2
Alguns Estimadores para IBNR

O objetivo desse capítulo é apresentar alguns métodos existentes para a determinação da reserva de IBNR. Todos os modelos a seguir baseiam-se no *triângulo de desenvolvimento*, que será definido a seguir.

O *triângulo de desenvolvimento* é de uma matriz. Os dados para a construção dessa matriz são os sinistros ocorridos e avisados até a data presente. Neste trabalho serão usados dados mensais durante um período de m meses. Define-se *data inicial* como a data de início das informações, ou seja, m meses antes da data presente. O *mês de ocorrência* de um certo sinistro será denotado por i se ele ocorreu i meses após a data inicial e o *mês de aviso* de um certo sinistro será chamado de j se ele foi avisado j meses após a data de ocorrência.

Por exemplo, se a data presente for 01 de janeiro de 2000 e um sinistro ocorreu no dia 05 de fevereiro de 2000 e foi avisado no dia 20 de fevereiro de 2005, o seu mês de ocorrência será 2 e seu mês aviso será 1. Pode-se observar que o mês de ocorrência varia entre 1 e m enquanto o mês de aviso varia entre 1 e $m-i+1$, onde i representa o mês de ocorrência, uma vez que se um sinistro ocorreu no mês i e só há informações até m a data de aviso máxima para esse sinistro é $m - i + 1$.

Seja $Y_{i,j}$ a soma dos valores dos sinistros com mês de ocorrência i e mês de aviso j. Como já foi comentado, só são conhecidos os valores de $Y_{i,j}$ para $1 \leq i \leq m$ e $1 \leq j \leq m - i + 1$. O *triângulo de desenvolvimento* é uma matriz $m \times m$ tal que a posição (i, j), para $1 \leq i \leq m$ e $1 \leq j \leq m - i + 1$, recebe o valor de $Y_{i,j}$ e as demais entradas são desconhecidas. Em alguns modelos é comum encontrar o termo *triângulo acumulado*, que trata-se da matriz com entradas $A_{i,j} = \sum_{k=1}^{j} Y_{i,k}$. Cada posição do triângulo acumulado representa o total de sinistros ocorridos em i e avisados em até j meses.

Exemplo 2.0.1. *Este exemplo apresenta um possível triângulo de desenvolvimento, Y, e seu respectivo triângulo acumulado, A. No caso $m = 10$.*

$$Y = \begin{bmatrix} 29660 & 38124 & 34786 & 14737 & 11401 & 6805 & 3161 & 7653 & 352 & 4586 \\ 11136 & 48999051 & 23075 & 9267 & 5873 & 352 & 3243 & 2187 & 2830 & \\ 30883 & 8302860 & 45778 & 22091 & 9073 & 18068 & 4344 & 1559 & & \\ 39943 & 61988574 & 26268 & 18381 & 5697 & 4491 & 2461 & & & \\ 60408574 & 7508098 & 24258 & 15868 & 5831 & 4313 & & & & \\ 37071 & 61697622 & 23545 & 15264 & 10593 & & & & & \\ 73483813 & 79448574 & 32452 & 31019 & & & & & & \\ 52241432 & 79178574 & 34127 & & & & & & & \\ 60503336 & 58051908 & & & & & & & & \\ 46014 & & & & & & & & & \end{bmatrix}$$

$$A = \begin{bmatrix} 29660 & 67784 & 102570 & 117307 & 128708 & 135513 & 138674 & 146327 & 146679 & 151265 \\ 11136 & 49010187 & 49033262 & 49042529 & 49048402 & 49048754 & 49051997 & 49054184 & 49057014 & \\ 30883 & 8333743 & 8379521 & 8401612 & 8410685 & 8428753 & 8433097 & 8434656 & & \\ 39943 & 62028517 & 62054785 & 62073166 & 62078863 & 62083354 & 62085815 & & & \\ 60408574 & 67916672 & 67940930 & 67956798 & 67962629 & 67966942 & & & & \\ 37071 & 61734693 & 61758238 & 61773502 & 61784095 & & & & & \\ 73483813 & 152932387 & 152964839 & 152995858 & & & & & & \\ 52241432 & 131420006 & 131454133 & & & & & & & \\ 60503336 & 118555244 & & & & & & & & \\ 46014 & & & & & & & & & \end{bmatrix}$$

A posição $Y_{2,4} = 9267$ fornece o total de sinistros com mês de ocorrência 2 e mês de aviso 4, ou seja, esse é o valor total dos sinistros ocorridos 2 meses após a data inicial e avisados 4 meses depois da sua ocorrência. Já a posição $A_{2,4} = 49042529$ fornece o total de sinistros com mês de ocorrência 2 e com mês de aviso menor ou igual a 4.

Neste trabalho, a letra Y será usada para o triângulo de desenvolvimento e a letra A para o triângulo acumulado. Dessa forma, ao longo do texto $Y_{i,j}$ e $A_{i,j}$ representam as entradas (i, j) do triângulo de desenvolvimento e acumulado, respectivamente.

A seguir serão apresentados três estimadores baseados em diferentes modelos probabilísticos. Os detalhes de cada estimador serão analisados em cada uma das próximas seções.

2.1 Primeiro Estimador

O primeiro método apresentado é o mais utilizado entre as seguradoras, chamado *Chain Ladder*. Trata-se de uma seqüência de cálculos realizados a partir do triângulo acumulado, cujas entradas são os únicos dados levados em consideração pelo estimador.

Esse método é realizado sem o conhecimento de um modelo probabilístico que justifique sua implementação. Por causa dessa lacuna alguns autores procuram modelos que possam estar por trás de tais contas, como fez Thomas Mack em [10]. Nessa seção será apresentada a forma de usar o *Chain Ladder* para estimar o IBNR, as questões probabilísticas não serão discutidas nesse trabalho e podem ser encontradas com detalhes em [10] e [11].

Apresentação do Modelo

A estimativa do IBNR pelo método do *Chain Ladder* sairá da estimativa das entradas desconhecidas do triângulo acumulado, mas precisamente da estimativa da sua última coluna. Como já foi mencionado, $A_{i,j}$ representa a posição (i, j) do triângulo acumulado. Esses valores são conhecidos sempre que $1 \leq i \leq m$ e $1 \leq j \leq m - i + 1$, onde m representa o número de meses de observação. Supondo que todo sinistro demora no máximo m meses para ser avisado, é verdade que $IBNR = \sum_{i=2}^{m} (A_{i,m} - A_{i,m-i+1})$. Então, é preciso estimar os elementos $A_{i,m}$ para poder determinar o valor do IBNR pela equação anterior.

A principal idéia desse método é que a razão $\frac{A_{i,j+1}}{A_{i,j}}$ é uma variável aleatória com média f_j, ou seja, o fator de crescimento entre as colunas j e $j + 1$ não depende da linha i. O primeiro passo será, utilizando os valores conhecidos do triângulo, estimar os fatores f_j para $1 \leq j \leq m - 1$. Isso pode ser feito de diversas maneiras, dando origem a diferentes formas de realizar o *Chain Ladder*.

Observe que o triângulo fornece alguns fatores, por exemplo para $j = 1$ existe uma amostra de tamanho $m-1$: $\left\{\frac{A_{1,2}}{A_{1,1}}, \frac{A_{2,2}}{A_{2,1}}, ..., \frac{A_{m-1,2}}{A_{m-1,1}}\right\}$. Com esses dados é possível estimar os f_j pela média amostral, pelo maior fator, pelo menor fator, pela mediana ou por qualquer outra estatística. A escolha da estatística causa uma mudança no estimador tornando-o mais ou menos otimista. Nesse trabalho o estimador de f_j, para $1 \leq j \leq m - 1$, está exposto na equação a seguir, trata-se do estimador não tendencioso e de variância mínima, como mostrou Thomas Mack em [11].

$$\hat{f}_j = \frac{\sum_{i=1}^{m-j} A_{i,j+1}}{\sum_{i=1}^{m-j} A_{i,j}} \tag{2-1}$$

Com os fatores já estimados é possível encontrar uma estimativa para as entradas desconhecidas do triângulo acumulado. Para $2 \leq i \leq m$ faça $\hat{A}_{i,m-i+2} = A_{i,m-i+1}\hat{f}_{m-i+1}$, em seguida, para $3 \leq i \leq m$ faça $\hat{A}_{i,m-i+3} = \hat{A}_{i,m-i+2}\hat{f}_{m-i+2}$ e assim por diante. Dessa forma todos os valores desconhecidos do triângulo podem ser estimados. A equação 2-2 apresenta a estimativa para os elementos da coluna m. Lembre-se que $2 \leq i \leq m$.

$$\hat{A}_{i,m} = A_{i,m-i+1} \prod_{j=m-i+1}^{m-1} \hat{f}_j \tag{2-2}$$

Apresentação do Estimador

As equações 2-1 e 2-2 permitem estimar os elementos da coluna m. Com esse resultado, o primeiro estimador para o IBNR é:

$$I B \hat{N} R_1 = \sum_{i=2}^{m} A_{i,m-i+1} \left(\left(\prod_{j=m-i+1}^{m-1} \hat{f}_j \right) - 1 \right) \tag{2-3}$$

2.2 Segundo Estimador

O modelo que será exposto nesta seção foi proposto por Hertig [5] e novamente apresentado por de Jong em [2]. O método é bem parecido com o *Chain Ladder*, uma vez que se baseia em fatores de crescimento para o triângulo de desenvolvimento. A principal diferença está na forma de calcular os fatores de cada coluna.

Apresentação do Modelo

Como já foi comentado na seção anterior, $A_{i,j}$ será conhecido sempre que $1 \leq i \leq m$ e $1 \leq j \leq m - i + 1$, onde m representa o número de meses de observação. Supondo que todo sinistro demora no máximo m meses para ser avisado, o IBNR na data presente será determinado por $IBNR = \sum_{i=2}^{m} (A_{i,m} - A_{i,m-i+1})$. A principal suposição do modelo é:

$$A_{i,m} - A_{i,m-i+1} = A_{i,m-i+1}(\mathrm{e}^{g_i} - 1) \quad \Rightarrow \quad g_i = \ln \left(\frac{A_{i,m}}{A_{i,m-i+1}} \right)$$

Dessa forma é possível escrever:

$$IBNR = \sum_{i=2}^{m} \left(A_{i,m-i+1}(\mathrm{e}^{g_i} - 1) \right) \tag{2-4}$$

A única variável desconhecida na equação 2-4 é g_i, e a sua média será estimada utilizando os valores conhecidos do triângulo acumulado. Defina a variável aleatória $\delta_{i,j}$ da seguinte maneira:

$$\delta_{i,j} = \ln \left(\frac{A_{i,j}}{A_{i,j-1}} \right) \tag{2-5}$$

Veja que $\delta_{i,j}$ representa a razão entre as colunas do triângulo. Suponha que $E[\delta_{i,j}] = \mu_j$, ou seja, a sua média depende apenas do mês de aviso e não do mês ocorrência. Observe que podemos escrever g_i em função dos $\delta_{i,j}$'s:

$$g_i = \ln\left(\frac{A_{i,m}}{A_{i,m-i+1}}\right)$$
$$g_i = \ln\left(\frac{A_{i,m}}{A_{i,m-1}}\frac{A_{i,m-1}}{A_{i,m-2}}...\frac{A_{i,m-i+2}}{A_{i,m-i+1}}\right)$$
$$g_i = \ln\left(\frac{A_{i,m}}{A_{i,m-1}}\right) + \ln\left(\frac{A_{i,m-1}}{A_{i,m-2}}\right) + ... + \ln\left(\frac{A_{i,m-i+2}}{A_{i,m-i+1}}\right)$$
$$g_i = \delta_{i,m} + \delta_{i,m-1} + ... + \delta_{i,m-i+2}$$

Então para estimar $E[g_i]$ basta estimar μ_j para $m - i + 2 \leq j \leq m$, o que pode ser feito pela média amostral:

$$\hat{\mu_j} = \frac{1}{m-j+1}\sum_{k=1}^{m-j+1}\delta_{k,j} \tag{2-6}$$

Observe que como o desejado é estimar g_i para $2 \leq i \leq m$, serão calculados $\hat{\mu_j}$ para $2 \leq j \leq m$. Nesses cálculos para cada j só serão usados os $\delta_{i,j}$ com $1 \leq i \leq m-j+1$, ou seja, os $A_{i,j}$ tais que $1 \leq j \leq m$ e $1 \leq i \leq m-j+1$. Com isso é garantido que só serão usadas as entradas conhecidas do triângulo acumulado.

Então a estimativa para a média de g_i, que será chamada de $\hat{g_i}$, será:

$$\hat{g_i} = \sum_{j=m-i+2}^{m}\hat{\mu_j} \tag{2-7}$$

Apresentação do Estimador

Juntando as equações 2-4, 2-5, 2-6 e 2-7 o segundo estimador para o IBNR é:

$$IB\hat{N}R_2 = \sum_{i=2}^{m}\left(A_{i,m-i+1}(\mathrm{e}^{\hat{g_i}} - 1)\right) \tag{2-8}$$

Veja que $\mathrm{e}^{\hat{g_i}}$ desempenha o mesmo papel da multiplicação dos fatores no modelo de *Chain Ladder*. A forma pela qual ele é determinado é que difere os dois modelos.

2.3 Terceiro Estimador

Nessa seção será exposto o modelo para o cálculo do IBNR apresentado por Doray [4]. Esse modelo terá como base de dados o triângulo de desenvolvimento e não mais o triângulo acumulado, como os anteriores. Mais uma vez será usada a hipótese de que qualquer sinistro demora no máximo m meses pra ser avisado. Lembrando $Y_{i,j}$ representa os elementos (i,j) do triângulo de desenvolvimento pode-se afirmar que o valor do IBNR é a soma dos $Y_{i,j}$ desconhecidos.

$$IBNR = \sum_{i=2}^{m} \sum_{j=m-i+2}^{m} Y_{i,j} \tag{2-9}$$

Então uma estimativa dos $Y_{i,j}$ desconhecidos fornecerá uma estimativa para o $IBNR$. Baseado em algumas hipóteses, este modelo fornece o estimador não tendencioso de variância mínima para o IBNR.

Apresentação do Modelo

A suposição principal desse modelo é que $Y_{i,j}$ pode ser escrito da seguinte forma:

$$Y_{i,j} = R_i C_j E_{i,j} \tag{2-10}$$

onde R_i representa o efeito do mês de ocorrência, C_j representa o efeito do mês de aviso e $E_{i,j}$ são variáveis aleatórias independentes de distribuição lognormal com parâmetro $(0,\sigma^2)$, $E_{i,j} \sim \text{Lognormal}(0,\sigma^2)$. Logo pode-se afirmar que $Y_{i,j} \sim \text{Lognormal}(\mu_{ij},\sigma^2)$. Definindo $Z_{i,j} = \ln(Y_{i,j})$ tem-se $Z_{i,j} \sim \text{N}(\mu_{ij},\sigma^2)$, variável aleatória normal de média μ_{ij} e variância σ^2. Além disso é possível escrever

$$Z_{i,j} = \alpha_i + \beta_j + \epsilon_{i,j} \tag{2-11}$$

onde $\alpha_i = \ln(R_i)$, $\beta_j = \ln(C_j)$ e $\epsilon_{i,j} = \ln(E_{i,j})$. Dessa forma $\epsilon_{i,j} \sim \text{N}(0,\sigma^2)$, consequentemente $E[Z_{i,j}] = \mu_{ij} = \alpha_i + \beta_j$ e $Var[Z_{i,j}] = \sigma^2$.

A equação 2-11 vale para todos os $Z_{i,j}$, conhecidos ou não (observe que $Z_{i,j}$ será conhecido sempre que $Y_{i,j}$ for conhecido). A partir dos $Z_{i,j}$ conhecidos serão estimados α_i e β_j . Com esse resultado será possível estimar $E[IBNR]$.

Partindo da equação 2-11 e usando somente os $Z_{i,j}$ conhecidos pode-se escrever a seguinte forma matricial

$$Z = Xv + \epsilon, \tag{2-12}$$

onde

$$\begin{aligned} Z^t &= (Z_{1,1}, Z_{1,2}, ..., Z_{1,m}, Z_{2,1}, ..., Z_{2,m-1}, ..., Z_{m,1}) \\ v^t &= (\alpha_1, \alpha_2, ..., \alpha_m, \beta_2, \beta_3, ..., \beta_m) \\ \epsilon^t &= (\epsilon_{1,1}, \epsilon_{1,2}, ..., \epsilon_{1,m}, \epsilon_{2,1}, ..., \epsilon_{2,m-1}, ..., \epsilon_{m,1}) \end{aligned}$$

A matriz X surge naturalmente da relação matricial acima. Ela está bem definida e pode ser representada da seguinte maneira.

$$X = \begin{pmatrix} \boxed{\begin{matrix} 1&0&0&...&0&0 \\ 1&0&0&...&0&0 \\ 1&0&0&...&0&0 \\ \vdots&\vdots&\vdots&\vdots&\vdots&\vdots \\ 1&0&0&...&0&0 \end{matrix}} & \boxed{\begin{matrix} 0&0&0&...&0&0 \\ 1&0&0&...&0&0 \\ 0&1&0&...&0&0 \\ \vdots&\vdots&\vdots&\vdots&\vdots&\vdots \\ 0&0&0&...&0&1 \end{matrix}} \\ \boxed{\begin{matrix} 0&1&0&...&0&0 \\ 0&1&0&...&0&0 \\ 0&1&0&...&0&0 \\ \vdots&\vdots&\vdots&\vdots&\vdots&\vdots \\ 0&1&0&...&0&0 \end{matrix}} & \boxed{\begin{matrix} 0&0&0&...&0&0 \\ 1&0&0&...&0&0 \\ 0&1&0&...&0&0 \\ \vdots&\vdots&\vdots&\vdots&\vdots&\vdots \\ 0&0&0&...&1&0 \end{matrix}} \\ \begin{matrix} \vdots&\vdots&\vdots&\vdots&\vdots&\vdots \\ \vdots&\vdots&\vdots&\vdots&\vdots&\vdots \end{matrix} & \begin{matrix} \vdots&\vdots&\vdots&\vdots&\vdots&\vdots \\ \vdots&\vdots&\vdots&\vdots&\vdots&\vdots \end{matrix} \\ \boxed{\begin{matrix} 0&0&0&...&1&0 \\ 0&0&0&...&1&0 \end{matrix}} & \boxed{\begin{matrix} 0&0&0&...&0&0 \\ 1&0&0&...&0&0 \end{matrix}} \\ \boxed{\begin{matrix} 0&0&0&...&0&1 \end{matrix}} & \boxed{\begin{matrix} 0&0&0&...&0&0 \end{matrix}} \end{pmatrix}$$

Veja que X é uma matriz $(\frac{m(m+1)}{2} \times 2m - 1)$. Observe que o vetor v não possui o elemento β_1, isso deve-se ao fato de que β_1 será zero para que a matriz de regressão não seja singular, veja [4].

As estimativas para v e σ^2 apresentadas em [4] são:

$$\hat{v} = (X^t X)^{-1} X^t Z \tag{2-13}$$

$$\hat{\sigma}^2 = \frac{(Z - X\hat{v})^t (Z - X\hat{v})}{n-p} = \frac{SS_z}{n-p} \tag{2-14}$$

onde $n = \frac{m(m+1)}{2}$ (dimensão do vetor Z) e $p = 2m - 1$ (dimensão do vetor v).

Agora que v e σ^2 já foram estimados é possível usar esses valores para determinar os $Z_{i,j}$ desconhecidos, que serão arrumados em um vetor Z^u como poderá ser visto a seguir. A letra u representa a palavra em inglês *unknown*.

Partindo da equação 2-11 e usando somente os $Z_{i,j}$ desconhecidos é possível, análogo ao que foi feito acima, escrever a seguinte forma matricial.

$$Z^u = Bv + \epsilon \tag{2-15}$$

onde

$$(Z^u)^t = (Z_{2,m}, Z_{3,m-1}, Z_{3,m}, Z_{4,m-2}, Z_{4,m-1}, Z_{4,m}, ..., Z_{m,m})$$

A matriz B também surge naturalmente da relação matricial e pode ser representada da seguinte maneira.

$$B = \begin{pmatrix} \boxed{0\ 1\ 0\ 0\ ...\ 0} & \boxed{0\ 0\ ...\ 0\ 0\ 1} \\ \boxed{\begin{matrix} 0\ 0\ 1\ 0\ ...\ 0 \\ 0\ 0\ 1\ 0\ ...\ 0 \end{matrix}} & \boxed{\begin{matrix} 0\ 0\ ...\ 0\ 1\ 0 \\ 0\ 0\ ...\ 0\ 0\ 1 \end{matrix}} \\ \boxed{\begin{matrix} 0\ 0\ 0\ 1\ ...\ 0 \\ 0\ 0\ 0\ 1\ ...\ 0 \\ 0\ 0\ 0\ 1\ ...\ 0 \end{matrix}} & \boxed{\begin{matrix} 0\ 0\ ...\ 1\ 0\ 0 \\ 0\ 0\ ...\ 0\ 1\ 0 \\ 0\ 0\ ...\ 0\ 0\ 1 \end{matrix}} \\ \vdots & \vdots \\ \vdots & \vdots \\ \boxed{\begin{matrix} 0\ 0\ 0\ 0\ ...\ 1 \\ 0\ 0\ 0\ 0\ ...\ 1 \\ \vdots \\ 0\ 0\ 0\ 0\ ...\ 1 \end{matrix}} & \boxed{\begin{matrix} 1\ 0\ ...\ 0\ 0\ 0 \\ 0\ 1\ ...\ 0\ 0\ 0 \\ \vdots \\ 0\ 0\ ...\ 0\ 0\ 1 \end{matrix}} \end{pmatrix}$$

A dimensão da matriz B é $(\frac{m(m-1)}{2} \times 2m - 1)$.

Define-se Z_i^u como sendo a i-ésima posição do vetor Z^u, b_i a i-ésima linha da matriz B e ϵ_i a i-ésima posição do vetor ϵ. Então é verdade que:

$$Z_i^u = b_i v + \epsilon_i \tag{2-16}$$

Apresentação do Estimador

O objetivo de Doray [4] no artigo é determinar o estimador UMVUE (*uniformly minimum variance unbiased estimator*) para a ḿedia do IBNR, também conhecido como *o melhor estimador não tendencioso*. Em seu artigo [4] Doray expões os cálculos e a demontração de que o único estimador com essas características é:

$$I\hat{BNR}_3 = {}_0F_1\left(\frac{n-p}{2}; \frac{(Z - X\hat{v})^t(Z - X\hat{v})}{4}\right) \sum_{i=1}^{\frac{m(m-1)}{2}} \mathrm{e}^{b_i\hat{v}} \tag{2-17}$$

onde ${}_0F_1(\alpha; z)$ é a função hipergeométrica, que é definida da seguinte maneira:

$${}_0F_1(\alpha; z) = \sum_{j=0}^{\infty} \frac{z^j}{j!(\alpha)_j} \text{ com } (\alpha)_j = \begin{cases} \alpha(\alpha+1)...(\alpha+j-1), & se\ j \geq 1 \\ 1, & se\ j = 0 \end{cases}$$

Vale a pena lembrar que para o estimador de um certo parâmetro ser um UMVUE é necessário e suficiente que ele seja não tendencioso e que entre todos os estimadores não tendenciosos desse parâmetro ele seja aquele com variância mínima. Para um estudo mais aprofundado sobre este estimador veja [12] ou [1].