Module 3 – Loss Development

Actuarial Data Manipulation with R – CAS Spring Meeting 2024

Denis Dreano
2024-05-05

Content

Goals: Using claim transaction data, calculate the loss development triangles and estimate ultimate losses.

Content:

- Data Manipulation with `dplyr` and `tidyr`:
    - `left_join()`
    - `pivot_wider()`
    - `crossing()`
- Matrices
    - subsetting
    - `apply()`
- Iterations with `purrr`:
    - `pmap_dbl()`

Data Manipulation with `dplyr`: `left_join()`

```
library(dplyr)
claim_df <- tibble(policy_id = c(1, 1, 2, 2, 3),
                   claim_amount = c(100, 200, 300, 400, 500))
policy_df <- tibble(policy_id = c(1, 2, 3),
                    policy_type = c("auto", "home", "auto"))
left_join(claim_df, policy_df, by = "policy_id")
```

```
# A tibble: 5 x 3
  policy_id claim_amount policy_type
      <dbl>        <dbl> <chr>
1         1          100 auto
2         1          200 auto
3         2          300 home
4         2          400 home
5         3          500 auto
```

Data Manipulation with `tidyr`: `crossing()`

```
library(tidyr)
crossing(
  accident_year = 2010:2011,
  maturity = seq(12, 36, by = 12)
)

# A tibble: 6 x 2
  accident_year maturity
          <int>    <dbl>
1          2010       12
2          2010       24
3          2010       36
4          2011       12
5          2011       24
6          2011       36
```

Data Manipulation with `tidyr`: `pivot_wider()`

```r
library(tidyr)
triangle_df <- tibble(
  accident_year = c(2010, 2010, 2010, 2011, 2011),
  maturity = c(12, 24, 36, 12, 24),
  cumulative_paid = c(100, 200, 300, 400, 500)
)
triangle_df |>
  pivot_wider(names_from = maturity, values_from = cumulative_paid)

# A tibble: 2 x 4
  accident_year  `12`  `24`  `36`
          <dbl> <dbl> <dbl> <dbl>
1          2010   100   200   300
2          2011   400   500    NA
```

Matrices: Creation

```
triangle_matrix <- matrix(
  c(100, 200, 300, 400, 500, NA), nrow = 2
)
triangle_matrix
```

```
     [,1] [,2] [,3]
[1,]  100  300  500
[2,]  200  400   NA
```

```
dimnames(triangle_matrix) <- list(
  accident_year = c(2010, 2011),
  maturity = c(12, 24, 36)
)
triangle_matrix
```

```
            maturity
accident_year  12  24  36
        2010 100 300 500
        2011 200 400  NA
```

```
rownames(triangle_matrix)
```

```
[1] "2010" "2011"
```

```
colnames(triangle_matrix)
```

```
[1] "12" "24" "36"
```

```
nrow(triangle_matrix)
```

```
[1] 2
```

```
ncol(triangle_matrix)
```

```
[1] 3
```

Matrices: Subsetting

```
triangle_matrix[1, 2]
```
```
[1] 300
```
```
triangle_matrix["2010", "24"]
```
```
[1] 300
```
```
triangle_matrix[1, ]
```
```
 12  24  36
100 300 500
```
```
triangle_matrix[, 2]
```
```
2010 2011
 300  400
```

```
triangle_matrix[1:2, 2:3]
```
```
          maturity
accident_year  24  36
       2010 300 500
       2011 400  NA
```
```
triangle_matrix[, -1]
```
```
          maturity
accident_year  24  36
       2010 300 500
       2011 400  NA
```
```
triangle_matrix[, -nrow(triangle_matrix)]
```
```
          maturity
accident_year  12  36
       2010 100 500
       2011 200  NA
```

```r
apply(triangle_matrix, 1, sum)
```

```
2010 2011
 900   NA
```

```r
apply(triangle_matrix, 2, sum)
```

```
 12  24  36
300 700  NA
```

```
library(purrr)
policy_df <- tibble(policy_id = c(1, 2, 3),
                    policy_type = c("auto", "home", "auto"))
calculate_premium <- function(policy_type, ...) {
  switch(policy_type,
         auto = 100,
         home = 200)
}
pmap_dbl(policy_df, calculate_premium)
```

```
[1] 100 200 100
```

References

- Cheat sheets in `References/`
  - `data-transformation.pdf`
  - `base-r-cheat-sheet.pdf`
  - `purrr.pdf`
- Wickham, H. et al. *R for Data Science* (r4ds.hadley.nz)
  - Chap. 3 Data Transformation
  - Chap. 5 Data Tidying
- Wickham, H. *Advanced R* (adv-r.hadley.nz)
  - Chap. 3 Vectors
  - Chap. 4 Subsetting
  - Chap. 9 Functionals