

## 4. Metodologia

A verificação da existência de diferença na taxa de mortalidade de participantes que abandonam um plano de seguro de vida ou previdência complementar será realizada mediante a comparação entre as probabilidades de morte por idade da população que deixou o plano e as da população do grupo remanescente.

O objetivo deste trabalho vai além da investigação quanto a diferenças entre as probabilidades de morte dos dois citados grupos. É de interesse também verificarmos se há diferenças nestas probabilidades de morte quando passamos a considerar o tempo decorrido desde o ano em que o indivíduo deixou o plano, ou seja, se pode-se concluir que a probabilidade de morte no 1º ano após a saída do plano é diferente da probabilidade de morte no 2º ano, e assim sucessivamente, até o 4º ano após o indivíduo ter deixado o plano. O ano de saída do plano é também conhecido tecnicamente como período de seleção.

A teoria econômica prediz que caso haja uma diferença nas taxas de mortalidade entre os grupos, a mesma deverá ser observada nos primeiros anos após a mudança de status, ou seja, nos primeiros anos após o indivíduo ter deixado o seu plano. Decorrido algum tempo deste acontecimento, a tendência é a de que as taxas voltem ao padrão inicial, que é o do grupo remanescente nos planos.

As probabilidades de morte são representadas pela função  $q_x$  da tábua de mortalidade. Para ser possível a comparação entre os diferentes anos de saída do plano, utilizaremos os conceitos de tábua seleta de mortalidade, a qual apresenta os valores das probabilidades de morte segregados pelo período de seleção. Abaixo serão apresentados maiores detalhes sobre a construção de uma tábua de mortalidade e suas funções.

A comparação entre as probabilidades de morte do grupo de pessoas que deixaram o plano com as do grupo remanescente será feita por meio da razão entre as probabilidades de morte dos dois grupos e será segregada pela cobertura do plano, sobrevivência ou morte, por sexo e pelo período de seleção. Na subseção 4.2 será apresentada a metodologia utilizada para a modelagem das razões utilizadas.

#### 4.1. Tábuas de mortalidade

Uma tábua de mortalidade consiste em uma tabela com registros do número de pessoas que vão atingindo determinada idade, partindo de um grupo inicial até a extinção do mesmo. A tábua construída desta forma é chamada tábua de vida de coorte, e é baseada na experiência de mortalidade de uma coorte específica. Tábuas deste tipo são muito difíceis de serem construídas, pois necessitam do acompanhamento da coorte por um longo período de tempo, que pode chegar a mais de um século.

Outra forma para a construção da tábua de mortalidade é a utilização do conceito de coorte sintética, no qual é observada a experiência da população durante um pequeno período de tempo onde a mortalidade tenha mantido padrões semelhantes. Neste caso são utilizadas informações de várias coortes, referentes a cada idade exata no ano de construção da tábua.

Independente do método utilizado para a sua construção, a tábua de mortalidade pode ser integralmente obtida a partir da função  ${}_nq_x$ , que representa a estimativa da probabilidade de que um indivíduo com idade exata  $x$  morra durante o intervalo  $(x, x+n)$ . A partir desta função obtém-se facilmente as demais funções básicas de uma tábua de mortalidade, que são a quantidade de sobreviventes à idade exata  $x$  ( $l_x$ ), o número de óbitos entre as idades  $x$  e  $x+n$  ( ${}_nd_x$ ), o número médio de sobreviventes entre  $x$  e  $x+n$  ( ${}_nL_x$ ), o número médio de anos que a população de idade  $x$  irá sobreviver ( $T_x$ ) e a expectativa de vida futura do indivíduo à idade exata  $x$  ( $e^0_x$ ).

A função  $l_x$  indica o número de sobreviventes à idade exata  $x$ , a partir de um valor inicial ou raiz  $l_0$ , normalmente definido como sendo 10.000 ou 100.000. A partir do valor inicial estabelecido os próximos valores são obtidos da seguinte forma<sup>1</sup>:

$$l_x = l_{x-1} - l_{x-1} * q_x.$$

O número de mortes ocorridas em determinado intervalo de  $n$  anos é representado por  ${}_nd_x$  e pode ser encontrado por meio da seguinte expressão:

$${}_nd_x = l_x - {}_nl_x.$$

---

<sup>1</sup> Quando o subscrito ( $n$ ) for omitido nas funções da tábua de mortalidade, subentende-se que  $n=1$ .

A função  ${}_nL_x$ , por sua vez, corresponde ao número de pessoas que, tendo completado  $x$  anos de idade, não completaram  $x+n$  anos de idade ainda. Se considerarmos a hipótese de que a ocorrência de mortes entre as idades  $x$  e  $x+1$  possui uma distribuição uniforme temos que

$$L_x = l_{x+1/2}, \text{ sendo } l_{x+1/2} = \frac{1}{2} * (l_x + l_{x+1}).$$

Portanto,  $L_x$  representa o número médio de sobreviventes entre as idades sucessivas  $x$  e  $x+1$ .

Por fim, as funções  $T_x$ ,  $e_x$  estão relacionadas à quantidade de anos vividos pela população, sendo que a primeira corresponde ao número de anos que, a partir da idade exata  $x$ , viverão todos os componentes do grupo (população) até que o mesmo se extinga. É obtida por:

$$T_x = \sum_{u=x}^{w-1} L_u.$$

A segunda função,  $e_x^0$ , representa uma média deste valor, ou seja, indica o número médio de anos que ainda serão vividos por um único indivíduo com idade exata  $x$ . É também conhecida como esperança de vida à idade  $x$  e representa uma das principais funções da tábua de mortalidade. Pode ser obtida pela seguinte fórmula:

$$e_x^0 = \frac{T_x}{l_x}.$$

No processo de construção da tábua de mortalidade é de suma importância fazer a conversão entre as taxas centrais de mortalidade ( ${}_n m_x$ ) em probabilidades de morte ( ${}_n q_x$ ). As taxas centrais de mortalidade são expressas por:

$${}_n m_x = \frac{{}_n d_x}{{}_n L_x} = \frac{l_x - l_{x+n}}{{}_n L_x}.$$

Já as probabilidades de morte são representadas por:

$${}_n q_x = \frac{{}_n d_x}{l_x} = \frac{l_x - l_{x+n}}{l_x}.$$

Como as duas funções são apresentadas em função de  $l_x$ , se conhecermos a forma funcional de  $l_x$  poderemos obter uma expressão que relaciona  ${}_n m_x$  e  ${}_n q_x$ . A suposição mais comum a respeito da função  $l_x$  é a de que a mesma varia linearmente, o que equivale a dizer que as mortes se distribuem uniformemente dentro do intervalo  $(x, x+n)$ . Com base nesta suposição, temos que:

$${}_n q_x = \frac{2n \cdot m_x}{2 + n \cdot m_x}.$$

Conforme dito no início desta seção, toda a tábua de mortalidade pode ser derivada a partir da estimativa da probabilidade de morte de um indivíduo com idade exata  $x$  no intervalo  $(x, x+n)$ . Desta forma, atenção especial deve ser dada à estimação da função  $q_x$ .

O número de óbitos ocorridos em determinada população para uma idade  $x$  pode ser considerado variável aleatória com distribuição binomial,  $\text{Bin}(E_x, q_x)$ , com o parâmetro de tamanho conhecido,  $E_x$ , a população exposta ao risco, e o parâmetro de probabilidade,  $q_x$ , desconhecido e a ser estimado (BELTRÃO E SUGAHARA, 2005). A função de verossimilhança para a binomial, dado  $\{E_x\}_{x=0, \dots, w}$ , a exposição e  $\{O_x\}_{x=0, \dots, w}$ , os óbitos, com  $q_x$  como taxa correspondente, é:

$$L(q) = \prod_{x=0}^w \binom{E_x}{O_x} (q_x)^{O_x} (1-q_x)^{E_x - O_x}$$

$$L(q) = \prod_{x=0}^w \binom{E_x}{O_x} \left( \frac{q_x}{1-q_x} \right)^{O_x} (1-q_x)^{E_x}$$

$$L(q) = \exp \left[ \sum_{x=0}^w \ln \binom{E_x}{O_x} + O_x \ln \left( \frac{q_x}{1-q_x} \right) + E_x \ln(1-q_x) \right]$$

Da função de verossimilhança acima podemos obter o estimador de máxima verossimilhança (EMV) de  $q_x$ , que é dado por:

$$\text{EMV}(q_x) = \frac{O_x}{E_x}.$$

Até o presente momento tratamos da construção da tábua de mortalidade sem levar em consideração os períodos de seleção que serão utilizados neste trabalho, os quais se referem ao ano em que o indivíduo abandonou o seu plano de seguro de vida ou de previdência complementar aberta. Para introduzir estas informações na tábua de mortalidade faremos uso de alguns conceitos largamente empregados para a construção das chamadas tábuas seletas de mortalidade.

São diversos os métodos que podem ser empregados para a construção de uma tábua seleta de mortalidade. Neste trabalho utilizaremos o método proposto por Benjamin e Pollard (1980), que consiste em um modelo de graduação por referência a uma tábua básica.

Quando os dados são provenientes de uma experiência similar para a qual uma tábua resumo já existe, é possível utilizar esta tábua como uma curva para graduar os novos dados (Arteaga, 2009). A tábua que será utilizada como referência nesta dissertação é a tábua “Experiência do Mercado Segurador Brasileiro (BR– EMS), que foi desenvolvida pelo Laboratório de Matemática Aplicada – LABMA, da UFRJ (Universidade Federal do Rio de Janeiro) em parceria com a Superintendência de Seguros Privados – SUSEP e a Federação Nacional de Previdência Privada e Vida – Fenaprevi. Esta tábua foi construída no mesmo molde que estamos utilizando neste trabalho, ou seja, segregada entre cobertura e sexo, o que torna ainda mais conveniente a utilização da mesma como tábua de referência.

Os dados utilizados para a construção da referida tábua e os utilizados nesta dissertação são provenientes da mesma fonte, com a única diferença que neste trabalho utilizamos somente as informações dos participantes que abandonaram o plano, enquanto que para a construção da tábua BR – EMS, o LABMA/UFRJ utilizou somente as informações referentes aos participantes ativos. Isso indica ser totalmente plausível a suposição de que os dados para a construção da tábua seleta possuem experiência similar à da tábua de referência.

Assim sendo, podemos escolher entre diversas formas para relacionar a experiência de mortalidade apresentada pela tábua de referência com a mortalidade de cada período de seleção. Optamos por utilizar uma forma simples e intuitiva, que é calcular a razão entre as taxas específicas de mortalidade,  $q_{x,s,cob}^i$ , derivada dos dados para cada ano de seleção (i), idade (x), sexo (s), e cobertura (cob) e as taxas específicas de mortalidade,  $q_{x,s,cob}^{h-p}$ , provenientes da

tábua de referência. É importante ressaltar que as taxas específicas brutas de mortalidade derivadas dos dados são obtidas por:

$$q_{x,s,cob}^i = \frac{O_{x,s,cob}^i}{E_{x,s,cob}^i + \frac{O_{x,s,cob}^i}{2}},$$

para  $i = 1, 2, 3$  e  $4$ ,  $s =$  feminino ou masculino e  $cob =$  morte ou

sobrevivência. O termo  $\frac{O_{x,s,cob}^i}{2}$  acrescido ao denominador é um fator de correção ao estimador de máxima verossimilhança de  $q_x$ , utilizado para transformar as taxas centrais de mortalidade em probabilidades de morte, pois

$$L_x = \frac{l_x + l_{x+1}}{2} = \frac{l_x + l_x - O_x}{2} = l_x - \frac{O_x}{2}$$

$$l_x = L_x + \frac{O_x}{2}$$

De posse da estimativa de  $q_{x,s,cob}^i$  a obtenção da razão para a construção da tábua seleta é obtida diretamente por:

$$r_{x,s,cob}^i = \frac{q_{x,s,cob}^i}{q_{x,s,cob}^{h-p}},$$

onde  $q_{x,s,cob}^{h-p}$  é a taxa de mortalidade proveniente da tábua de referência desenvolvida pelo LABMA/UFRJ para a idade ( $x$ ), sexo ( $s$ ) e cobertura ( $cob$ ), sendo que o sobrescrito  $h-p$  faz referência ao modelo de Heligman and Pollard (1980) utilizado para a construção da tábua de referência.

Como esperamos que o comportamento dos nossos dados seja bastante semelhante ao apresentado pela tábua de referência, a tendência é a de que nenhum valor encontrado para as razões mencionadas acima se afaste muito da unidade, a não ser devido ao fato de existir uma seleção adversa nas saídas dos planos, que é a hipótese que esta sendo investigada neste trabalho.

Um aspecto importante a respeito da construção de tábuas de mortalidade que ainda não foi tratado neste capítulo refere-se à graduação ou suavização das taxas de mortalidade por idade, uma vez que as taxas brutas geralmente apresentam mudanças bruscas de uma idade para outra, o que não corresponde à hipótese plausível de que as probabilidades de morte para duas idades consecutivas devem ser muito próximas.

Para contornar esse problema, ajustamos um modelo estatístico para as razões utilizadas para a obtenção da tábua seleta, obtendo assim um valor estimado para a razão suavizada,  $\hat{r}_{x,s,cob}^i$ . A partir deste valor, as probabilidades de morte suavizadas, para cada período de seleção, sexo, cobertura e idade, podem ser estimadas pela seguinte expressão:

$$\hat{q}_{x,s,cob}^i = q_{x,s,cob}^{h-p} * \hat{r}_{x,s,cob}^i$$

A subseção seguinte aborda o tema da modelagem das razões.

## 4.2. Ajuste do modelo

A partir desta etapa do trabalho só consideramos as idades entre 20 e 100 anos, pois nestas idades estão concentrados os consumidores de produtos de seguro de vida e previdência complementar aberta, e os dados fora destes limites são muito escassos.

Abaixo apresentamos os gráficos de dispersão das razões as quais desejamos modelar, separadas por cobertura e sexo:

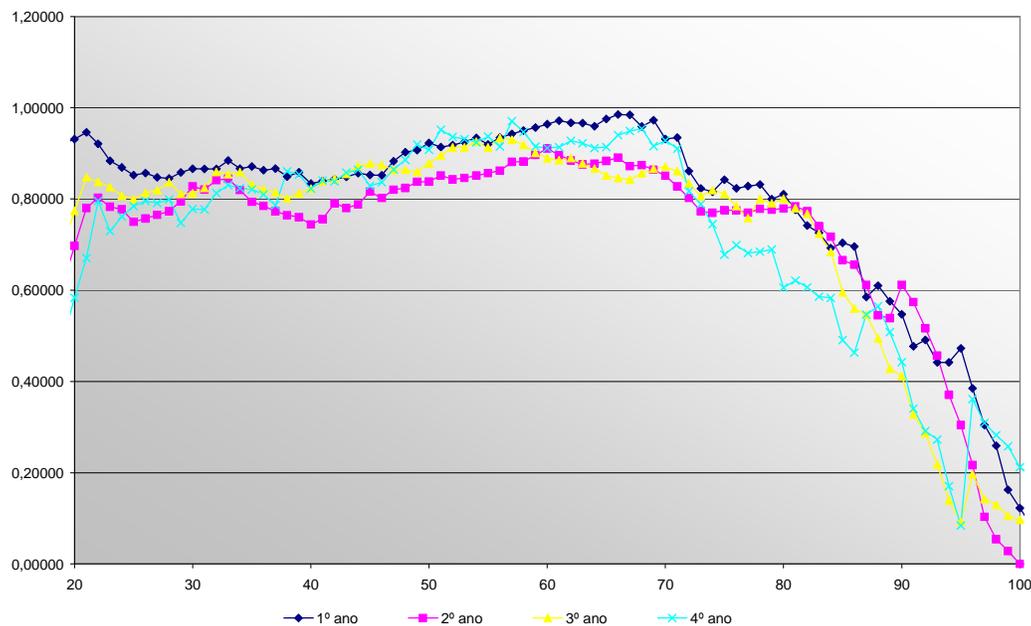


Gráfico 1 - Razão entre as taxas específicas brutas encontradas nos dados e as da tábua BR – EMS, por período de seleção, para a cobertura "Morte" e sexo "Masculino".

Fonte: Base de microdados de segurados do projeto FenaPrevi/Labma-UFRJ.

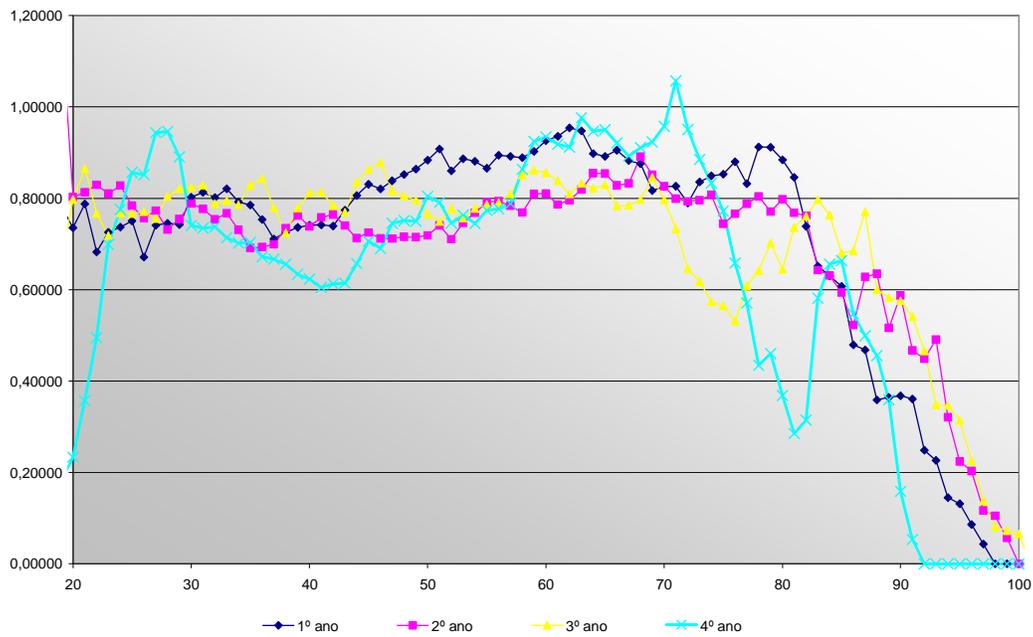


Gráfico 2- Razão entre as taxas específicas brutas encontradas nos dados e as da tábua BR – EMS, por período de seleção, para a cobertura "Morte" e sexo "Feminino".

Fonte: Base de microdados de segurados do projeto FenaPrevi/Labma-UFRJ.

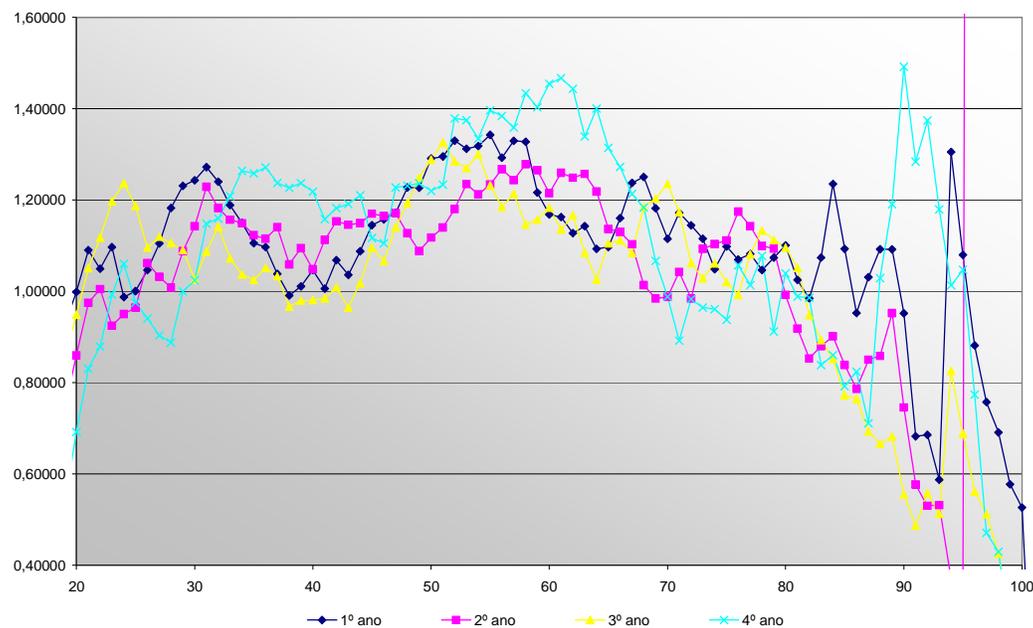


Gráfico 3 - Razão entre as taxas específicas brutas encontradas nos dados e as da tábua BR – EMS, por período de seleção, para a cobertura "Sobrevivência" e sexo "Masculino".

Fonte: Base de microdados de segurados do projeto FenaPrevi/Labma-UFRJ.

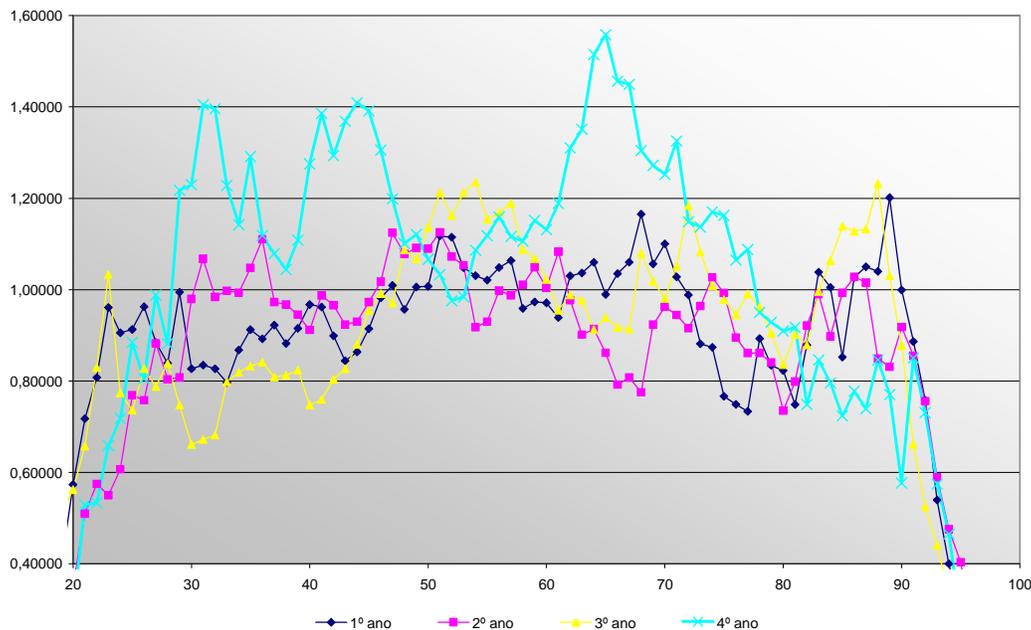


Gráfico 4 - Razão entre as taxas específicas brutas encontradas nos dados e as tábuas BR – EMS, por período de seleção, para a cobertura “Sobrevivência” e sexo “Feminino”.

Fonte: Base de microdados de segurados do projeto FenaPrevi/Labma-UFRJ.

O primeiro passo para a modelagem das razões é a determinação da forma funcional do modelo. Devido ao aspecto apresentado pelas mesmas e ao fato de os polinômios representarem uma família densa de modelos, decidimos testar o ajuste do seguinte polinômio<sup>2</sup>:

$$\hat{r}_{x,s,cob}^i = \alpha x_{s,cob} + \beta x_{s,cob}^2 + \gamma x_{s,cob}^3 + (a_1 I_1 + a_2 I_2 + a_3 I_3 + a_4 I_4) + (b_2 x_{s,cob} I_2 + b_3 x_{s,cob} I_3 + b_4 x_{s,cob} I_4) + (c_2 x_{s,cob}^2 I_2 + c_3 x_{s,cob}^2 I_3 + c_4 x_{s,cob}^2 I_4)$$

Equação 1

onde  $x_{s,cob}$  é a idade referente ao sexo (s) e à cobertura (cob),  $\alpha, \beta, \gamma, a_1, a_2, a_3, a_4, b_2, b_3, b_4, c_2, c_3, c_4$  são parâmetros a serem estimados e  $I_i$  é uma variável dummy para indicar o período de seleção.

<sup>2</sup> A primeira tentativa foi a de ajustar o modelo logito  $\left(\frac{C}{1+e^{(ax-b)}}\right)$  para as razões, porém o mesmo não se apresentou adequado para todos os subgrupos em estudo, o que nos fez optar pelo polinômio.

Com esse modelo estamos supondo que o termo dominante do polinômio ( $x_{s,cob}^3$ ) é obrigatoriamente o mesmo para todos os períodos de seleção e que estas razões podem diferir nos termos de ordem menor. Testamos a igualdade obrigatória para os dois ( $x_{s,cob}^2, x_{s,cob}^3$ ) e três ( $x_{s,cob}, x_{s,cob}^2, x_{s,cob}^3$ ) termos de ordem maior, mas o ajuste não foi adequado para todas as combinações de sexo e cobertura. O ajuste com os três termos de ordem maior iguais é equivalente a ajustar polinômios paralelos entre si, diferindo apenas no termo constante, que seria diferente para cada período de seleção e é representado pela variável dummy.

A distribuição das observações do número de órbitos  $O_{x,s,cob}^i$  é Bin( $E_{x,s,cob}^i, q_{x,s,cob}^i$ ), onde o estimador de  $q_{x,s,cob}^i$  é:

$$\hat{q}_{x,s,cob}^i = q_{x,s,cob}^{h-p} * \hat{r}_{x,s,cob}^i .$$

A razão  $\hat{r}_{x,s,cob}^i$  é parametrizada segundo a equação 1, e são calculados os estimadores de máxima verossimilhança dos parâmetros iterativamente. A princípio estima-se os parâmetros da equação 1 e em seguida, calcula-se a variância de  $O_x$ . O inverso da variância estimada é utilizado como peso num processo de mínimos quadrados ponderados para se estimar um novo conjunto de parâmetros. Calcula-se novamente a variância e as iterações terminam quando não há diferenças nos parâmetros estimados em dois passos consecutivos. Desta forma, fica evidente que, em modelos lineares generalizados, os estimadores de máxima verossimilhança são obtidos por um processo iterativo de mínimos quadrados ponderados.<sup>3</sup>

Procedendo à estimação dos parâmetros para cada combinação entre sexo e cobertura separadamente, em um total de quatro casos distintos, pudemos testar a significância de cada parâmetro para chegarmos, enfim, ao modelo mais significativo e parcimonioso para cada caso.

A significância dos parâmetros foi testada pelo método tradicional, com as seguintes hipóteses:

$$\begin{aligned} H_0 : p &= 0 \\ H_1 : p &\neq 0 \end{aligned}$$

<sup>3</sup> A estimação dos parâmetros foi realizada com o auxílio do software SPSS, por meio da procedure Analyze – Regression – Nonlinear.

Sendo  $H_0$  a hipótese nula,  $H_1$  a hipótese alternativa e  $(p)$  um símbolo para representar qualquer um dos parâmetros estimados. Caso a hipótese nula não seja rejeitada, podemos afirmar a um determinado nível de significância, que no caso do presente trabalho foi de 5%, que o parâmetro que está sendo testado não é significativo.

Após a estimação e teste de cada um dos parâmetros, chegamos a diferentes modelos para cada uma das combinações, que são “cobertura por morte e sexo masculino”, “cobertura por morte e sexo feminino”, “cobertura por sobrevivência e sexo masculino”, “cobertura por sobrevivência e sexo feminino”. Tais modelos serão apresentados detalhadamente no próximo capítulo, que trata sobre os resultados encontrados.