Introduction to Actuarial Data Manipulation with R – Workshop

CAS Spring Meeting 2024

Denis Dreano

2024-05-05

Denis Dreano

- Pricing actuary at Swiss Re since 2018
- Currently, Lead Actuarial Models and Architecture
  - Focus on data models for actuarial applications
- R application and package writer
- LinkedIn: https://www.linkedin.com/in/denis-dreano/

Why this Workshop

- Programming is part of the actuarial profession
- Actuaries lack formal training in programming
- This leads to inefficiencies and errors
- R is a popular language in the actuarial profession

Learning Objectives

1. Select the most appropriate R data structure to represent different types of actuarial concepts: vector, list, data frame, array, date, etc.
2. Write R scripts that load, manipulate and export insurance data, using state-of-the-art libraries: tidyverse, dplyr, lubridate, purrr, etc.
3. Perform simple standard actuarial procedures in R: aggregate claim/policy data, calculate summary ratios, calculate and develop claim triangles, etc.

Goal

Reproduce parts of a realistic actuarial process:

- Automobile Rate Indication in Appendix A from Werner and Modlin, *Basic Ratemaking* (Syllabus of Exam 5)
- See `References/` for the original source

By the end of the session you should have:

- A better knowledge of the R language and modern R libraries
- A better understanding of actuarial data manipulation in R

It will allow you to:

- Be more efficient when programming in R
- Automate more parts of your workflow
- Collaborate more effectively with application developers and data engineers

Disclaimers

- You will need some programming experience in R or in another language (Python, VBA, etc.)
  - If you're not familiar with R or need a refresher see: `Slides/r_refresher.pdf`
- I recommend using RStudio for the workshop
  - I will use RStudio for the workshop
  - I won't be able to help you with other environments
- The workshop will focus on data structures and data manipulation
  - No coverage of charting libraries report building, etc.

Organization

Three parts:

- 12:30pm Module 1: Rate Indication
- 1:30pm Module 2: Earned Premium
- 2:30pm Module 3: Claim Development
- 3:30pm Discussion and Conclusion

Each module will be organized as follows:

- 10 minutes of introduction
- 40 minutes of exercises
  - Open the "question" R scripts in `Workshop/` folder
  - Try to solve the questions individually or in a small group
  - Don't spend too much time on each question. After a few minutes, look at the "solution" script and move on to the next question.
- 10 minutes of discussions

Resources

- `Data/`: Raw data you will need to load
- `References/`: Original source for the rate indication
- `Slides/`: This slide and other resources
- `Workshop/`: The exercises and the solutions
- `Cheatsheets/`: Helpful R cheatsheets for reference