

2 Referencial Teórico

De acordo com Bowers et al (1997) a tábua biométrica é uma forma de sumarizar a vida dos indivíduos de uma população, onde uma probabilidade aferida será dependente de covariáveis como a idade, gênero, raça, profissão, renda ou tabagismo. A idade é o fator de risco mais importante sendo fundamental e inerente à construção de qualquer tábua biométrica.

São utilizadas normalmente quatro tipos de tábuas representativas das probabilidades relacionadas à mortalidade e invalidez. Para o evento gerador mortalidade de válidos, denomina-se tábua de sobrevivência quando construída ou aplicada a uma população cujo plano é pago em razão da sobrevivência do indivíduo. Já quando se dá em razão da morte é chamada de tábua de mortalidade. As tábuas de entrada em invalidez e de sobrevivência de inválidos mensuram a probabilidade de um válido se invalidar e de um inválido morrer, respectivamente.

Uma das primeiras questões que são abordadas é a escolha de qual tipo de segregação da população será feita. O que se observa normalmente é que para as tábuas de sobrevivência/mortalidade esta é dada por sexo, mas para as tábuas de entrada em invalidez e sobrevivência de inválidos não costuma haver este tipo de segregação por existir pouca quantidade de dados ou influência menor do fator sexo. Na hipótese de existir um volume de dados suficiente, é interessante avaliar o efeito da segregação entre fumantes e não fumantes, como na elaboração da CSO-2001, bem como a segregação por escolaridade, renda ou profissão, esta última principalmente com relação ao evento de entrada em invalidez.

Uma vez definida a segregação, cabe utilizar-se de diversas técnicas e identificar qual revela a melhor graduação para a tábua construída, através de testes estatísticos de validação. A graduação é uma metodologia de suavização das taxas brutas de mortalidade para que as probabilidades estimadas sejam suavizadas em relação às idades, e muitas vezes também se tornem monótonas e crescentes pois tais probabilidades crescem a partir de certa idade, geralmente na fase adulta.

2.1. Definições do Evento Gerador

Para o estudo da sobrevivência, em termos estatísticos define-se inicialmente X como a variável aleatória que define a idade no momento de morte de um recém-nascido. Outro evento gerador como a entrada em invalidez ou a mortalidade de inválido poderia ser analogamente definido.

A função de distribuição de X é dada por $F(x) = P(X \leq x), \forall x \in \mathbb{R}^+$. Equivalentemente, a função de sobrevivência de X é definida por $S(x) = 1 - F(x) = P(X > x)$. Deste modo, tem-se que $F(0) = 0$ e $S(0) = 1$. Conseqüentemente, a função de sobrevivência $S(x)$ será igual a ${}_x p_0$.

A probabilidade de o indivíduo de idade exata x sobreviver até completar $x+t$ anos de idade é denotada por ${}_t p_x = \frac{S(x+t)}{S(x)}$, que também pode ser expressa por $P(X > x+t | X > x) = \frac{P(X > x+t)}{P(X > x)}$.

A força de mortalidade ou taxa instantânea de mortalidade em x é dada por $\mu_x = -\frac{S'(x)}{S(x)}$. Ela também pode ser definida como $\lim_{z \rightarrow x^+} \frac{F(z) - F(x)}{1 - F(x)}$.

Estas probabilidades e taxas podem ser relacionadas na forma a seguir:

$${}_t q_x = \int_0^t {}_t p_x \mu_{x+s} ds \quad (1)$$

$${}_t p_x = \exp\left(-\int_0^t \mu_{x+s} ds\right) \quad (2)$$

Uma tábua de sobrevivência normalmente possui tabulações por idades das funções básicas q_x, d_x e l_x , onde ω é a idade máxima alcançada e:

l_0 é a raiz da tábua que representa o número inicial de pessoas no grupo.

l_x é o número esperado de sobreviventes à idade x , $x = 0, 1, \dots, \omega$.

${}_nq_x = 1 - {}_np_x = \frac{l_x - l_{x+n}}{l_x}$ é a probabilidade de o indivíduo de idade x

falecer antes de completar $x+n$ anos de idade. Para $n=1$, a notação simplificada é ${}_1q_x = q_x$, e equivalentemente, ${}_1p_x = p_x$.

$d_x = l_x - l_{x+1}$ é o número esperado de mortes na idade x .

Ao se obter q_x para todas as idades $x=0, \dots, \omega$ e se definir l_0 pode-se então calcular os valores de l_x . Normalmente a idade máxima ω é definida de acordo com a exposição da população estudada e um método de construção a ser escolhido, sendo este problema abordado na seção 2.3. Atualmente as tábuas biométricas atingem a idade máxima de até 120 anos.

O estudo da sobrevivência passa por compreender a variável aleatória A_x definida como o número de mortes em certa idade x que será uma função de q_x ou μ_x . Ao se observar uma população e se obter o número de mortes em cada idade pode-se então fazer inferência sobre q_x ou μ_x para cada uma das idades. Após essa primeira fase procede-se então com a segunda que é a suavização dos parâmetros entre as idades. Este é o processo que completa a graduação da tábua biométrica.

Assim, a primeira fase descrita passa por compreender a variável aleatória A_x . Hipóteses iniciais do processo do evento gerador (mortalidade de válido, entrada em invalidez ou mortalidade de inválido) devem ser feitas para se definir um modelo. Os principais modelos utilizados são o Binomial e o de Poisson que serão descritos nas seções 2.1.1 e 2.1.2. Já a segunda fase do processo de graduação, a suavização, terá suas metodologias descritas na seção 2.2.

Adicionalmente, deve ser observado que uma importante peculiaridade ocorre para o evento gerador mortalidade de inválidos. As definições descritas até então consideram que as probabilidades dependem somente da idade x alcançada. No entanto, para aposentados por invalidez as probabilidades de morte são funções da idade de entrada em invalidez e também da duração da mesma, visto que esta mortalidade nos primeiros anos após a entrada em invalidez é maior do que a mortalidade de aposentados com mesma idade mas que se aposentaram há mais tempo. Os indivíduos com piores condições de saúde tendem a falecer nos primeiros anos de invalidez (Bowers et al, 1997).

Este efeito seletivo tende a desaparecer com o aumento da duração da invalidez. O período onde ele está presente é chamado de período de seleção (*selection period*). Para um período de seleção de D anos, taxas de mortalidade são calculadas em cada idade para categorias de duração discretizadas $0, 1, \dots, D-1$, constituindo assim uma tábua bidimensional, também conhecida como tábua de dupla entrada. As categorias de duração $0, 1, \dots, D-1$ referem-se aos benefícios com durações desde a entrada em invalidez pertencentes aos intervalos $[0; 1), [1; 2), \dots, [D-1; D)$. Para durações maiores ou iguais a D , considera-se que o efeito desaparece e que as taxas de mortalidade são uma função somente da idade alcançada pelo segurado, sendo então denominadas de taxas últimas ou *ultimate rates*.

Nas tábuas seletas, $[x]$ é a idade de seleção (entrada em benefício) e ${}_t q_{[x]+z}$ a probabilidade do indivíduo morrer entre as idades $x+z$ e $x+z+t$, dado que ele se invalidou à idade x e encontra-se inválido há z anos. As probabilidades ${}_t q_{[x]+z}$ de cada uma das funções da tábua de mortalidade dependem então de duas variáveis: a idade de entrada em invalidez e a duração da invalidez. Estas probabilidades podem ser organizadas num arranjo bi-dimensional em $[x]$ e z , configurando uma tábua biométrica seleta.

Como o impacto da duração da invalidez sobre as probabilidades de morte tende a diminuir, assume-se que além do período de seleção D as probabilidades de morte dependem significativamente somente da idade alcançada pelo grupo de sobreviventes, isto é, ocorre que $q_{[x-j]+D+j} \cong q_{[x]+D}$, $j > 0$, recaindo no caso unidimensional onde a tábua biométrica é função somente da idade x .

Sendo assim, o arranjo bi-dimensional pode ser truncado ao fim do período de seleção, resultando num conjunto de dois tipos de tábuas de mortalidade, chamados de tábuas seletas e tábuas últimas (*select and ultimate tables*).

2.1.1. Modelo Binomial

A hipótese que possui uma forte característica intuitiva é a do modelo Binomial e tem como parâmetro a probabilidade de morte na idade. Ela parte do

princípio de que cada uma das idades possui n vidas independentes e identicamente distribuídas todas no início preciso da idade x . Seja a variável aleatória A_x representativa do número de óbitos e d'_x seu valor observado, pode-se então definir este modelo da seguinte forma:

$$A_x \sim \text{Binomial}(n, q_x) \rightarrow P(A_x = d'_x) = \binom{n}{d'_x} \cdot q_x^{d'_x} \cdot (1 - q_x)^{n-d'_x} \quad d'_x = 0, \dots, n \quad (3)$$

$$E(A_x) = n \cdot q_x \quad e \quad \text{Var}(A_x) = n \cdot q_x \cdot (1 - q_x) \quad (4)$$

A função de verossimilhança $L(\cdot)$ pode ser maximizada através da aplicação da função monótona logarítmica, facilitando os cálculos e chegando-se à expressão $\text{Log}(L(q_x)) = \log\left(\binom{n}{d'_x}\right) + d'_x \cdot \log(q_x) + (n - d'_x) \cdot \log(1 - q_x)$.

Derivando o logaritmo da verossimilhança em relação a q_x chega-se ao resultado $\frac{\partial}{\partial q_x} \log(L(q_x)) = \frac{d'_x}{q_x} - \frac{n - d'_x}{1 - q_x}$.

O estimador de máxima verossimilhança de q_x é dado quando se encontra o zero da função primeira derivada, maximizando a função de verossimilhança, sendo definido então por $\hat{q}_x = \frac{d'_x}{n}$.

Assim, o estimador \hat{q}_x possui média igual a q_x e variância $\frac{q_x \cdot (1 - q_x)}{n}$, além de ter distribuição assintótica Normal como qualquer estimador de máxima verossimilhança.

Infelizmente existem problemas que fazem com que este modelo não seja realístico e necessite de adaptações devido à censura nos dados. Primeiramente nem todos os indivíduos observados irão estar no mesmo intervalo de idade, o que limita as observações a indivíduos com idade no início do estudo entre x e $x+1$. Considera-se também que nem todos os indivíduos serão observados por um ano inteiro, ou seja, a vida estará no experimento somente no intervalo $[x + t_i, x + u_i]$, com $0 \leq t_i < u_i \leq 1$. Além disso, nem todos os indivíduos saem somente por morte, havendo o evento de entrada em invalidez.

Tais questões alteram o modelo Binomial de forma que a função de verossimilhança não mais poderá ser simplificada, devendo ser derivada

diretamente da distribuição Bernoulli, sendo escrita com base nas probabilidades de morte em períodos fracionários da seguinte forma:

$$\propto \prod_{i=1}^N ({}_{u_i-t_i}q_{x+t_i})^{d_i} (1 - {}_{u_i-t_i}q_{x+t_i})^{1-d_i}, \quad d_i = 0,1 \quad (5)$$

E se encontrar o estimador de q_x tem-se que utilizar alguma relação entre ${}_{1-t}q_{x+t}$ e q_x , o que pode ser feito de várias formas de aproximação, como a uniforme e a exponencial (Bowers et al, 1997). A mais utilizada é a de Balducci, assumindo que ${}_{1-u}q_{x+u}$ é uma interpolação harmônica das probabilidades de morte em u , $0 \leq u \leq 1$. Logo, pode-se escrever ${}_{1-u}q_{x+u} = (1-u) \cdot q_x$ e utiliza-se a seguinte relação:

$${}_{u_i-t_i}q_{x+t_i} = {}_{1-t_i}q_{x+t_i} - u_i P_{x+t_i} \cdot {}_{1-u_i}q_{x+u_i} \quad (6)$$

Pode-se finalmente substituir na função de verossimilhança e encontrar o estimador de máxima verossimilhança, também chamado de estimador atuarial, que assume a forma já conhecida de $\hat{q}_x = \frac{d'_x}{E_x}$, onde E_x é a exposição inicial ao

risco e possui a seguinte expressão:

$$E_x = \sum_{i=1}^n (u_i - t_i) + \sum_{i:\text{saídas}} (1 - u_i) \quad (7)$$

Observa-se que esta expressão adiciona à exposição que considera somente o período exposto do indivíduo realmente $(x + u_i) - (x + t_i) = u_i - t_i$, o tempo restante de exposição na idade para os indivíduos que saíram da observação.

É interessante observar que a expressão da exposição inicial ao risco pode ser escrita também como $E_x = E_x^C + \sum_{i:\text{saídas}} (1 - u_i)$, onde E_x^C é chamada de exposição central ao risco, sendo a soma das exposições “puras” individuais. A exposição inicial pode ser aproximada por $E_x = E_x^C + \frac{d'_x}{2}$, assumindo que as mortes ocorrem segundo uma distribuição uniforme para cada idade x .

Com estas definições, o modelo Binomial é geralmente escrito diretamente na forma da distribuição de probabilidade do número de óbitos:

$$A_x \sim \text{Binomial}(E_x, q_x)$$

No entanto o modelo Binomial não é de amplo uso atualmente. Ele pressupõe que as vidas são como elementos sob reamostragem com reposição, o que não é condizente com a realidade, além de considerar somente um estado de transição. Sendo assim, parte-se para o modelo de Poisson que estuda a força de mortalidade diretamente.

2.1.2. Modelo Poisson

No modelo Poisson, como o nome já diz, assume-se que o número de mortes numa determinada idade segue uma distribuição de Poisson. A distribuição é utilizada para modelar eventos raros a ocorrer num pequeno intervalo de tempo e só assume valores inteiros e não-negativos. Não há chance de ocorrer dois ou mais sinistros no mesmo instante, e a ocorrência de um independe daqueles já observados. Essa distribuição se adequa melhor à definição dos eventos de mortalidade ou de entrada em invalidez.

Ao se assumir μ_x constante ao longo da idade x , $\mu_{x+s} = \mu_x$, $1 < s \leq 0$, a distribuição de Poisson terá a seguinte forma, onde d'_x é o valor observado de A_x :

$$A_x \sim \text{Poisson}(E_x^C \cdot \mu_x) \rightarrow P(A_x = d'_x) = \frac{(E_x^C \mu_x)^{d'_x} e^{-E_x^C \mu_x}}{d'_x!}, \quad d'_x = 0, 1, \dots \quad (8)$$

$$E(A_x) = E_x^C \cdot \mu_x \quad e \quad \text{Var}(A_x) = E_x^C \cdot \mu_x \quad (9)$$

A função de verossimilhança pode ser maximizada através da aplicação da função monótona logarítmica.

$$\text{Log}(L(\mu_x)) = d'_x \cdot (\log E_x^C + \log \mu_x) - (E_x^C \cdot \mu_x) - \log d'_x! \quad (10)$$

Derivando a função em relação a μ_x chega-se a:

$$\frac{\partial}{\partial \mu_x} \log(L(\mu_x)) = \frac{d'_x}{\mu_x} - E_x^C \quad (11)$$

O estimador de máxima verossimilhança de μ_x é dado quando se encontra o zero da função primeira derivada, maximizando a função de verossimilhança, sendo definido então por $\mu_x = \frac{d'_x}{E_x^C}$.

Assim, o estimador $\hat{\mu}_x$ possui média igual a μ_x e variância $\frac{\mu_x}{E_x^C}$, além de ter distribuição assintótica Normal como qualquer estimador de máxima verossimilhança.

O modelo de Poisson, assim como o Binomial, pressupõe somente uma forma de transição de estado. Para ambos também ocorre que a exposição é conhecida e não considerada como uma variável aleatória. Para casos de mais de uma transição de estado, chega-se à extensão natural que é o modelo de Cadeias de Markov.

Assim, pode-se analisar mais de uma forma de decremento da população, por exemplo, no estudo do comportamento de invalidez temporária. Isto porque no caso das tábuas de entrada em invalidez e de mortalidade de inválidos, uma atenção especial deve ser dada ao fato que o indivíduo pode retornar à atividade. Ou seja, o modelo é útil para quando se deseja estudar além das transições *ATIVO* \longrightarrow *INVÁLIDO*, *ATIVO* \longrightarrow *MORTO* e *INVÁLIDO* \longrightarrow *MORTO*, também a movimentação reversa *INVÁLIDO* \longrightarrow *ATIVO*. Outros estados não absorventes também podem ser adicionados ao modelo, como o de desligado do plano.

Neste caso, a teoria de múltiplos decrementos descrita por Haberman & Pitacco (2001) deve ser utilizada. Havendo retorno à atividade, seu comportamento deverá ser estudado de forma a necessariamente participar do estudo de transição de estados. Os modelos não-paramétricos de Kaplan-Meier (1958) e de Nelson-Aalen (Nelson, 1972 e Aalen, 1978) podem ser aplicados quando E_x^C é considerada uma variável aleatória.

Quando existem somente dois estados (vivo e morto, por exemplo) o modelo de Markov é denominado modelo de Markov de dois estados, sendo igual ao modelo de Poisson: *ATIVO* $\xrightarrow{\mu_x}$ *MORTO*. O primeiro estado é *ATIVO* e o segundo estado é *MORTO* onde a transição se dá em uma única direção, pois o estado *MORTO* é absorvente. Por definição, o modelo assume que a probabilidade de uma vida se situar na idade seguinte, num dos estados, depende somente das idades envolvidas, do gênero do indivíduo e do estado atualmente ocupado (Hipótese de Markov). Ou seja, o modelo é “sem memória”. Também

assume que a probabilidade de morte num intervalo infinitesimal de tempo tem uma relação linear com a força de mortalidade: ${}_du q_{x+u} = du \cdot \mu_{x+u}$.

No entanto todos estes modelos (Binomial, Poisson e Cadeias de Markov) assumem independência entre as idades. Ou seja, os estimadores são calculados individualmente por idade e nenhum tipo de suavização é feito entre elas. Então os modelos de graduação são inseridos para garantir a suavização e com isso que nenhum comportamento brusco entre as idades seja observado.

2.2. Formas de Graduação Tradicionais

Após obter as probabilidades ou taxas brutas de mortalidade em cada idade, inicia-se o processo de graduação que é a técnica de suavização de curvas aplicado à construção da tábua biométrica. A razão mais importante do processo de suavização é que os estimadores não devem variar bruscamente com as idades. Desta forma os valores de contribuições/prêmios não se alteram bruscamente ao se aumentar ou diminuir a idade em um ano. No caso de um plano em repartição simples cujo benefício é pago pelo evento gerador morte, não seria justificável cobrar menos de um segurado mais idoso, exposto a um risco teórico maior.

O processo de graduação assume que o estimador para certa idade x possui informações sobre as idades adjacentes. Por exemplo, a idade x nos traz algumas informações sobre idades anteriores ($x-1, x-2, \dots$) e posteriores ($x+1, x+2, \dots$). O procedimento envolve um *trade-off* (balanceamento) entre suavização e ajuste, sendo que o objetivo é produzir taxas suavizadas, mas não excessivamente ao ponto de se perderem as características intrínsecas da população. Ao mesmo tempo, o grau de aderência (ajuste) aos dados não deve ser excessivo, para que não seja construída uma tábua representando a aleatoriedade amostral inerente a qualquer processo de inferência estatística.

As formas de graduação são divididas em dois tipos básicos: paramétrico (global) e não-paramétrico (local). Modelos paramétricos são indicados principalmente para quando há poucos dados disponíveis. Estes normalmente devem ser bem analisados para evitar sobreparametrização, refletindo excessivamente as especificidades amostrais, sendo a suavização preterida em relação ao ajuste. Já os não-paramétricos são mais flexíveis, justamente por não

estarem vinculados às funções pré-definidas. Neles, os dados “falam por si” e por isto funcionam para quando há muitos dados. Entretanto, estes modelos têm como problema a subjetividade na escolha do grau de suavização.

2.2.1. Graduação Paramétrica

Existem algumas funções matemáticas já reconhecidas como representativas do comportamento da mortalidade humana ao longo das idades. Estas representam os modelos paramétricos, conhecidos também como modelos globais, justamente por considerarem a experiência de todas as idades observadas e incorporarem essa informação em uma única função.

A vantagem de se trabalhar com uma forma analítica está no fato de que a partir da função matemática que descreve a tábua biométrica se pode facilmente obter fórmulas explícitas que representam as probabilidades de ocorrência do evento gerador para as vidas envolvidas.

Porém, pode ocorrer para as idades iniciais e finais da tábua uma graduação bem diferente dos dados brutos observados. Para solucionar essa distorção, uma mistura de modelos paramétricos pode ser utilizada para cada grupo de idade adjacente, por exemplo. O problema nesse caso é como determinar a transição de uma função para outra de forma a evitar descontinuidade. Uma forma usual de contornar o problema é forçar numericamente que as diferentes funções tenham o mesmo valor nas idades de intersecção.

A primeira tentativa de se modelar as probabilidades relacionadas à sobrevivência foi sugerida por De Moivre em 1724, representando uma aproximação excessivamente simples. Em seu modelo, a força de mortalidade era

modelada como $\mu_x = \frac{1}{\omega - x}$, o que equivale a ${}_x p_0 = 1 - \frac{x}{\omega}$. Essa fórmula foi

utilizada por algumas décadas devido a sua fácil implementação. Diversos modelos também foram propostos, como o de Babbage (1823), Young (1826), Littrow (1832), Moser (1839), Opperman (1870), Thiele (1871), Perks (1932), Weibull (1951), Beard (1971), Barnett (CMI Bureau, 1974, p. 141) e Wilkie (1976), porém se tornaram menos populares.

Os três modelos atualmente mais utilizados em graduação paramétrica, Gompertz-Makeham, Heligman-Pollard e o Modelo Linear Generalizado são explicados nas subseções posteriores.

2.2.1.1.

Gompertz-Makeham

O primeiro modelo a se tornar amplamente reconhecido para fins atuariais foi o modelo de Gompertz (1825) que possui a limitação de só considerar o fator de morte natural Bc^x em seu modelo matemático. Posteriormente, Makeham (1867) incluiu na função uma parcela, A , correspondente a mortes acidentais, a qual independe da idade da pessoa. Os modelos de Gompertz e de Makeham consideram que a força de mortalidade cresce exponencialmente com a idade e podem ser descritos por:

$$\text{Gompertz: } \mu_x = Bc^x \quad (12)$$

$$\text{Makeham } \mu_x = A + Bc^x \quad (13)$$

Makeham (1890) também propôs outro acréscimo, contendo uma terceira componente que aumenta em progressão aritmética com o passar das idades, implicando $\mu_x = A + Bc^x + Hx$.

A Lei de Makeham foi usada na construção das conhecidas tábuas CSO-41 e AT-49. A importância desses modelos foi tamanha que deu origem aos atualmente conhecidos modelos de Gompertz-Makeham (GM), conforme proposto por Forfar, McCutcheon e Wilkie (1988). O modelo GM de ordem (r,s) possui a forma geral dada por:

$$\mu_x = GM(r, s) = \sum_{i=0}^{r-1} \alpha_i x^i + \exp\left(\sum_{i=0}^{s-1} \beta_i x^i\right) e \quad q_x = \frac{GM(r, s)}{1 + GM(r, s)} \quad (14)$$

Nota-se que a fórmula GM é um polinômio de grau $r-1$ somado à exponencial de um polinômio de grau $s-1$. Então uma linha reta é ajustada se for definido $r=2$ ou então $s=2$ numa escala logarítmica natural. O modelo original de Gompertz pode ser escrito como $e^{(\log B + x \log C)} = e^{(\beta_0 + \beta_1 x)}$. Já o modelo inicial de Makeham é um GM(1,2) onde $A = \alpha_0$, $B = e^{\beta_0}$ e $C = e^{\beta_1}$, enquanto que o modificado é um GM(2,2) onde $H = \alpha_1$.

O procedimento para escolher a ordem do GM é partir de um modelo mais simples para modelos de ordem mais altas. Cabe então decidir se o aumento da ordem do modelo gera melhora significativa no ajuste da função. Para avaliar a intensidade dessa melhora existem testes estatísticos que ajudam a investigar a inclusão de novos parâmetros aos modelos. O teste mais utilizado é a estatística de razão de verossimilhança.

Suponha que um modelo esteja ajustado com p parâmetros e L_p seja a log-verossimilhança. O que se pretende é avaliar o efeito no ajuste devido à inclusão de mais q parâmetros, decidindo-se então por um modelo com $p+q$ parâmetros, cuja log-verossimilhança será L_{p+q} . Sob a hipótese nula de que os q parâmetros extras não têm efeito na presença dos p parâmetros originais, a estatística $-2(L_p - L_{p+q})$ tem uma distribuição assintótica χ^2 com q graus de liberdade.

Os parâmetros podem ser estimados por máxima verossimilhança, por mínimos quadrados, ou ainda minimizando a estatística χ^2 de valor

$$\sum_{i=0}^n \frac{\left(d'_i - \overset{\circ}{d}_i\right)^2}{\text{Var}\left(\overset{\circ}{d}_i\right)}, \text{ onde } \overset{\circ}{d}_i \text{ é o número de óbitos de acordo com o modelo graduado e}$$

d'_i é o número de óbitos observados.

Observe que se pode aplicar GM tanto em q_x quanto em μ_x . Caso tenha sido eleito o modelo de Poisson, em função da força de mortalidade, a distribuição de probabilidade e a função de verossimilhança serão dadas por:

$$A_x \sim \text{Poisson}(E_x^C \cdot \mu_x) \rightarrow P(A_x = d'_x) \propto e^{-E_x^C \mu_x} (\mu_x)^{d'_x} \quad (15)$$

Aplicando GM em μ_x , a log-verossimilhança será proporcional a $d'_x \log(GM) - E_x^C GM$, que no caso de Gompertz se simplifica a $d'_x(\alpha_1 + \alpha_2 x) - E_x^C e^{\alpha_1 + \alpha_2 x}$, $x = 0, 1, \dots, n$.

Já para o modelo Binomial, em função da probabilidade de morte anual, as estatísticas são:

$$A_x \sim \text{Binomial}(E_x, q_x) \rightarrow P(A_x = d'_x) \propto q_x^{d'_x} \cdot (1 - q_x)^{E_x - d'_x} \quad (16)$$

Aplicando GM em $\frac{q_x}{1-q_x}$, a log-verossimilhança será proporcional a

$$d'_x \log\left(\frac{q_x}{1-q_x}\right) - E_x \log\left(\frac{1}{1-q_x}\right) = d'_x \log(GM) - E_x \log(1-GM) \quad \text{que no caso de}$$

Gompertz se simplifica a $d'_x(\alpha_1 + \alpha_2 x) - E_x \log(1 - e^{\alpha_1 + \alpha_2 x})$, $x = 0, 1, \dots, n$.

Deste modo, partindo das log-verossimilhanças chega-se à estatística com distribuição χ^2 que fornecerá evidência estatística sobre a inclusão ou não de mais variáveis no modelo.

2.2.1.2.

Heligman-Pollard

O modelo de Heligman e Pollard (1980) é uma equação composta pela soma de três parcelas correspondendo, respectivamente, à mortalidade infantil, mortalidade por causas externas e mortalidade por senescência, semelhante à divisão proposta por Thiele (1871). Este é o modelo utilizado no Brasil para o cálculo das tábuas de mortalidade da população (IBGE, 2007).

O modelo Heligman-Pollard (HP) do tipo 1 assume a forma:

$$\mu_x = A^{(x+B)^C} + D \cdot \exp\left(-E \cdot (\log x - \log F)^2\right) + \frac{GH^x}{1+IGH^x} \quad (17)$$

Já para o modelo tipo 2, a parcela de mortalidade por senescência sofre alteração, como exposta a seguir:

$$\mu_x = A^{(x+B)^C} + D \cdot \exp\left(-E \cdot (\log x - \log F)^2\right) + \frac{GH^{x^I}}{1+GH^{x^I}} \quad (18)$$

O parâmetro D balanceia a mortalidade por causas externas, definida pelos valores E e F . Quanto maior o D , maior será o valor da parcela de mortalidade correspondente aos adultos. Por estar multiplicado pela função exponencial atinge principalmente as idades iniciais, pouco influenciando as maiores que 80 anos.

O parâmetro G é também um fator multiplicador, da terceira parcela, no caso. Seu efeito é provocar uma soma em $\log(G)$ na abscissa (idade), transladando a curva de mortalidade para cima e para esquerda, conforme seu aumento.

Já o parâmetro H é responsável pela definição do ponto de inflexão da curva, passando de concavidade acima para uma concavidade abaixo. O ponto de inflexão ocorre no zero da segunda derivada da função, onde $x = -\frac{\log(IG)}{\log(H)}$.

Por fim, o parâmetro I define a assíntota horizontal para as idades superiores, que será igual à $D + \frac{1}{I}$. Portanto, quanto menor for I , positivo, maior será o valor da assíntota. Se I assumir um valor negativo não haverá assíntota horizontal, permanecendo a curva sempre côncava.

2.2.1.3. Modelo Linear Generalizado

O modelo de Gompertz-Makeham pode ser visto como um caso particular do Modelo Linear Generalizado (GLM). O GLM é descrito como uma regressão na soma linear de funções pré-definidas, segundo à eq. (19).

$$Y_j = \psi_{x_j} \beta_j \quad (19)$$

Assume-se que o vetor Y composto de n observações é a realização de uma variável aleatória cujos componentes são independentes e identicamente distribuídos (i.i.d.), com média igual a μ . O GLM é definido por três componentes: uma variável aleatória, uma determinística e uma função de ligação.

A variável aleatória é representada pelos elementos de Y que pertencem à família exponencial, sendo i.i.d. com $E(Y) = \mu$.

A componente determinística é formada pelas variáveis explicativas x_1, x_2, \dots, x_n que definem um preditor linear da forma $\eta = \sum_{i=1}^p x_i \beta_i$, onde β_i representa os parâmetros desconhecidos a serem estimados pelo modelo.

A função de ligação relaciona a variável aleatória à componente sistemática, sendo uma função monótona, diferenciável em seu domínio e que possui inversa definida.

Conseqüentemente, é possível definir mais de uma função de ligação para cada família de distribuição de probabilidade. As escolhas elencadas a seguir

levam a estimativas em uma função com estatística suficiente e minimal para os parâmetros do preditor linear.

- Distribuição Normal: função Identidade
- Distribuição Binomial: função Logito
- Distribuição Poisson: função Logarítmica
- Distribuição Gama: função Inversa

Estas funções de ligação são conhecidas como funções canônicas. Além de possuírem as propriedades estatísticas desejáveis, atuam de forma que a componente sistemática torna-se aditiva na escala da própria função, facilitando o objetivo de aplicação prática.

É comum se definir que os elementos Y são normalmente distribuídos com variância constante e função de ligação Identidade, como na eq. (20).

$$Y_i \sim Normal(\mu, \sigma^2) \quad i.i.d. \rightarrow \mu = \sum_{i=1}^p x_i \beta_i \quad (20)$$

Mas para a graduação de tábuas biométricas, considera-se geralmente que o número de mortes segue uma distribuição Poisson quando a exposição central ao risco é conhecida. Utiliza-se, portanto, a função de ligação logarítmica. Outra forma utilizada é a da distribuição Binomial quando a exposição inicial é conhecida, utilizando-se a função de ligação Logito.

Regra geral, nas distribuições pertencentes à família exponencial a variância da variável dependente pode ser expressa em função da média. A constante ϕ representa o parâmetro de escala para a variância e possui valor constante, mas não necessariamente conhecido.

$$Var(Y) = \phi \cdot f(\mu) \quad (21)$$

Na distribuição Gama, a função variância é proporcional ao quadrado da média. Na Binomial, a variância é igual à média multiplicada por um menos a probabilidade de ocorrência do evento. Já na distribuição de Poisson, a variância é igual à média.

Por vezes, pode haver sub ou sobredisposição nos dados, refletido pelo valor de ϕ menor ou maior que 1, respectivamente. Renshaw e Haberman (1996) mostram que a modelagem por apólice, e não por vida, gera ϕ maior que 1, pois uma vida pode ter mais de uma apólice. Diversos testes podem ser aplicados para identificar a sobredisposição nos dados, como os testes apresentados por Ganio e

Schafer (1992), Dean (1992) e Smith e Heitjan (1993), além do tradicional teste Qui-quadrado de Pearson.

Já para testar a adequação dos modelos, utiliza-se o teste Qui-quadrado quando a amostra não é pequena. Num modelo ajustado com p parâmetros, o *Residual Deviance* possui distribuição Qui-quadrado com p graus de liberdade. Analogamente, um modelo de $p+q$ parâmetros possui distribuição com $p+q$ graus de liberdade. Sendo assim, a diferença entre o *Residual Deviance* destes dois modelos possui também distribuição Qui-quadrado, mas com q graus de liberdade. A estatística pode ser utilizada para testar a hipótese nula de que não há melhora estatisticamente significativa partindo de um modelo com menos (p) para outro com mais ($p+q$) parâmetros.

2.2.2. Graduação Não-Paramétrica

Se os modelos paramétricos são os mais utilizados no Brasil (Conde, 1991, Beltrão et al, 1995, Ribeiro e Pires, 2001, Beltrão e Sugahara, 2002a e 2002b, Neves, 2004 e Ribeiro, 2006) e possuem uma grande aceitação pela sua estrutura formal, os modelos não-paramétricos, também conhecidos como modelos locais, foram objeto de estudo no Brasil por Neves (2004) e também utilizados pela Society of Actuaries (2001) na elaboração da última tábua de sobrevivência RP-2000 pelo fato de que geram resultados satisfatórios para a situação em que existe uma grande quantidade de dados.

2.2.2.1. Whittaker-Henderson

Whittaker-Henderson (1952) é o modelo não-paramétrico mais utilizado ultimamente, sendo adotado pela Society of Actuaries (2001) e American Academy of Actuaries (2002). Este método consiste em minimizar a seguinte expressão, onde o primeiro termo mede a qualidade do ajuste e o segundo a suavização:

$$\sum_{t=0}^{\overline{w}-x} w_{x+t} (q'_{x+t} - q_{x+t})^2 + \sum_{t=0}^{\overline{w}-x-z} K_{x+t} (\Delta^z q_{x+t})^2 \quad (22)$$

w_{x+t} : coeficiente de ponderação do ajuste.

q'_x : probabilidades brutas (observadas).

q_x : probabilidades graduadas.

Δ^z : diferença finita de ordem z .

$$\Delta^z q_x = \binom{z}{0} q_x - \binom{z}{1} q_{x-1} + \binom{z}{2} q_{x-2} - \binom{z}{3} q_{x-3} + \dots + (-1)^z \binom{z}{z} q_{x-z}$$

K_{x+t} : peso positivo da suavização.

Este método representa uma ponderação clara entre o ajuste do modelo e a sua suavização. O primeiro termo mede a proximidade entre as taxas graduadas e as brutas (ajuste) e o segundo a suavidade com que as taxas graduadas se alteram com as idades (suavização). Ele também possui um forte componente subjetivo, já que as regras de definição dos parâmetros z , K_{x+t} e w_{x+t} não são rígidas e nem possuem critérios de comparação pré-estabelecidos.

Os estimadores graduados são obtidos minimizando a expressão definida anteriormente, em função dos q_x , obtendo-se um sistema de equações lineares.

Existem dois tipos de métodos de Whittaker-Henderson (WH): tipo A e tipo B, assumindo hipóteses quanto ao coeficiente de ponderação w_{x+t} . No tipo A assume-se que este é inexistente ($w_{x+t} = 1$ para $0 \leq x+t \leq \omega, t > 0$). No tipo B, mais indicado para quando há um grande volume de dados, os coeficientes serão definidos de acordo com a exposição. Uma possível formulação é

$$w_{x+t} = \frac{E_{x+t}}{q_{x+t}} (1 - q_{x+t}).$$

Muitas vezes os parâmetros do modelo são definidos de forma a garantirem consistência teórica na tábua. Por exemplo, a SOA afirma que como critério para seleção dos parâmetros, a graduação final deverá apresentar poucos casos em que $q_x > q_{x+1}$ e nenhum em que $q_x < 0$ ou $q_x > 1$, para $x = 0, 1, \dots, \omega$.

Tais definições são muito vagas e por isso normalmente envolvem um grupo de especialistas para fazê-las. No entanto, se do ponto de vista acadêmico tal método parece ser menos robusto por seu caráter discricionário, ele certamente apresenta resultados coerentes se utilizado corretamente.

2.2.2.2. Graduação de Kernel

Já a Graduação de Kernel é outra forma de graduação não-paramétrica que representa um processo de utilização de médias móveis ponderadas. Foi inicialmente aplicado ao processo de graduação por Copas e Haberman (1983) e Ramlau-Hansen (1983). A técnica estima a probabilidade q_x como:

$$q_x = \frac{\sum_{i=1}^n \frac{d'_{x_i}}{E_{x_i}} K_b(x - x_i)}{\sum_{i=1}^n K_b(x - x_i)} \quad (23)$$

Uma forma alternativa para a graduação é dada pela estimação da taxa por:

$$q_x = \frac{\sum_{i=1}^n \frac{d'^2_{x_i}}{E_{x_i}} K_b(x - x_i)}{\sum_{i=1}^n K_b(x - x_i)} \quad (24)$$

$K_b = K\left(\frac{x}{b}\right)/b$ e $K(\cdot)$ é uma função Kernel dada por $\int_{-\infty}^{+\infty} K(x) dx = 1$,

sendo comumente escolhida a função de densidade da distribuição Normal Padrão.

A constante b mede o grau de suavização do modelo, podendo ser arbitrado ou calculado por validação cruzada, quando terá seu valor dado na

situação em que se minimiza a função $\frac{1}{n} \sum_{i=1}^n \left(\frac{d'_x}{E_{x_i}} - (q_{x_i})^{-i} \right)^2$, onde $(q_{x_i})^{-i}$ é o

valor estimado para o caso em que não se utiliza a informação da idade x_i .

O problema deste modelo consiste na sua aplicação aos extremos da tábua biométrica, onde o valor de b pode não revelar uma constante que suavize satisfatoriamente os dados.

2.2.2.3 Graduação por Polinômios Locais

A técnica proposta por Cleveland (1979) consiste em se traçar vários polinômios em cada segmento de idades, utilizando-se regressão. A vantagem

deste método em relação à Graduação de Kernel é de não haver o problema do mau ajuste com relação às idades extremas da tábua biométrica (Verrall, 1996).

Define-se a vizinhança $N(x_0)$ como o conjunto dos k vizinhos a $x_0 = \frac{d'_{x_0}}{E_{x_0}}$.

Por exemplo, se $k=4$, a vizinhança será $N(x_0) = \{x_{-2}, x_{-1}, x_0, x_1, x_2\}$. Após, calcula-

se $\Delta(x_0) = \max_{x_i \in N(x_0)} |x_0 - x_i|$ e derivam-se os pesos $w_i = T\left(\frac{|x_0 - x_i|}{\Delta(x_0)}\right)$, onde

$$T(u) = (1 - u^3)^3.$$

Por fim, traça-se a regressão polinomial (linear, quadrática, cúbica, etc.) na região $N(x_0)$ usando os pesos $\{w_1, w_2, \dots, w_{k+1}\}$. O valor estimado \hat{q}_{x_0} é $f(x_0)$, onde $f(\cdot)$ é a função da regressão polinomial ajustada.

2.3. Graduação do Fim da Tábua

O fim da tábua biométrica é representado pela última idade ω , sendo suficientemente alta pois interfere na precificação dos produtos de aposentadoria e pensão. Existe praticamente um consenso sobre a idade ω de 120 anos ser considerada satisfatória para a elaboração das tábuas biométricas atuais, uma vez que há no mundo e em toda a história, comprovados documentalmente, menos de uma dezena de indivíduos que vieram a falecer em idades superiores a 120.

Hustead (2005) apresenta as técnicas utilizadas para a construção do final da tábua e demonstra baseado nas taxas brutas da RP-2000 que os efeitos financeiros da utilização de diferentes métodos provocam impacto financeiro relevante apenas na hipótese de planos contratados por indivíduos centenários, o que, no entanto, sempre é vedado pela política de subscrição das entidades.

Existem basicamente quatro formas de construção do fim da tábua biométrica. O *Forced Method* simplesmente define a taxa imediatamente posterior à última taxa bruta suavizada como valor de probabilidade igual a um. Essa teoria assume que a duração da vida tem um limite (que aumenta com o passar dos anos), justificando esse salto para a probabilidade unitária. O problema é que essa descontinuidade da penúltima para a última idade gera variações abruptas para o cálculo de contribuições/prêmios e provisões/reservas de indivíduos com idades

avançadas. As tábuas construídas pelo Instituto Brasileiro de Geografia e Estatística (IBGE, 2007) para a população brasileira se baseiam neste método forçando a última idade $\omega=80$ a ter probabilidade unitária, o que corresponde a uma idade excessivamente baixa devido ao fato de a população do país oferecer uma grande quantidade de óbitos para idades superiores a este patamar.

No *Blended Method* utilizam-se somente as últimas idades (95 ou mais, por exemplo) para gerar um padrão de graduação que é estendido até a probabilidade atingir o valor um. A CSO-2001 utilizou este método capturando informações desde a idade 95 para a construção da tábua até $\omega=120$. Este método foi eleito para a graduação não-paramétrica do fim das tábuas deste estudo, pois a partir de 95 anos havia algumas idades cujas taxas brutas eram zero.

O *Pattern Method* é semelhante ao *Blended Method*, mas considerando todas as idades para o estabelecimento do padrão matemático e não só as últimas. Segundo essa teoria, a curva deve mudar de concavidade para baixo nas idades altas até tangenciar a probabilidade unitária, assumindo que em tese a morte nunca é certa ($q_x < 1, \forall x \in \square$) e a vida não tem limite, como implícito no *Forced Method*. Este método foi utilizado para a graduação paramétrica aplicada neste estudo.

O *Lass-Than-One Method* considera que a probabilidade para idades avançadas tange a um valor menor que um. Este método é normalmente utilizado em tábuas censitárias elaboradas pelos órgãos governamentais, onde não há preocupação quanto à sua utilização em um plano pago pelo evento gerador sobrevivência ou morte. A tábua RP-2000 utiliza esta metodologia, onde a probabilidade atinge 0,4 na idade 115 para o sexo feminino e 106 para o masculino.

2.4. Testes de Adequação

Para decidir se a graduação é satisfatória deve-se analisar se o resultado é aderente e também se suaviza os dados observados nas idades. No entanto, essas características são conflitantes: a suavização dos dados deve ser suficiente apenas para não haver “saltos” na função construída.

Portanto, testes estatísticos são normalmente definidos, sendo que para a modelagem paramétrica os principais são os que avaliam a qualidade de ajuste do modelo diante dos dados. A gradação paramétrica já produz estimadores suavizados, então os testes de suavização de dados são menos importantes.

2.4.1. Teste das Diferenças

O Teste das Diferenças é um teste de suavização, normalmente utilizado em modelagem não-paramétrica, possibilitando a identificação de um modelo que apresente oscilações excessivas nas taxas graduadas. Para medir a suavização, a terceira diferença entre os estimadores deve ser calculada e ser pequena em magnitude comparada com os próprios estimadores, além de progredir regularmente. A terceira diferença é a diferença entre as segundas diferenças, que por sua vez é a diferença entre as primeiras diferenças.

$$\nabla^3 \overset{\circ}{q}_x = \nabla^2 \overset{\circ}{q}_x - \nabla^2 \overset{\circ}{q}_{x-1} \text{ e } \nabla^2 \overset{\circ}{q}_x = \overset{\circ}{q}_x - \overset{\circ}{q}_{x-1} \quad (25)$$

Assim, tem-se que a terceira diferença é dada por:

$$\nabla^3 \overset{\circ}{q}_x = \overset{\circ}{q}_x - 3\overset{\circ}{q}_{x-1} + 3\overset{\circ}{q}_{x-2} - \overset{\circ}{q}_{x-3} \quad (26)$$

Já para mensurar o grau de ajuste (aderência) aos dados, diversos testes podem ser utilizados sobre as taxas graduadas comparando-as às taxas brutas, sendo apresentados nas seções 2.4.2 a 2.4.9.

2.4.2. Análise de Resíduos sob Normalidade

Para avaliar o ajuste de uma gradação os resíduos devem ser examinados. É usual analisar os desvios divididos pelo estimador de seu erro padrão:

$$Z_x = \frac{d'_x - \overset{\circ}{d}_x}{\sqrt{\text{Var}\left(\overset{\circ}{d}_x\right)}} \quad (27)$$

Examinam-se qualitativamente os gráficos dos resíduos, sendo muito útil a observação do gráfico qq-plot ou mesmo do histograma dos resíduos que podem deflagrar uma não-normalidade visualmente clara nos resultados do modelo ajustado, conforme exemplificado na Figura 1.

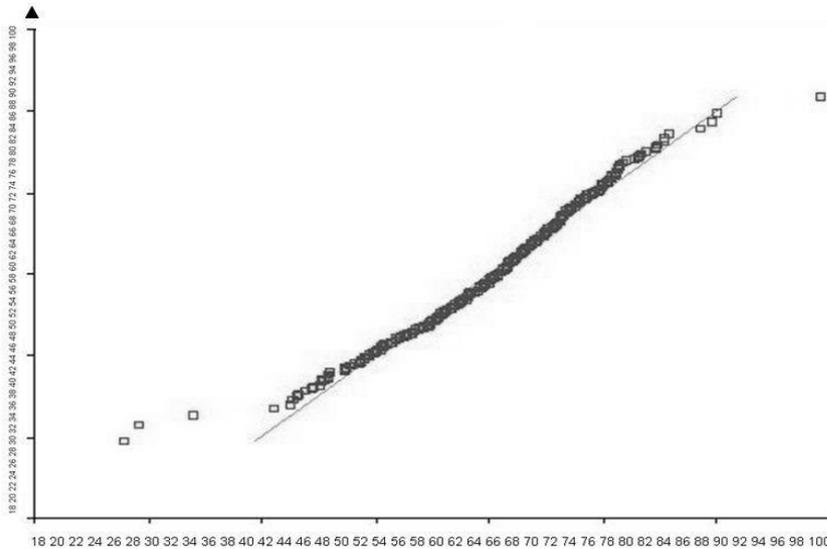


Figura 1 – Exemplo de qq-plot de resíduos não-Normais

Utilizando a mesma idéia pode-se verificar se os resíduos Z_x possuem distribuição Normal Padrão através de testes estatísticos de qualidade de ajuste.

O teste de Kolmogorov-Smirnov (K-S), cuja hipótese nula é a variável seguir a distribuição de probabilidade pela qual os dados foram ajustados, compara as diferenças máximas D_n , em módulo, entre a função de distribuição acumulada teórica (Normal, para o caso) e a obtida pelos dados. A hipótese nula de normalidade é rejeitada a um nível de significância α se $\sqrt{n} \cdot D_n > K_\alpha$.

Quando n é grande, K_α possui valor $\sqrt{-\frac{\ln[(1-\alpha)/2]}{2n}} - \frac{1}{6n}$.

Já o teste Jarque-Bera (Jarque e Bera, 1980), cuja hipótese nula é a variável seguir a distribuição Normal de probabilidade, leva em consideração a assimetria (A) e o excesso de curtose (C) nos dados através da estatística

$$JB = \frac{n}{6} \left(A^2 + \frac{C^2}{4} \right)$$

que possui distribuição assintótica Qui-Quadrado com dois

graus de liberdade.

Adicionalmente, uma medida quantitativa que pode ser aplicada é o Teste Qui-Quadrado. Sob a hipótese nula de que o modelo adere bem aos dados, os resíduos possuem distribuição Normal Padrão e então a soma quadrática dos mesmos possui uma distribuição Qui-Quadrado cujo número de graus de liberdade é dado pelo número de resíduos menos o número de parâmetros do modelo.

$$\sum_{i=0}^n \frac{\left(d'_i - \overset{\circ}{d}_i\right)^2}{\text{Var}\left(\overset{\circ}{d}_i\right)} \sim \chi^2_{(n-p)} \quad (28)$$

2.4.3. Teste Qui-Quadrado de Pearson

Um teste de aderência usual é o que compara o número observado de eventos geradores (óbitos, por exemplo) com o número esperado, de acordo com a tábua biométrica graduada construída. Sob a hipótese nula de que os números observados e esperados são iguais para todas as idades, pode-se construir um teste unilateral à direita sobre a distribuição qui-quadrado (χ^2) cujo número de graus de liberdade é dado pelo número idades menos um.

$$\sum_{i=0}^n \frac{\left(d'_i - \overset{\circ}{d}_i\right)^2}{\overset{\circ}{d}_i} \sim \chi^2_{(n-1)} \quad (29)$$

2.4.4. Teste dos Desvios Acumulados

Neste teste, as idades devem ser divididas em um número aleatório de grupos de tamanhos iguais. É importante que este número seja realmente aleatório para não ser influenciado pelas observações. O teste procura por vícios ou um grande período de desvios do mesmo sinal.

Assumindo que os resíduos absolutos (diferença entre os números graduados e observados de mortes) são independentes e possuem média zero, a soma dos resíduos de um grupo terá média zero e variância igual à soma das variâncias. Para a graduação de μ_x e q_x , ocorre que para cada grupo:

$$\frac{\sum_{i=0}^n \left(d'_x - E_x^C \cdot \overset{\circ}{\mu}_x\right)}{\sqrt{\sum_{i=0}^n E_x^C \cdot \overset{\circ}{\mu}_x}} \sim \text{Normal}(0,1) \quad \frac{\sum_{i=0}^n \left(d'_x - E_x \cdot \overset{\circ}{q}_x\right)}{\sqrt{\sum_{i=0}^n E_x \cdot \overset{\circ}{q}_x \cdot \left(1 - \overset{\circ}{q}_x\right)}} \sim \text{Normal}(0,1) \quad (30)$$

A maior estatística (em módulo) dentre os grupos pode ser confrontada com a distribuição Normal Padrão, a fim de se realizar um teste bicaudal cuja hipótese nula é que não há blocos que apresentam desvios acumulados.

2.4.5. Teste dos Sinais de Mann-Whitney

O Teste dos Sinais também é conhecido como Teste de Mann-Whitney. Se os resíduos são aleatórios, espera-se que haja aproximadamente tantos resíduos positivos quanto negativos. Para verificar tal afirmação define-se que a distribuição do número de resíduos positivos n_1 (ou negativos n_2) seguirá a seguinte distribuição de probabilidade:

$$n_1 \sim \text{Binomial}\left(n, \frac{1}{2}\right) \text{ onde } n = n_1 + n_2 \text{ é o número de resíduos observados}$$

Com $n > 50$, pode-se utilizar a aproximação pela distribuição Normal:

$$n_1 \sim \text{Normal}\left(\frac{n}{2}, \frac{n}{4}\right)$$

Logo, a estatística $\frac{n_1 - n/2}{\sqrt{n}/2}$ tem distribuição Normal Padrão e pode ser

utilizada a fim de se realizar o teste bicaudal cuja hipótese nula é que não há excesso ou falta de resíduos positivos em relação aos negativos.

2.4.6. Teste de Agrupamento dos Sinais de Stevens

O Teste de Agrupamento dos Sinais também é conhecido como Teste de Stevens. Seja G o número de grupos formados por resíduos positivos, onde foram observados n_1 e n_2 números de resíduos positivos e negativos respectivamente. A estatística G convergirá para a seguinte distribuição de probabilidade:

$$G \sim \text{Normal}\left(\frac{n_1(n_2 + 1)}{n_1 + n_2}, \frac{(n_1 \cdot n_2)^2}{(n_1 + n_2)^3}\right) \quad (31)$$

Pode-se então efetuar o teste a fim de se rejeitar ou não a hipótese nula de que os grupos de sinais são formados de acordo com o comportamento estatístico esperado.

2.4.7. Teste da Rodada

Seguindo a idéia de análise de sinais, pode ser observada a seqüência de resíduos positivos e negativos. Define-se então o número de rodadas como o número de vezes em que há troca de sinais nos resíduos (ordenados por idade), mais um. Sendo assim, é ilustrada na Figura 2, como exemplo, uma seqüência de sinais de resíduos.

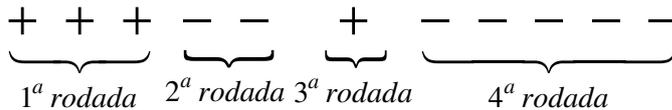


Figura 2 – Exemplo de onze resíduos perfazendo quatro rodadas

Se estes forem independentes o número de rodadas tem média e variância dadas respectivamente por $\frac{2n_1n_2}{n_1+n_2} + 1$ e $\frac{2n_1n_2[2n_1n_2 - (n_1+n_2)]}{(n_1+n_2)^2 \cdot (n_1+n_2-1)}$, onde novamente pode-se utilizar a estatística da distribuição Normal Padrão para o teste de hipótese bicaudal sob hipótese nula de que as rodadas são formadas de acordo com o comportamento estatístico esperado.

2.4.8. Teste de Correlação Serial

As seqüências z_i não devem ser autocorrelacionadas, como a exemplificada a seguir que possui tamanho $m-1$ e lag igual a 1 ($j=1$):

$$z_1, z_2, z_3, \dots, z_{m-2}, z_{m-1} \quad (32)$$

$$z_2, z_3, z_4, \dots, z_{m-1}, z_m$$

E assim para todos os $lags$, seja o coeficiente de correlação da seqüência dos j -ésimos $lags$ dado pela expressão a seguir, onde $\bar{z}^{(1)} = \frac{1}{m-j} \sum_{i=1}^{m-j} z_i$ e

$$\bar{z}^{(2)} = \frac{1}{m-j} \sum_{i=1}^{m-j} z_{i+j}.$$

$$r_j = \frac{\sum_{i=1}^{m-j} (z_i - \bar{z}^{(1)})(z_{i+j} - \bar{z}^{(2)})}{\sqrt{\sum_{i=1}^{m-j} (z_i - \bar{z}^{(1)})^2 \cdot \sum_{i=1}^{m-j} (z_{i+j} - \bar{z}^{(2)})^2}} \quad (33)$$

Se m é grande o suficiente aproxima-se o coeficiente de correlação para:

$$r_j \cong \frac{\sum_{i=1}^{m-j} (z_i - \bar{z})(z_{i+j} - \bar{z})}{\sqrt{\sum_{i=1}^{m-j} (z_i - \bar{z})^2 \cdot \sum_{i=1}^{m-j} (z_{i+j} - \bar{z})^2}} \quad (34)$$

Então é possível utilizar a distribuição Normal para testar o ajuste do modelo. Valores altos indicam uma tendência para os desvios do mesmo sinal se agruparem, rejeitando a hipótese de aleatoriedade.

Outra forma é aplicar o teste de Ljung-Box (Ljung e Box, 1978) também conhecido como teste Portmanteau, cuja hipótese nula é a ausência de autocorrelação. A estatística do teste é definida como $LB = n(n+2) \sum_{j=1}^h \frac{r_j^2}{n-j}$ e possui distribuição Qui-Quadrado com h graus de liberdade, onde h é a quantidade de *lags* a serem testados.

2.4.9. Teste dos Limites

Este teste consiste em construir um intervalo de confiança de $\gamma\%$ em torno das taxas graduadas e checar se as taxas brutas estão dentro destes limites em pelo menos $\gamma\%$ das vezes. A variância deve ser calculada de acordo com o modelo eleito, se Poisson ou Binomial. Como a distribuição assintótica do estimador é Normal, constrói-se o intervalo de confiança a partir a função de distribuição acumulada da mesma.

É interessante também procurar violações sistemáticas investigando faixas de idade com mau ajuste. Na Figura 3 não há violações sistemáticas e existem três pontos fora do intervalo de confiança de 95%. A ocorrência de 3 pontos em 81 observações (3,7%) está dentro no esperado para o grau de confiança escolhido.

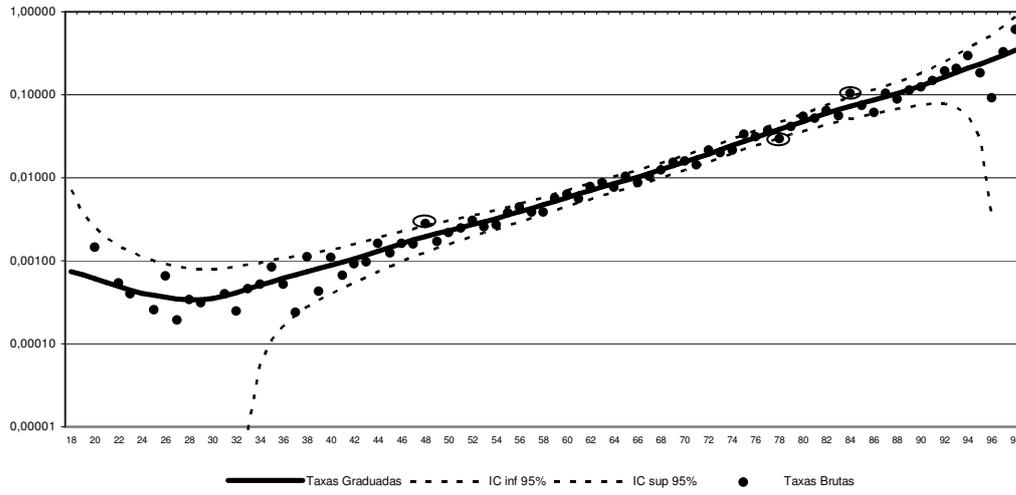


Figura 3 - Exemplo de Teste dos Limites para um intervalo de confiança de 95%