}

<u>In Excel:</u>
CORREL(X values, Y values)

Lag (years)	Correlation	
1	0.598	←Correlate Years 2011-2020 and Years 2010-2019 (1-year lag)
2	0.325	
3	0.101	←Correlate Years 2013-2020 and Years 2010-2017 (3-year lag)
4	-0.185	
5	0.006	

2) Test if the correlations fall as the lag increases. A falling correlation indicates the parameter is shifting over time.

The correlation is relatively high for a 1-year lag but decreases as the lag increases.
This indicates the loss ratio is shifting over time.

Discussion

The key idea here is that if there is significant correlation between years close in time (like the 1-year lagged values here), the recent years can be used to help predict <u>*future values*</u>. We'll see this in the next recipe, "Credibility-Weighted Predictions."

If parameters were stable over time, the correlations wouldn't vary by the length of the lag in years. Higher correlations for years closer together than for years farther apart indicates that the parameters are shifting.

The correlations in this problem are noisy because of the small number of data points. The example in Mahler (pg. 237-238) shows lagged correlations with a smoother trend from high correlations for shorter lags to negligible correlations for longer lags.

An exam problem may give a table of correlations, like in the Mahler example, and ask you to interpret whether parameters are shifting over time.

Source

Mahler – pg. 235, 237-239

Credibility-Weighted Predictions X_{est}

Mahler

Problem

Given the following historical collision claims frequency for a book of Motorcycle policies:

Year	Earned Car Years	Frequency
2014	6,500	0.033
2015	6,760	0.042
2016	7,030	0.035
2017	7,410	0.028
2018	7,630	0.039
2019	7,820	0.032
2020	7,900	0.025
2021	8,120	0.030

- Credibility (Z) 0.4

Calculate the predicted 2022 claims frequency for the book using the following methods:

i. Credibility weight of the most recent year and the overall mean frequency

ii. Equal weight to the three most recent years of data

iii. Credibility weight of the most recent year and the previous estimate

- Assume a grand mean of 0.03 is an appropriate a priori estimated value of 2014 frequency

iv. Credibility weight using varying weights on the data and the grand mean

- 40% weight on the first prior value, 20% on the second prior value and the rest on the overall mean

Solution Recipe

1) **Calculate the overall mean using all the years prior to the year being estimated. If there is an exposure basis or earned premium, make sure to use a weighted average.**

$$\mu = \frac{\sum Exposure_i \cdot Freq_i}{\sum Exposure_i}$$

$$\mu = \frac{6{,}500 \times .033 + 6{,}760 \times .042 + \cdots}{6{,}500 + 6{,}760 + \cdots}$$

$$= .0328$$

Note:
Since we're estimating 2022 claims frequency, the overall mean should reflect all years prior to 2022. If we were back testing and estimating the 2018 frequency, the overall mean would include years prior to 2018.

If a problem provides an overall mean or "grand mean", use that for μ.

2) Calculate the estimated value X using the appropriate credibility-weighting from Mahler:

Credibility Weight of the Most Recent Year and the Grand Mean

$$\boxed{X_{est} = Z \cdot Y_1 + (1-Z) \cdot \mu}$$

$$X_{est} = .4 \times .030 + (1 - .4) \times .0328$$
$$= \boxed{.0317}$$

Equal Weight to the N Most Recent Years

$$\boxed{X_{est} = \frac{Z}{N} \cdot \sum_{i=1}^{N} Y_i + (1-Z) \cdot \mu}$$

$$X_{est} = \frac{.4}{3} \times (.032 + .025 + .030) + (1 - .4) \times .0328$$
$$= \boxed{.0313}$$

Credibility Weight of the Most Recent Data and Previous Estimate

$$\boxed{X_{est,i+1} = Z \cdot Y_i + (1-Z) \cdot X_{est,i}} \qquad X_{est,i+1} = Z \cdot Y_i + Z(1-Z) \cdot Y_{i-1} + Z(1-Z)^2 \cdot Y_{i-2} + \cdots + (1-Z)^N \cdot \mu$$

$$X_{est,2022} = .4 \times .03 + (1 - .4) \times .0303$$
$$= \boxed{.0302}$$

Year	Frequency	Estimate	
2014	0.033	0.0300	<- Given in problem
2015	0.042	0.0312	$= .4 \times .033 + (1 - .4) \times .030$
2016	0.035	0.0355	$= .4 \times .042 + (1 - .4) \times .0312$
2017	0.028	0.0353	$= .4 \times .035 + (1 - .4) \times .0355$
2018	0.039	0.0324	
2019	0.032	0.0350	
2020	0.025	0.0338	
2021	0.030	0.0303	
2022		0.0302	

Credibility Weight with Varying Weights

$$\boxed{X_{est} = \sum_{i=1}^{N} Z_i \cdot Y_i + \left(1 - \sum Z_i\right) \cdot \mu}$$

$$X_{est} = .4 \times .03 + .2 \times .025 + (1 - .4 - .2) \times .0328$$
$$= \boxed{.0301}$$

Discussion

This section in Mahler shows different ways of how recent years of experience can be used to predict future experience. This is relevant for prospective experience rating. If parameters are shifting over time, we want to use a credibility-weighting formula that applies more credibility on recent years and less credibility on older years.

Mahler shows several different ways to credibility-weight past experience and the overall mean. In the next recipe, we'll see how to find the optimal credibility weighting (Z) in order to get the best estimate of future experience.

Special Cases of the Credibility Formula

$$\boxed{X_{est} = Z \cdot Y_1 + (1 - Z)\mu}$$

Every Risk is Average:
0% credibility on recent data (except as part of the overall mean)

$Z = 0 \rightarrow X_{est} = \mu \quad X_{est} = \boxed{0.0328}$

The Most Recent Year Repeats:
100% credibility on past data

$Z = 1 \rightarrow X_{est} = Y_1 \quad X_{est} = \boxed{0.0300}$

CBT Spreadsheet Tips

For the third method, "Credibility Weight of the Most Recent Data and Previous Estimate", the easiest approach is to set up a table and calculate all the estimates from the beginning instead of attempting a direct calculation of the final estimate.

Source

Mahler – pg. 239-241, 255-256

More Practice

CAS 2018 – 1

Finding Optimal Z: MSE Criteria

Mahler

Problem

Given the following historical loss ratios for a book of business:

Year	Loss Ratio
2011	0.832
2012	0.738
2013	0.825
2014	0.870
2015	0.838
2016	0.736
2017	0.911
2018	0.941
2019	1.207
2020	0.872

- Complement of credibility loss ratio 85%

Predicted loss ratios are estimated as a credibility-weighting of the most recent year and the complement of credibility loss ratio.

Use the least squared error criteria to determine which credibility weight on the most recent year produces the best predicted loss ratio estimates. Select the optimal Z between 0% and 100% in 20% intervals.

Solution Recipe

1) **Calculate the estimated values for each credibility scenario and set up a table of estimated and actual values for each scenario.**

$X_{est} = Z \cdot Y_1 + (1 - Z)\mu$ | $X_{est,2020\ \&\ Z=.2} = .2 \times 1.207 + (1 - .2) \times .85 = .921$

Year	Actual Loss Ratio(Y_i)	Predicted Loss Ratio (X_{est}) Z = 0%	20%	40%	60%	80%	100%
2011	0.832						
2012	0.738	0.850	0.846	0.843	0.839	0.836	0.832
2013	0.825	0.850	0.828	0.805	0.783	0.760	0.738
2014	0.87	0.850	0.845	0.840	0.835	0.830	0.825
2015	0.838	0.850	0.854	0.858	0.862	0.866	0.870
2016	0.736	0.850	0.848	0.845	0.843	0.840	0.838
2017	0.911	0.850	0.827	0.804	0.782	0.759	0.736
2018	0.941	0.850	0.862	0.874	0.887	0.899	0.911
2019	1.207	0.850	0.868	0.886	0.905	0.923	0.941
2020	0.872	0.850	0.921	0.993	1.064	1.136	1.207

Note:
We don't calculate an estimated loss ratio for 2011 since the estimate requires a loss ratio from the prior year.

2) Calculate the Mean Squared Error (MSE) for each credibility scenario.

$$\boxed{MSE = \frac{\sum \left(Y_i - X_{est,i}\right)^2}{n}}$$

$$MSE_{Z=.2} = \frac{(.738 - .846)^2 + (.825 - .828)^2 + \cdots}{9}$$

$$= .01728$$

	0%	**20%**	40%	60%	80%	100%
MSE	0.01852	**0.01728**	0.01753	0.01926	0.02246	0.02715

<u>MSE Calculation using Array formulas</u>
The quickest way to calculate MSE is by using an array formula as `AVERAGE( (Y Values - X Values)^2 )`. This allows you to calculate the MSE directly without needing to set up a whole table of "helper" calculations in the spreadsheet. See the Excel version for this.

Important: In Pearson Vue, make sure to press CTRL + SHIFT + ENTER to calculate array formulas.

3) Select the credibility with the smallest MSE.

Z = 20% has the smallest MSE of the two credibility scenarios and therefore produces better loss ratio estimates according the least squares error criteria.

Discussion

Even when we know what formula we want to use to calculate the predicted X_{est} values, we need to select a credibility weighting (Z) to use. The goal here is to back test on historical data to find the Z value that minimizes the MSE. This is the credibility weighting we'd use to estimate <u>*future*</u> loss ratios.

Calculating the Reduction in MSE

Based on the MSE at the extremes (Z = 0 and Z = 1), we can calculate how much the MSE is reduced to using the optimal Z (see Mahler pg. 252).

$$\boxed{\%\ Reduced = \frac{MSE_{optimal\ Z}}{min\left(MSE_{Z=0,}\ MSE_{Z=1}\right)}}$$

$$\%Reduced = \frac{.01728}{Min(.01852, .02715)} = 93\%$$

MSE is reduced to 93% of its previous value.

Using the Average as the Complement

For this problem we're using a selected complement of credibility. One twist to the problem is to use the average loss ratio as the complement instead. With the average loss ratio, make sure to use a <u>*running average*</u> as opposed to the all-year average (unless otherwise specified).

For example, below are the calculations for Z = 20%. Note the calculation of μ that goes into the X_{est} formula:

$$X_{est,2015} = .2 \times .87 + (1 - .2) \times \frac{.832 + .738 + .825 + .87}{4} =$$

Year	Actual Loss Ratio (Y_i)	μ Running Avg	X_{est} (Z=20%)
2011	0.832		
2012	0.738	0.832	0.832
2013	0.825	0.785	0.776
2014	0.87	0.798	0.804
2015	0.838	0.816	0.827
2016	0.736	0.821	0.824
2017	0.911	0.807	0.792
2018	0.941	0.821	0.839
2019	1.207	0.836	0.857
2020	0.872	0.878	0.943

Source

Mahler – pg. 242-246, 250-252

More Practice

CAS 2018 – 1

Finding Optimal Z: Limited Fluctuation

Mahler

Problem

Given the following loss frequency information for a book of business:

Year	Loss Frequency	Year	Loss Frequency
2010	0.83%	2016	0.75%
2011	0.83%	2017	0.90%
2012	0.59%	2018	1.04%
2013	0.67%	2019	0.83%
2014	0.77%	2020	0.83%
2015	0.95%	2021	0.90%

Predicted loss frequency is estimated as a credibility-weighting of the average of the two most-recent years and the selected overall frequency 0.8%.

A large error threshold of 8% is selected.

Use the small chance of large errors criterion to determine which credibility weight on the most recent two years produces the best predicted loss frequency estimates. Select the optimal Z between 0% and 100% in 20% intervals.

Solution Recipe

1) **Calculate the estimated values for each credibility scenario and set up a table of estimated and actual values for each scenario.**

$$X_{est,i} = \frac{Z}{2} \cdot \left(Y_{i-1} + Y_{i-2}\right) + (1 - Z)\mu$$

$$X_{est,2020,Z=20\%} = \frac{.2}{2} \times (.83\% + 1.04\%) + (1 - .2) \times .8\%$$

Year	Actual Loss Freq (Y_i)	0%	20%	40%	60%	80%	100%
2010	0.830%						
2011	0.830%						
2012	0.590%	0.800%	0.806%	0.812%	0.818%	0.824%	0.830%
2013	0.670%	0.800%	0.782%	0.764%	0.746%	0.728%	0.710%
2014	0.770%	0.800%	0.766%	0.732%	0.698%	0.664%	0.630%
2015	0.950%	0.800%	0.784%	0.768%	0.752%	0.736%	0.720%
2016	0.750%	0.800%	0.812%	0.824%	0.836%	0.848%	0.860%
2017	0.900%	0.800%	0.810%	0.820%	0.830%	0.840%	0.850%
2018	1.040%	0.800%	0.805%	0.810%	0.815%	0.820%	0.825%
2019	0.830%	0.800%	0.834%	0.868%	0.902%	0.936%	0.970%
2020	0.830%	0.800%	0.827%	0.854%	0.881%	0.908%	0.935%
2021	0.900%	0.800%	0.806%	0.812%	0.818%	0.824%	0.830%

Note:
We don't calculate an estimated loss frequency for 2010/2011 since the estimate requires a loss frequency from the prior two years.

2) **Calculate the absolute errors between the actual and expected values for each year and credibility scenario. Use the ABS(number) spreadsheet formula.**

$$\boxed{|Error\%| = \frac{|Y_i - X_{est,i}|}{X_{est,i}}}$$

$$|Error\%| = \frac{|actual - expected|}{expected}$$

$$|Error\%_{2020,Z=20\%}| = \frac{|.83\% - .827\%|}{.827\%}$$
$$= 0.36\%$$

Absolute Errors for Credibility Level Z

Year	0%	20%	40%	60%	80%	100%
2012	26.25%	26.80%	27.34%	27.87%	28.40%	28.92%
2013	16.25%	14.32%	12.30%	10.19%	7.97%	5.63%
2014	3.75%	0.52%	5.19%	10.32%	15.96%	22.22%
2015	18.75%	21.17%	23.70%	26.33%	29.08%	31.94%
2016	6.25%	7.64%	8.98%	10.29%	11.56%	12.79%
2017	12.50%	11.11%	9.76%	8.43%	7.14%	5.88%
2018	30.00%	29.19%	28.40%	27.61%	26.83%	26.06%
2019	3.75%	0.48%	4.38%	7.98%	11.32%	14.43%
2020	3.75%	0.36%	2.81%	5.79%	8.59%	11.23%
2021	12.50%	11.66%	10.84%	10.02%	9.22%	8.43%

3) **Calculate the empirical probability that the absolute error is greater than the threshold criteria. The quickest way to do this is to use the** COUNTIF(range, criteria) **formula.**

$$\Pr\left(|Error\%| > threshold\right) = \frac{\#|Error\%| > threshold}{n}$$

$$\Pr(|Error\%_{Z=20\%}| > 10\%) = \frac{6}{10}$$
$$= 60\%$$

Probability of Large Error

Credibility (Z)	0%	20%	**40%**	60%	80%	100%
Pr(Large Error)	60%	60%	**50%**	70%	60%	70%

In Excel
=COUNTIF(Absolute Error Values, ">"& Threshold) / COUNT(Absolute Error Values).

4) **Select the credibility that minimizes the probability of large errors.**

The probability of large errors over the threshold of 10% is minimized when the credibility Z is 40%. Therefore, Z = 40% produces the better estimates.

Discussion

The large error threshold is an arbitrary choice. In the paper, Mahler looks at 5%, 10%, and 20% as threshold values for this method.

Another point Mahler makes is that there's not a single "correct" optimal credibility value for Z. Instead, based on the different criteria, there will be a range of values for the credibility that work well.

Discussion

For the calculation of the probability of large error, use the COUNTIF(range, criteria) formula to count how many of the absolute errors (from Step 2) are greater than the Large Error Threshold.

Important: The setup of the COUNTIF( ) formula is tricky because it uses a logical operator. The correct criteria is: ">"& Threshold.

The full formula is:
COUNTIF(Absolute Error Values, ">"& Threshold)

Source

Mahler – pg. 242

Finding Optimal Z: Meyers-Dorweiler

Mahler

Problem

Given the following information for a book of business:

Year	Actual Loss Ratio
2011	83.2%
2012	73.8%
2013	82.5%
2014	87.0%
2015	83.8%
2016	73.6%
2017	91.1%
2018	94.1%
2019	120.7%
2020	87.2%
Overall Average	87.7%

For years 2013-2020, a predicted loss ratio is estimated using a credibility weighting of the prior years and the average loss ratio to-date for the year estimated as the complement of credibility.

Two credibility-weighting methods are analyzed:

- Scenario 1: 25% weight on the prior year
- Scenario 2: 25% weight on the average of the prior two years

Use the Meyers-Dorweiler criterion with the least squares correlation to determine whether one or two years of prior experience should be used to estimate a predicted loss ratio for the upcoming year.

Solution Recipe

1) **Calculate the estimated values for each credibility scenario and set up a table of estimated and actual values for each scenario.**

$$X_{est} = \sum_{i=1}^{N} Z_i \cdot Y_i + \left(1 - \sum Z_i\right) \cdot \mu$$

N = 2, Z = 25%

$$Running\ Avg\ \mu_{2018} = \frac{.832 + .738 + \cdots + .911}{7} = .821$$

$$X_{est,2018} = .25 \times \frac{.911 + .736}{2} + (1 - .25) \times .821 = .822$$

Note:
We use a running average of the prior years' loss ratios for μ to calculate the predicted loss ratio. Use this approach unless otherwise specified.

Year	Actual Loss Ratio (Y_i)	μ Running Avg	X_{est} (N=1, Z=25%)	X_{est} (N=2, Z=25%)
2011	0.832			
2012	0.738			
2013	0.825	0.785	0.773	0.785
2014	0.870	0.798	0.805	0.794
2015	0.838	0.816	0.830	0.824
2016	0.736	0.821	0.825	0.829
2017	0.911	0.807	0.789	0.802
2018	0.941	0.821	0.844	0.822
2019	1.207	0.836	0.863	0.859
2020	0.872	0.878	0.960	0.927

2) **Calculate the two quantities for the Meyers-Dorweiler criterion by year. The first is Actual/Predicted and the second is Predicted/Overall Average.**

$$\boxed{Y = \frac{Actual\ Loss\ \%}{Predicted\ Loss\ \%}}$$

$$\boxed{X = \frac{Predicted\ Loss\ \%}{Overall\ Average}}$$

2018 quantities, N = 2

$$Y = \frac{.941}{.822} = 1.1448$$

$$X = \frac{.822}{.877} = .9372$$

Scenario 1: N = 1, Z = 25%

Year	Y	X
2013	1.0669	0.8817
2014	1.0807	0.9179
2015	1.0100	0.9461
2016	0.8922	0.9406
2017	1.1548	0.8995
2018	1.1152	0.9622
2019	1.3994	0.9835
2020	0.9084	1.0945

Scenario 2: N = 2, Z = 25%

Year	Y	X
2013	1.0510	0.8951
2014	1.0955	0.9055
2015	1.0169	0.9396
2016	0.8879	0.9452
2017	1.1364	0.9141
2018	1.1448	0.9372
2019	1.4055	0.9792
2020	0.9410	1.0566

3) **Calculate the Meyers-Dorweiler criterion as the correlation between the Actual/Predicted and the Predicted/Overall Average. For this problem, the correlation statistic is the traditional least-squares correlation.**

$$\boxed{r = \frac{\overline{XY} - \overline{X} \cdot \overline{Y}}{\sigma_X \cdot \sigma_Y}}$$

Y = Actual / Predicted

X = Predicted / Overall

In Excel

= CORREL(X Values, Y Values)

Meyers-Dorweiler criterion

	Correlation
Scenario 1	-0.1968
Scenario 2	-0.0972

4) Select the credibility that results in a correlation closest to zero.

Using the prior two years (scenario 2), the correlation statistic is closer to zero than using one prior year.

Therefore, I select using the prior two years of experience for calculating the estimated loss ratios.

Discussion

For this type of problem, I would understand how the Meyer-Dorweiler criterion is calculated, but focus more on how to interpret the results and the difference from the first two criteria.

Mahler uses the Kendall Tau statistic for the correlation calculation. The Kendall Tau calculation is off the syllabus since it's in the Appendix, but he mentions that other correlation statistics can be used.

Use the CORREL(array 1, array 2) function to calculate the correlation on the CBT spreadsheet.

Contrasting Meyers-Dorweiler vs. Least Squares & Limited Fluctuation

The first two criteria (Least Squared Error and Limited Fluctuation) are focused on limiting large errors. In contrast, the Meyers-Dorweiler criterion is focused on the pattern (or correlation) between the errors (actual / predicted values) and the experience modification (predicted / overall average).

A situation where the errors (actual / predicted) are small but correlated with the predicted-to-overall average ratio would be preferable for the first two criteria, but not for Meyers-Dorweiler.

Pg. 270-271 in the paper has a good discussion about the differences that is worth reading.

Source

Mahler – pg. 243-244, 270-271

Credibility-Wtd Class Excess Ratios (ELFs)

Robertson

Problem

Given the current excess ratios by Hazard Group:

Limit ($000)	HG A	HG B	HG C
100	.26	.30	.35
500	.11	.14	.17
1,000	.03	.05	.07

Class 3383 is currently mapped to Hazard Group B and has the following countrywide excess ratios based on the most recent five years of data, weighted by injury type:

Limit ($000)	Class 3383
100	.32
500	.15
1,000	.04

- The number of claims in class 3383 is 2,860
- The average number of claims per class is 3,300

Calculate the credibility-weighted vector of excess ratios for class 3383 that could be used in a hazard group cluster analysis, as described by Robertson.

Solution Recipe

1) **Calculate the credibility of the class excess ratio using the number of claims in the class, *n*, and the average number of claims per class, *k*.**

$$\boxed{z = \min\left(\frac{n}{n+k}\times 1.5,\, 1\right)}$$

$$z = \min\left(\frac{2{,}860}{2{,}860+3{,}300}\times 1.5, 1\right)$$

$$= 0.696$$

2) **Calculate the vector of excess ratios for the class as a credibility-weighting of the class excess ratios and the current hazard group excess ratios.**

$$\boxed{R_{cred} = z\cdot R_c + (1-z)\cdot R_{HG}}$$

$$XS\ Ratio_{cred} = z\cdot XS\ Ratio_{class} + (1-z)\cdot XS\ Ratio_{HG}$$

$$R_{cred} = 0.696\times\begin{bmatrix}.32\\.15\\.04\end{bmatrix} + (1-0.696)\times\begin{bmatrix}.30\\.14\\.05\end{bmatrix} = \begin{bmatrix}.314\\.147\\.043\end{bmatrix}$$

Discussion

The goal of calculating countrywide class excess ratios is to sort classes by excess ratio. Using credibility-weighted class excess ratios, we can use cluster analysis to group classes into Hazard Groups.

The number of claims by class is highly skewed with few classes having the majority of claims. Robertson notes that if k was defined as the median number of claims rather than the mean, then there would be a very large increase in credibility.

Square Root Rule as an Alternative Credibility Measure

In the paper, they considered a few square root rules for calculating credibility. As one example, the **simple square root rule** is below:

$$z = \sqrt{\frac{n}{384}}$$

In the formula, n is the number of claims, 384 is the full credibility standard, and z is capped at 1. For the class in the problem above, z would be 100% under this approach. See pg. 199 in the paper for a few other variations. In the end, the square root rule wasn't used for the Hazard Group analysis.

Source

Robertson – pg. 197-199

More Practice

CAS 2014 – 2

Weighted *k*-means Cluster Analysis

Robertson

Problem

The following information is used in a weighted *k*-means cluster analysis:

Class	Premium (000s)	Class Excess Ratio (R_C) 100k	Class Excess Ratio (R_C) 1,000k	Randomly Assigned Cluster
1	50	.5	.15	A
2	200	.4	.13	B
3	600	.8	.19	A
4	150	.7	.17	B

Calculate the reassigned clusters for each class after one iteration of the weighted *k*-means algorithm.

Solution Recipe

1) **Calculate the centroid for each of the k clusters as a premium-weighted average of the excess ratios of the classes in the cluster. See the Excel version for the spreadsheet calculations.**

$$\boxed{\bar{R}_i = \frac{\sum w_c R_c}{\sum w_c}}$$

$w_c = \%$ of total premium in class c

$$\overline{XS\ Ratio}_i = \frac{\sum weight_{class} \cdot XS\ Ratio_{class}}{\sum weight_{class}}$$

$$w_1 = \frac{50}{50+200+600+150} = .05$$

$$\bar{R}_A = \frac{.05 \times \begin{bmatrix} .5 \\ .15 \end{bmatrix} + .6 \times \begin{bmatrix} .8 \\ .19 \end{bmatrix}}{.05+.6} = \begin{bmatrix} .777 \\ .187 \end{bmatrix}$$

$$\bar{R}_B = \begin{bmatrix} .529 \\ .147 \end{bmatrix}$$

Calculation with Array Formula

$$= \frac{\text{SUM(IF(Class Clusters=Criteria, } \mathbf{Prem_c \times XS\ Ratio_c}\text{))}}{\text{SUMIF(Class Clusters, Criteria, } \mathbf{Prem_c}\text{)}}$$

Calculation *without* Array Formula

$$= \frac{\text{SUMIF(Class Clusters, Criteria, } \mathbf{w_c R_c\ values}\text{)}}{\text{SUMIF(Class Clusters, Criteria, } \mathbf{Prem_c}\text{)}}$$

- Use nested Sum(IF( )) as an array formula (see CBT Spreadsheet Tips)

Important: Press CTRL+SHIFT+ENTER to enter the formula as an array formula in Pearson Vue.

2) **Calculate the Euclidean Distance between each class's excess ratio vector and each cluster's excess ratio vector. Reassign each class to the cluster with the closest centroid based on the Euclidean distance between the excess ratio vectors. Repeat steps 1 and 2 until there are no reassignments.**

$$\boxed{\|R_c - R_{HG}\|_2 = \sqrt{\sum_{Limits}\left[R_c(L_i) - R_{HG}(L_i)\right]^2}}$$

$$distance = \sqrt{\left[R_{class}(L_1) - R_{HG}(L_1)\right]^2 + \cdots + \left[R_{class}(L_n) - R_{HG}(L_n)\right]^2}$$

Class 1:

$$\|R_1 - R_A\|_2 = \sqrt{(.5 - .777)^2 + (.15 - .187)^2)}$$
$$= .279$$
$$\|R_1 - R_B\|_2 = \sqrt{(.5 - .529)^2 + (.15 - .147)^2)}$$
$$= .029$$

Class	$\|R_c - R_A\|_2$	$\|R_c - R_B\|_2$	New Cluster
1	0.279	0.029	B
2	0.381	0.130	B
3	0.023	0.275	A
4	0.079	0.173	A

Calculation with Array Formula

=SQRT(SUM((**XS Ratios$_c$**-**XS Ratios$_{HG}$**)^2))

Calculation *without* Array Formula

=SQRT((**XS Ratios$_{c,100k}$**-**XS Ratios$_{HG,100k}$**)^2 +(**XS Ratios$_{c,1000k}$**-**XS Ratios$_{HG,1000k}$**)^2)

Important: Press CTRL+SHIFT+ENTER to enter the formula as an array formula in Pearson Vue.

Discussion

The goal of a cluster analysis is to group classes with similar excess ratios into new hazard groups.

A cluster analysis can be either non-hierarchical or hierarchical:

- **Non-hierarchical** – Analysis seeks the best partition of clusters for a pre-specified number of clusters
- **Hierarchical** – An analysis with *k*+1 clusters is the same as an analysis with *k* clusters, but with one of the clusters being subdivided into two clusters.

Robertson uses a non-hierarchical approach, like this problem.

The *k*-means clustering algorithm groups classes into *k* clusters that both:

- *Minimizes the within-cluster variance*, resulting in homogenous groups
- *Maximizes the between-cluster variance*, resulting in well-separated groups

L^1 vs L^2 Distance Metrics for Vectors

L^2 – Euclidean Distance

$$\boxed{\|x-y\|_2=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2}}$$

- Used extensively in statistics and preferred by the Robertson paper
- Penalizes large deviations (large deviations between class and HG excess ratios)

L^1 – Equal-Weighted Distance

$$\boxed{\|x-y\|_1=\sum_{i=1}^{n}|x_i-y_i|}$$

- Doesn't penalize large deviations more than the sum of many small deviations
- Minimizes the relative error in estimating the excess premium (between the HG and class)

I would use L^2, the Euclidean distance, which is preferred by Robertson, unless a problem clearly specifies or suggests otherwise.

CBT Spreadsheet Tips

In both steps, you can use array formulas as a shortcut.

An array formula allows us to put the `(Premium array)×(Excess Ratio array)` calculation directly in the formula when calculating the cluster centroids. The tricky part is that you can't use this directly in a `SUMIF()` formula. Instead, you need to use a `SUM(IF( ))` nested formula. Play around with the formulas above.

Without an array formula, you need to set up a helper calculation table. This is a perfectly fine solution too if it makes more intuitive sense to you. See the Excel version for how these calculations work in the spreadsheet.

When calculating the Euclidean Distance, use an array formula so that you can calculate the differences between the class and Hazard group excess ratio vectors for ALL limits at the same time. Without an array formula, you need a term for each excess ratio limit.

Source

Robertson – pg. 203-205

More Practice

CAS 2019 – 4
CAS 2017 – 2
CAS 2015 – 6
CAS 2013 – 4
CAS 2011 – 4

Calinski-Harabasz Statistic

Robertson

Problem

The k-means algorithm is used to create clusters of Workers Compensation classes based on excess loss factors. The following statistics are from cluster analyses using different numbers of clusters:

Number of Clusters	trace(W)
4	0.0048
5	0.0031
6	0.0024
7	0.0019
8	0.0018

- The trace of the dispersion matrix of class excess loss factors is 0.0470
- The number of classes in the dataset is 870

a. For each cluster analysis, calculate the Calinski-Harabasz statistic.

b. Interpret the results from part a and determine the optimal number of Hazard Groups for the dataset.

Solution Recipe

Part a – Calinski-Harabasz Statistic

1) Calculate trace(B) and trace(W) for each cluster analysis based on the T = B + W relationship.

$$\boxed{trace(T) = trace(B) + trace(W)}$$

$$\text{Var}(Total) = \text{Var}(Between\ clusters) + \text{Var}(Within\ clusters)$$

$$0.047 = trace(B_4) + 0.0048$$

$$\rightarrow trace(B_4) = 0.0422$$

k	trace(W)	trace(B)
4	0.0048	0.0422
5	0.0031	0.0439
6	0.0024	0.0446
7	0.0019	0.0451
8	0.0018	0.0452

2) Calculate the Calinski-Harabasz statistic for each cluster analysis. A higher value is better.

$$\boxed{C\text{-}H\ Statistic = \frac{trace(B)/(k-1)}{trace(W)/(n-k)}}$$

$$C\text{-}H = \frac{\text{Var}(Between\ clusters)/(\#\ clusters - 1)}{\text{Var}(Within\ clusters)/(\#\ classes - \#\ clusters)}$$

$$C\text{-}H_4 = \frac{.0422/(4-1)}{.0048/(870-4)}$$

$$= 2{,}538$$

k	C-H$_k$
4	2,538
5	3,062
6	3,211
7	**3,414**
8	3,092

Part b – Optimal Number of Clusters

A higher Calinski-Harabasz statistic indicates better clusters with higher between-cluster variance and lower within-cluster variance.

The analysis with seven clusters has the highest statistic, so seven is the optimal number of Hazard Groups.

Discussion

For a cluster analysis, we want:

- Low variance within clusters, trace(W), meaning clusters are homogenous
- High variance between clusters, trace(B), meaning clusters are well-separated

A key thing to understand is that the total sample variance, trace(T), is the sum of the total variance between clusters, trace(B), and the total variance of the class excess loss factors within clusters, trace(W).

Cubic Clustering Criterion (CCC)

The Cubic Clustering Criterion is another measure to find the optimal number of hazard groups. This measure compares the variance explained by a set of clusters to that expected if clusters are formed randomly. The greater the amount of variance explained by the clusters compared to the expected amount, the greater the CCC statistic.

Like the Calinski-Harabasz statistic, **the optimal number of clusters is indicated by the maximum CCC statistic.**

Source

Robertson – pg. 204-207

More Practice

CAS 2013 – 4

Estimated Frequency Relativities to TT

Couret & Venter

Problem

Given the following information for Class 836, assigned to Hazard Group D, based on Workers Compensation data from 2017-2021:

Claim Counts by Injury Type – Class 836

Year	F	PT	Major	Minor	TT
2017	6	7	55	220	540
2018	9	8	60	210	560
2019	7	10	58	240	600
2020	10	6	63	210	550
2021	5	9	56	230	570

Claim Frequency Relative to TT

	F	PT	Major	Minor	TT
Hazard Group D	.005	.007	.090	.380	1.00

Optimal Weights for Estimating the PT:TT Ratio

Ratio to Estimate	F	PT	Major	Minor
F:TT Ratio	0.12	0.13	0.20	0.22
PT:TT Ratio	0.1	0.16	0.23	0.25

Calculate the predicted ratio of F:TT and PT:TT claims for Class 836 using multi-dimensional credibility.

Solution Recipe

1) For class *c*, calculate the frequency relativity for each injury type as a ratio to TT claims.

$$\boxed{\textit{Frequency Relativity} = \frac{\#\,\textit{Injury-type Claims}}{\#\,\textit{TT Claims}}}$$

$V_c = F:TT\ ratio$ $\quad X_c = Major:TT\ ratio$

$W_c = PT:TT\ ratio$ $\quad Y_c = Minor:TT\ ratio$

$$V_{836} = \frac{6+9+\cdots+5}{540+560+\cdots+570} = .0131$$

$W_{836} = .0142$

$X_{836} = .1035$

$Y_{836} = .3936$

Claim Frequency Relative to TT

	V_c	W_c	X_c	Y_c
Class 836	0.0131	0.0142	0.1035	0.3936

2) **Calculate the estimated class frequency relativity for each injury type using the credibility-weights specific to the injury type, the frequency relativities for the class, and the frequency relativities for the Hazard Group.**

Estimating the F:TT frequency relativity

$$\boxed{v_c{}^{pred} = E[V] + b_v(V_c - E[V]) + c_v(W_c - E[W]) + d_v(X_c - E[X]) + e_v(Y_c - E[Y])}$$

$$v_{class}^{pred} = V_{HG} + b_v(V_{class} - V_{HG}) + c_v(W_{class} - W_{HG}) + d_v(X_{class} - X_{HG}) + e_v(Y_{class} - Y_{HG})$$

$$v_{836}^{pred} = .005 + .12(.0131 - .005) + .13(.0142 - .007) + .20(.1035 - .09) + .22(.3936 - .38)$$
$$= \boxed{.0126}$$

In Excel:
v_{class} = **V_{HG}** + SUM(**Credibility Weights$_V$** * (Freq Relativities$_{class}$-Freq Relativities$_{HG}$))

Estimating the PT:TT frequency relativity

$$w_{class}^{pred} = W_{HG} + b_w(V_{class} - V_{HG}) + c_w(W_{class} - W_{HG}) + d_w(X_{class} - X_{HG}) + e_w(Y_{class} - Y_{HG})$$

$$w_{class}^{pred} = .007 + .1(.0131 - .005) + .16(.0142 - .007) + .23(.1035 - .09) + .25(.3936 - .38)$$
$$= \boxed{.0155}$$

In Excel:
w_{class} = **W_{HG}** + SUM(**Credibility Weights$_W$** * (Freq Relativities$_{class}$-Freq Relativities$_{HG}$))

Important: Press CTRL+SHIFT+ENTER to enter the formula as an array formula in Pearson Vue.

Discussion

Couret & Venter point out that claim counts between different injury types are correlated. This means that if a class has a higher frequency of Major and Minor claims relative to the Hazard Group average, it probably has a higher frequency of Fatal and Permanent Total claims.

The idea of multi-dimensional credibility is to take advantage of that extra claim frequency information for a class instead of simply relying on the Hazard Group average. Doing this results in more accurate predictions of claim frequencies for a class.

One thing to remember is that the optimal weights will be different for estimating the ratio of each of the different injury types. You can see this by comparing the formulas in step 2 between the Fatal:TT and PT:TT ratio predictions. To calculate w^{pred}, we use weights b_w, c_w, d_w and e_w. To calculate v^{pred}, we use weights b_v, c_v, d_v and e_v.

CBT Spreadsheet Tips

The multi-dimensional credibility weighting calculation in step 2 is simpler using an array formula. Pay attention to the formulation which has a SUM( ) formula in it to sum up all the credibility calculations over all the injury-types.

You can calculate it without the array formula, but there are a lot of terms and you need to pay attention to make sure you reference all the different values correctly.

Without an Array Formula the calculation has the following term:
C21*(C35-C17)+D21*(D35-D17)+E21*(E35-E17)+F21*(F35-F17)

With an Array Formula this is the equivalent term:
SUM(C21:F21*(C35:F35-C17:F17))

Note how the array formulation does all four injury-type credibility calculations together. See the Excel version to follow the calculations in the spreadsheet.

Source

Couret & Venter – pg. 76-77

More Practice

CAS 2015 – 5

Quintiles Test

Couret & Venter

Problem

An actuary is using a quintiles test approach to test a multi-dimensional credibility procedure for Workers Compensation injury-type relativities. Using six years of data, even years were used to predict the ratio of Major:TT claims and odd years were used as a holdout sample.

Class	Even Years Major Claims	Odd Years Major Claims	Even Years TT Claims	Odd Years TT Claims	Predicted Major:TT Ratio
1	100	78	500	600	0.21
2	336	570	2,800	2,850	0.16
3	160	285	2,000	1,900	0.12
4	72	58	1,200	1,150	0.10
5	1,050	1,000	3,500	4,000	0.23
Total	1,718	1,991	10,000	10,500	

- Assume the Hazard Group only has these five classes

a. Perform a quintiles test with three groupings instead of five and calculate the Sum of Squared Errors for predictions based on the Hazard Group, Raw Even-Year data and the Credibility Procedure.

b. Assess whether the credibility procedure is an improvement over the Hazard Group predictions.

Solution Recipe

<u>Part a – Quintiles Test</u>

1) Sort classes by ascending predicted ratio from the multi-dimensional credibility procedure. Assign classes into groupings with similar TT claim counts.

Class	Even-year TT Claims	Predicted Major:TT
4	1,200	.10
3	2,000	.12
2	2,800	.16
1	500	.21
5	3,500	.23

Quintile Group	Classes	TT Claims
1	4, 3	3,200
2	2, 1	3,300
3	5	3,500

2) **For each quintile, calculate the average injury-type:TT ratio for the holdout data, the raw test data, and the credibility-weighted prediction. Also, calculate the holdout/test injury-type:TT ratios for the Hazard Group as a whole.**

With Claim Counts:

$$X_{quintile} = \frac{\sum \# Major_i}{\sum \# TT_i}$$

$$X_{1,holdout} = \frac{285+58}{1,900+1,150} = .1125$$

$$X_{1,even} = \frac{160+72}{1,200+2,000} = .0725$$

$$x_{1,cred} = \frac{.10 \times 1,200 + .12 \times 2,000}{1,200+2,000} = .1125$$

With Predicted Major:TTratio:

$$x_{quintile} = \frac{\sum x_i \cdot \# TT_i}{\sum \# TT_i}$$

$$X_{HG,odd} = \frac{1,991}{10,500} = .1896$$

$$X_{HG,even} = \frac{1,718}{10,000} = .1718$$

$$x_{HG,cred} = \frac{.21 \times 500 + \cdots + .23 \times 3,500}{10,000} = .1718$$

Quintile	Holdout Ratio	Raw Even Ratio	Cred-Wtd Ratio
1	.1125	.0725	.1125
2	.1878	.1321	.1676
3	.2500	.3000	.2300
HG	.1896	.1718	.1718

3) **For each quintile, calculate the relativities of the average injury-type:TT ratio (Step 2) divided by the average injury-type:TT ratio for all quintiles combined (the Hazard Group ratio).**

$$\boxed{Rel^X_{quintile} = \frac{x_{quintile}}{X_{HG}}}$$

$$X\ Relativity_{quintile} = \frac{Major:TT\ Ratio_{quintile}}{Major:TT\ Ratio_{HG}}$$

$$Rel^X_{1,holdout} = \frac{.1125}{.1896} = .593$$

$$Rel^X_{1,even} = .422$$

$$Rel^X_{1,cred} = .655$$

Quintile Group	Holdout Relativity	HG Relativity	Raw Even Relativity	Cred-Wtd Relativity
1	.593	1	.422	0.655
2	.991	1	.769	0.975
3	1.318	1	1.746	1.339

4) **Calculate the SSE for each of the predictions.**

$$\boxed{SSE = \sum_{quintiles} \left(Rel_{quintile} - Rel_{holdout}\right)^2}$$

$$SSE_{HG} = (1-.593)^2 + (1-.991)^2 + (1-1.318)^2 = .267$$

$$SSE_{even} = (.422-.593)^2 + (.769-.991)^2 + (1.746-1.318)^2 = .261$$

$$SSE_{cred} = (.655-.593)^2 + (.975-.991)^2 + (1.339-1.318)^2 = .004$$

In Excel

SUMXMY2(array x, array y) is a neat shortcut to calculate the sum squared difference of two arrays that works in PearsonVue. You can also calculate the SSE for the full arrays using an array formula to simplify the calculation. See the Excel Version for the calculations.

Part b – Quintiles Test Assessment

The quintiles test shows increasing relativities for the holdout data, indicating that the credibility-weighted approach does a good job at separating out classes into quintiles.

The Hazard Group predictions are too high for the group 1 and too low for group 3. The slope of the relativities for the Raw Even-year relativities is too steep and the predictions are too low for group 1 and too high for group 3. The Credibility-Weighted relativities are more closely in line with the holdout data.

The SSE corroborates this, since the SSE of the Credibility-Weighted relativities is significantly lower than for the Hazard Group and Even-year relativities. This shows that the credibility procedure is an improvement over the Hazard Group predictions.

Discussion

On pg. 80 of Couret & Venter, they show the SSE calculated across *all* classes. With this simplistic approach, the multi-dimensional credibility procedure only shows a modest improvement over the Hazard Group predictions.

The purpose of using the quintiles test is to remove a lot of the random noise when calculating the SSE at the class level. We do this by grouping the classes into quintiles based on the credibility procedure predictions and *then* calculating the SSE on the quintiles.

Note:
The even-year data would be used to calculate the predicted Major:TT ratios using the Couret & Venter multi-dimensional credibility method. Because of this, I believe the ***overall*** hazard group cred-wtd ratio (.1718 here) would be the same as the overall raw even ratio (.1718). Unfortunately though, Couret & Venter don't show a detailed example of the calculations or elaborate to verify this.

CBT Spreadsheet Tips

In my opinion, the best approach on the CBT exam is to sort the table at the start directly. There isn't an automatic way to sort, but you can copy/paste the rows or reference them directly to create a sorted table. Then you can easily use SUMIF( ) formulas to lookup values for each quintile when doing the calculations.

In both steps 1 and 2, array formulas can be used to simplify the calculations. See the Excel version for the spreadsheet calculations.

Remember, when you use an array formula, press CTRL+SHIFT+ENTER to enter the formula as an array formula in Pearson Vue. If you enter it normally without pressing CTRL+SHIFT+ENTER, you'll get a #VALUE! error.

Source

Couret & Venter – pg. 81-83

More Practice

CAS 2015 – 5
CAS 2014 – 1
CAS 2013 – 3
CAS 2012 – 5
CAS 2011 – 2

Multi-Dimensional Credibility

Couret & Venter

Problem

Given the following information for Class 294, which is part of Hazard Group B:

Claim Counts by Injury Type - Class 294

Year	F	PT	TT
2014	5	15	600
2015	10	20	650
2016	5	20	660

- Hazard Group B ratio F:TT $V_B = .005$
- Hazard Group B ratio PT:TT $W_B = .010$

Variance of Injury-Type Ratio Across HG B Classes

Measure	V	W
VHM	.00026	.00043
EPV	.460	.880

- Covariance across classes for HG B Cov(V, W) = .00035

a. Calculate the multiplicative factors b and c that can be used with multi-dimensional credibility to estimate the F:TT ratio, v_i, for Class 294.

b. Calculate the predicted F:TT ratio, v_i, using multi-dimensional credibility and factors b and c from part a.

Solution Recipe

Part a – Multiplicative Factors

1) **Solve for the unconditional variance of each variable along the matrix diagonal, Var(V_i) and Var(W_i).**

$$\boxed{\mathrm{Var}(V_i) = \frac{EPV_V}{m_i} + VHM_V}$$

$$\mathrm{Var}(F:TT) = \frac{Exp.\,Process\,Var}{\#\,TT\,Claims_{Class\,i}} + Var\ Hypothetical\ Mean$$

$$\mathrm{Var}(V_i) = \frac{.460}{600+650+660} + .00026 = .000501$$

$$\mathrm{Var}(W_i) = \frac{.880}{600+650+660} + .00043 = .000891$$

2) **Set up the matrix equation with the covariance matrix to solve for the multiplicative factors.**

$$\underset{VHM_V \rightarrow}{}\boxed{\begin{pmatrix}\mathrm{Cov}(V_i, v_i)\\ \mathrm{Cov}(W_i, v_i)\end{pmatrix} = \begin{pmatrix}\mathrm{Var}(V_i) & \mathrm{Cov}(V_i, W_i)\\ \mathrm{Cov}(V_i, W_i) & \mathrm{Var}(W_i)\end{pmatrix} \cdot \begin{pmatrix} b_v \\ c_v\end{pmatrix}}$$

B = **A** x **X**

$$\begin{pmatrix}.00026\\ .00035\end{pmatrix} = \begin{pmatrix}.000501 & .00035\\ .00035 & .000891\end{pmatrix} \times \begin{pmatrix} b_v \\ c_v\end{pmatrix}$$

3) **Solve the system of equations for the multiplicative factors. See the Excel version to follow the Matrix calculations.**

In Excel

The solution for the factors is: **X** = **A**$^{-1}$ * **B**

A^{-1} = MINVERSE(array **A**)
X = MMULT(array **A**$^{-1}$, array **B**)

MINVERSE(A) - Returns the inverse of A
MMULT(A, B) - Multiplies matrices A and B

Note:
Hold CTRL + SHIFT + ENTER for array formulas

$$.00026 = .000501 b_v + .00035 c_v$$
$$.00035 = .00035 b_v + .000891 c_v$$
$$b_v \rightarrow .337$$
$$c_v \rightarrow .260$$

Part b – Predicted Class F:TT Ratio

4) **Calculate the frequency relativity for each injury type as a ratio to TT claims.**

$$\textit{Frequency Relativity} = \frac{\#\,\textit{Injury-type Claims}}{\#\,\textit{TT Claims}}$$

$$V_c = \frac{\#\,Fatal}{\#\,TT} = \frac{5+10+5}{600+650+660} = .0105$$

$$W_c = \frac{15+20+20}{600+650+660} = .0288$$

5) **Use the multiplicative factors to calculate the predicted frequency relativity for the injury type to TT claims. Remember that the multiplicative factors will vary by injury type.**

$$\boxed{v_c^{pred} = E[V] + b_v(V_c - E[V]) + c_v(W_c - E[W])}$$

$$v_{class}^{pred} = V_{HG} + b_v(V_{class} - V_{HG}) + c_v(W_{class} - W_{HG})$$

$$v_{class}^{pred} = .005 + .337(.0105 - .005) + .260(.0288 - .010)$$
$$= \boxed{.0117}$$

Discussion

One tricky thing to remember in step two is that Cov(V_i, v_i) is set to the VHM$_V$. In order to solve for the b_w and c_w parameters for the w_i ratio, we would use the matrix equation below and Cov(W_i, w_i) would be set to the VHM$_W$. This is how we get different credibility parameters for calculating the different predicted v_i, w_i, x_i and y_i ratios.

To solve for the w_i ratios we would use the formulas below:

$$VHM_W \rightarrow \boxed{\begin{pmatrix} \mathrm{Cov}(V_i, w_i) \\ \mathrm{Cov}(W_i, w_i) \end{pmatrix} = \begin{pmatrix} \mathrm{Var}(V_i) & \mathrm{Cov}(V_i, W_i) \\ \mathrm{Cov}(V_i, W_i) & \mathrm{Var}(W_i) \end{pmatrix} \cdot \begin{pmatrix} b_w \\ c_w \end{pmatrix}} \quad \Bigg| \quad \begin{pmatrix} .00035 \\ .00043 \end{pmatrix} = \begin{pmatrix} .000501 & .00035 \\ .00035 & .000891 \end{pmatrix} \times \begin{pmatrix} b_w \\ c_w \end{pmatrix}$$

<u>Couret & Venter Assumptions</u>

The uppercase V_i, W_i, X_i, and Y_i indicate the ***<u>observed</u>*** ratios for injury-types. The lowercase v_i, w_i, x_i, and y_i indicate the population ***<u>hypothetical mean</u>*** ratios for the i^{th} class.

Couret & Venter use some simplifying assumptions in the model, discussed on the bottom of pg. 78 in the paper. The observed ratio W_i is assumed to be the hypothetical mean w_i plus a random error term. Based on this simplifying assumption, we use Cov(W_i, w_i) = Cov(w_i, w_i) = VHM_W.

CBT Spreadsheet Tips

Solve the system of linear equations quickly using matrices. See the Excel version to follow the spreadsheet calculations

Here's a quick recap from your old Linear Algebra class:

In matrix notation, this system is $\mathbf{A} \times \mathbf{X} = \mathbf{B}$ where:

- **A** = Covariance Matrix
- **X** = The multiplicative factor array
- **B** = The matrix (array) of the covariances with v_i

To solve for **X** (factors), we need to get the inverse of **A**, denoted $\mathbf{A}^{-1}$.

The solution for the factors is:
$\mathbf{X} = \mathbf{A}^{-1} \times \mathbf{B}$

In Excel, use the following formulas:

MINVERSE(array A) - Returns the inverse of A
MMULT(array A, array B) - Multiplies matrices A and B

Note: Hold CTRL + SHIFT + ENTER for array formulas

Source

Couret & Venter – pg. 77-79

More Practice

CAS 2019 – 1