



1150022642



IMECC

T/UNICAMP Z11m

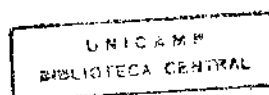
MODELOS LINEARES GENERALIZADOS NA MODELAGEM DA FECUNDIDADE MARITAL NO PERU

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida por **Daniela Rosa Zacharias** e aprovada pela Comissão Julgadora.

Campinas, 09 de agosto de 1994

Prof. Dra. ~~Aída~~ C.G. Verdugo Lazo

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para obtenção do Título de MESTRE em Estatística



AGRADECIMENTOS

Gostaria de expressar os meus sinceros agradecimentos a aquelas pessoas e instituições que ajudaram na realização deste trabalho.

-Ao Conselho Nacional de Pesquisa (CNPq) pelo apoio financeiro otorgado para este curso de pós-graduação.

-Ao IMECC, donde pasei dois anos e meio, disfrutando da sua biblioteca, laboratório e cafezinho.

-Aos professores que me cederam os seus conhecimentos e que sempre tiveram um tempo para solucionar as minhas dúvidas.

-A Ana Paula e Nívea, por tirar meu "sotaque" da minha tese.

-Aos amigos de longe e aos amigos de perto. Os primeiros com a sua amizade forte e antiga, os segundos com aquela novinha, mas que me ensinou que existem pessoas queríveis.

-A minha família de longe e de perto. Pelas cartas e ligações telefónicas, que sempre me trouxeram os carinhos e forças necessárias para me sentir acompanhada.

-A Carola, minha "gêmea", por sua tranquilidade para tomar decisões importantes (se não fosse por isso, eu jamás teria vindo a Campinas). Obrigada também por a sua paciência, compreensão e amizade que fizeram deste tempo, um tempo feliz ainda nos momentos de desespero.

-Por último, mas não de final, a minha orientadora Aida, cujo nome significa "ajuda", pela sua confiança, pela sua força, e todas as conversas estatísticas... e nem tanto.

Sumário

Lista de Figuras	3
Lista de Tabelas	4
0.1 INTRODUÇÃO	5
1 MODELOS LINEARES GENERALIZADOS	8
1.1 Componentes dos Modelos Lineares Generalizados	8
1.1.1 Componente Aleatória	9
1.1.2 Componente Sistemática	11
1.1.3 Função de Ligação	13
1.2 Análise da Componente Aleatória	14
1.3 Estimação do Modelo	16
1.3.1 Estimadores de Máxima Verossimilhança	17
1.3.2 Considerações Específicas sobre os Estimadores de Máxima Verossimilhança	23
1.4 Avaliação do Ajuste do Modelo	24
1.4.1 Função Desvio	25
1.4.2 Estatística de Pearson Generalizada	28
1.4.3 Resíduos	28
1.5 Modelos Log-Lineares	30
1.5.1 Modelo Log-Poisson	31
1.6 Comentários	34
2 PADRÕES DE FECUNDIDADE MARITAL	35
2.1 Modelo de Nupcialidade de Coale	36
2.1.1 Função Densidade Padrão de Primeiros Casamentos	37
2.1.2 Padronização das Distribuições de Frequências de Primeiros Casamentos	44
2.1.3 Reparametrização da Função de Densidade Padrão de Primeiros Casamentos	45
2.2 Modelo de Fecundidade	47

2.2.1	Modelo de Fecundidade de Coale e Trussell	47
2.2.2	Modelo de Fecundidade Marital Page	52
2.3	Comentários	57
3	MODELO LOG-POISSON PARA A FECUNDIDADE MARITAL	59
3.1	Aproximação Log-Linear	59
3.1.1	Bondade de Ajuste do Modelo	61
3.2	Interpretação dos Parâmetros	62
3.2.1	Parâmetro α	63
3.2.2	Parâmetro β	64
3.3	Comentários	64
4	ANÁLISE DA FECUNDIDADE MARITAL NO PERU ATRAVÉS DO MODELO LOG-POISSON	66
4.1	Peru: Principais Aspectos de Interesse	67
4.1.1	Regiões Geográficas	67
4.1.2	Área de Residência	68
4.1.3	Educação	72
4.2	Os Dados	74
4.2.1	Fontes	74
4.2.2	Processamento	77
4.3	Aplicação do Modelo	79
4.4	Resultados	80
4.4.1	Resultados Gerais	80
4.4.2	Efeito do Tipo de União	83
4.4.3	Efeito da Área de Residência	84
4.4.4	Efeito da Região de Residência	85
4.4.5	Efeito da Educação	90
4.5	Comentários	91
5	CONSIDERAÇÕES FINAIS	93
A	Teoria Estatística	95
B	Aspectos Computacionais	104
C	A Evolução da Fecundidade no Peru	111
D	Mapas	114
	Referências Bibliográficas	117

Lista de Figuras

2.1	Frequência de primeiros casamentos por idade, para quatro populações selecionadas.	37
2.2	Frequência de primeiros casamentos em populações selecionadas, com escala e origem ajustados.	38
2.3	Risco empírico , baseado na nupcialidade da Suécia (1865-1869), e ajustado por uma função dupla-exponencial.	40
2.4	Comparação das frequências dos primeiros casamentos de Suécia (1865-1869) e a função de densidade padrão $g(x)$	43
2.5	Valores de m , por idade, para populações selecionadas.	53
2.6	Valores de m , por idade. Peru, anos 1977, 1986, 1992.	54
4.1	Taxas de fecundidade marital por idade. Peru 1977-1992.	82
D.1	Peru: Divisão Política	114
D.2	Peru: Regiões Geográficas	115
D.3	Peru: Regiões de Planejamento (1977)	116

Lista de Tabelas

4.1	Distribuição urbano-rural da população. Peru 1960-1990.	69
4.2	Evolução dos domicílios em situação de pobreza e indigência, segundo área de residência. Peru 1970-1991	71
4.3	Taxa de fecundidade total observada e desejada, segundo área. Peru, 1986-1992	72
4.4	Índices de espaçamento, de limitação, bondade de ajuste, e $TFMT$, por tipo de união. Peru 1977-1992	81
4.5	Índices de espaçamento, de limitação, bondade de ajuste, e $TFMT$, por área de residência e tipo de união. Peru 1977-1992	85
4.6	Índices de espaçamento, de limitação, bondade de ajuste, e $TFMT$, por região de residência e tipo de união. Peru 1977-1992	87
4.7	Índices de espaçamento, de limitação, bondade de ajuste, e $TFMT$, por região de residência e tipo de união. Peru 1977.	88
4.8	Índices de espaçamento, de limitação, bondade de ajuste, e $TFMT$, por educação e tipo de união. Peru 1977-1992.	89
C.1	Taxa Fecundidade Total (TGF). Peru 1950-1990.	112
C.2	Fecundidade segundo fatores diferenciais. Peru 1977- 1992	113
C.3	Fecundidade segundo grupos etários	113

0.1 INTRODUÇÃO

Seguindo a estrutura da presente tese, apresenta-se nesta introdução as motivações “estatísticas” e as motivações “demográficas” que lhe deram origem. Estas duas abordagens serão logo “unidas” formando aquele “casamento” tão útil, mas não muitas vezes realizado, entre as ciências exatas e as ciências sociais. Origina-se então o objetivo principal deste trabalho: a aplicação dos modelos lineares generalizados para modelar a fecundidade marital.

Os modelos lineares generalizados têm sido um tanto proletados no ensino e estudo da estatística, onde os modelos lineares clássicos ocupam uma maior posição de destaque. Porém, como será mostrado nos capítulos seguintes, os modelos lineares clássicos não são senão um caso particular da classe dos modelos lineares generalizados. Esta classe está constituída por um imenso conjunto de modelos, onde a única condição para a função de distribuição da variável resposta é a de pertencer à família exponencial. Além disso, o conjunto de modelos que esta classe inclui cresce muito mais, ao permitir que a relação entre o valor esperado da variável resposta e o preditor linear (determinado pela combinação linear do conjunto de covariáveis) esteja dada por uma função de ligação entre eles, com a única condição de ser monótona e duas vezes diferenciável.

Nelder e Wedderburn (1972) notam a variedade de tipos de variáveis resposta que podem ser representadas ao fazer a extensão mencionada acima. Estes autores também mostram que apesar das condições dos modelos lineares clássicos terem sido “suavizadas”, as propriedades dos estimadores de máxima verossimilhança não se alteram. Nesse mesmo artigo os autores mostram como estimar os parâmetros envolvidos (através do método de mínimos quadrados ponderados iterativo), definem a *deviance*, que neste trabalho será chamada *função desvio*, e analisam conjuntos de dados sob esta óptica. A partir do artigo de Nelder e Wedderburn (1972) surge na literatura uma série de trabalhos sobre este tema. Nelder e McCullagh publicam, em 1983, o livro *Generalized Linear Models*, que representa uma excelente bibliografia sobre o tema.

Dentre dos autores brasileiros, Gauss Cordeiro é quem mais se dedicou ao

estudo destes modelos. No seu livro *Modelos Lineares Generalizados*, apresentado no VIII SINAPE (1986), além de incluir uma completa introdução teórica, fornece os programas para ser rodados com a utilização do pacote GLIM (Generalized Linear Interative Models) para uma grande quantidade de exercícios apresentados.

O livro mais recente constando na revisão bibliográfica deste tema é *An Introduction to Generalized Linear Models* de Anette Dobson (1990) que, apesar de utilizar uma linguagem simples, consegue ir além de uma introdução no tema dos modelos lineares generalizados.

Em relação às motivações "demográficas", o interesse pela modelagem da fecundidade marital surge como uma inquietude pessoal, de tipo social, frente às mudanças que vem acontecendo no mundo, e como estas se refletem no comportamento reprodutivo da população e, em particular, na população do Peru, país de residência da autora desta tese.

Rodriguez e Cleland (1988), num trabalho baseado em estudos anteriores sobre o modelo de fecundidade marital de Page (1977), utilizam a teoria dos modelos lineares generalizados, para representar o comportamento reprodutivo das mulheres que vivem em união. A análise destes autores foi reconstruída e utilizada para modelar a fecundidade marital no Peru.

Resultou uma sensação diferenciada, desde o ponto de vista estatístico-matemático, trabalhar com dados reais, onde cada observação da variável resposta não provinha de uma simulação realizada com o computador, mas que representava o número de filhos que teve uma mulher com uma idade e duração de união determinada, inserida em um certo contexto econômico, social e cultural, que estaria afetando a sua fecundidade. Esta sensação distinta foi porém reconfortante, ao poder utilizar um tópico da teoria estatística para intentar explicar um comportamento humano, como o comportamento reproductivo das mulheres, e assim poder contribuir com estes resultados à elaboração de possíveis planos nacionais que, de alguma forma, diminuam a desigualdade existente entre as mulheres peruanas no que se refere às oportunidades que têm de determinar sua fecundidade.

O trabalho apresentado a seguir está dividido em quatro capítulos. No primeiro capítulo apresenta-se a teoria dos modelos lineares generalizados, as suas

componentes e métodos de estimação dos parâmetros, discutindo particularmente o modelo log-Poisson.

No segundo capítulo apresenta-se a teoria dos modelos demográficos referente aos tópicos de nupcialidade e fecundidade marital. Mostra-se a evolução destes modelos, começando com o modelo de Coale (1971) até o modelo de Page (1977). No terceiro capítulo relacionam-se os temas discutidos nos dois capítulos anteriores, através da aproximação log-linear que Rodriguez e Cleland (1988) fazem do modelo de fecundidade marital de Page.

No quarto e último capítulo aplica-se o modelo log-Poisson com o fim de modelar a fecundidade marital no Peru, utilizando dados coletados em três levantamentos realizados no país, entre os anos 1977 e 1992. Os resultados obtidos são analisados segundo os efeitos do tipo de união, da área de residência, da região de residência, e do nível de instrução das mulheres entrevistadas.

Finalmente, espera-se que este trabalho tenha uma boa acolhida tanto por pesquisadores da área de ciências exatas como da área de ciências sociais, em particular da demografia. Foi com este intuito que se tentou utilizar uma linguagem acessível a ambas as áreas simultaneamente. Portanto, pede-se paciência aos leitores que encontrem explicações redundantes.

Capítulo 1

MODELOS LINEARES GENERALIZADOS

Nelder & Wedderburn (1972) desenvolveram a teoria de uma classe de modelos que denominaram como a classe dos *modelos lineares generalizados*. Esta classe reúne os modelos caracterizados por uma função de probabilidade membro da família exponencial para a variável resposta (ou variável independente), um conjunto de covariáveis (ou variáveis independentes) que descrevem a estrutura linear do modelo e uma função de ligação entre o valor esperado da variável resposta e a estrutura linear.

1.1 Componentes dos Modelos Lineares Generalizados

Os modelos lineares generalizados estão caracterizados por duas componentes: uma **aleatória** e uma **sistemática** e, uma função de relação entre elas que é chamada **função de ligação**. Estas três características que definem os modelos lineares generalizados são explicadas nesta seção.

1.1.1 Componente Aleatória

A parte aleatória dos modelos lineares generalizados está definida pela função de distribuição da variável resposta Y .

Esta variável é, na realidade, um vetor aleatório, cujas componentes Y_i , $i = 1, \dots, n$ são variáveis aleatórias independentes com distribuição pertencente à *família exponencial*.

Considere-se uma variável aleatória Y_i , cuja função de probabilidade (se Y_i é discreta) ou função de densidade (se Y_i é contínua) ¹ depende dos parâmetros θ_i e ϕ_i , e se denota $f(y_i; \theta_i, \phi_i)$. A densidade ou função de probabilidade da variável aleatória Y_i pertence à família exponencial se pode ser escrita da forma:

$$f(y_i; \theta_i, \phi_i) = \exp\{\alpha(\phi_i)[y_i \theta_i - g(\theta_i)] + c(y_i, \phi_i)\}, \quad (1.1)$$

onde:

y_i : realização da variável aleatória Y_i ,²

$\theta = (\theta_1, \dots, \theta_n)$ parâmetro que caracteriza a função dada em (1.1) e que é chamado parâmetro *canônico* ou *natural*,

$\phi = (\phi_1, \dots, \phi_n)$ parâmetro de *escala*, geralmente conhecido,

$\alpha(\cdot)$, $g(\cdot)$, $c(\cdot)$ são funções conhecidas e $\alpha(\phi_i) > 0$, $\forall i = 1, \dots, n$.

As distribuições normal, exponencial, Poisson, binomial, binomial negativa, gama, inversa Gaussiana, entre outras, podem ser escritas da forma (1.1). Portanto, se o modelo de probabilidade³ da variável resposta Y corresponde a um dos anteriores, estará incluído dentro da família exponencial. (Ver por exemplo: Cordeiro (1986), Nelder e McCullagh (1989), Dobson (1990)).

¹Ver apêndice A1

²Utiliza-se também: $y = (y_1, \dots, y_n)$, vetor de realização do vetor aleatório $Y = (Y_1, \dots, Y_n)$

³Ver apêndice A1.

Considere-se como exemplo, uma variável aleatória discreta Z , com modelo de probabilidade Poisson, cuja função de probabilidade é dada por

$$f(z; \mu) = \frac{\mu^z e^{-\mu}}{z!}$$

onde z assume os valores $0, 1, 2, \dots$.

Note-se que $f(z; \mu)$ pode ser escrita como

$$f(z, \mu) = \exp [z \ln(\mu) - \mu - \ln(z!)].$$

Identificando $\theta = \ln(\mu)$, $g(\theta) = \mu = e^\theta$, $\alpha(\phi) = 1$ e $c(z, \phi) = -\ln(z!)$, tem-se que a função de probabilidade da variável aleatória Z pode ser escrita da forma dada em (1.1). Portanto, a modelo de probabilidade Poisson pertence à família exponencial.

Como outro exemplo, pode se considerar a distribuição binomial, correspondente a uma variável aleatória discreta W , que representa o número de *sucessos* ocorridos em n ensaios independentes, onde a probabilidade de sucesso (π) é a mesma em todos os ensaios. A função de probabilidade da variável aleatória W é dada por:

$$f(w; \pi) = \binom{n}{w} \pi^w (1 - \pi)^{n-w},$$

onde w assume os valores $0, 1, 2, \dots, n$.

Esta função de probabilidade pode ser escrita da seguinte forma:

$$f(w; \pi) = \exp \left[w \ln \pi - w \ln(1 - \pi) + n \ln(1 - \pi) + \ln \binom{n}{w} \right]$$

$$f(w; \pi) = \exp \left[w \ln \frac{\pi}{1 - \pi} + n \ln(1 - \pi) + \ln \binom{n}{w} \right]$$

Identificando $\theta = \ln \frac{\pi}{1 - \pi}$, $g(\theta) = -n \ln(1 - \pi) = n \ln(1 + e^\theta)^4$, $\alpha(\phi) = 1$ e $c(w, \phi) = \ln \binom{n}{w}$, tem-se que o modelo binomial pertence à família exponencial já que a função de probabilidade pode ser escrita da forma dada em (1.1).

O conjunto de distribuições pertencentes à família exponencial permite trabalhar com uma grande gama de variáveis resposta que representam diversos tipos

⁴Pode-se obter $\pi = \frac{e^\theta}{1 + e^\theta}$ a partir da primeira relação e substituindo este resultado na equação de $g(\theta)$, se obtém $-n \ln(1 - \pi) = -n \ln(1 + e^\theta)$.

de fenômenos, como os modelos com observações contínuas e variância constante (modelos lineares clássicos), observações binárias ou categóricas, contagens, análise de sobrevivência, etc.

Dentro deste grande conjunto de situações que podem ser estudadas sob a ótica dos modelos lineares generalizados, encontra-se a variável resposta que será tratada e modelada nos próximos capítulos e que capta o comportamento da fecundidade marital, tema que será abordado nos capítulos seguintes. Essa variável resposta refere-se ao número de filhos tidos por mulher, num período de tempo determinado. As contagens dos nascimentos por mulher são variáveis independentes com distribuição Poisson (mais detalhes serão dados no capítulo 3) e, portanto, o seu estudo pode ser feito através de um modelo linear generalizado.

1.1.2 Componente Sistemática

A componente sistemática dos modelos lineares generalizados constitui-se de um conjunto de *covariáveis*⁵, x_{11}, \dots, x_{1p} , para cada componente y_i da variável resposta y , que atuam só como uma soma linear de seus efeitos, definindo o *preditor linear* $\eta = (\eta_1, \dots, \eta_n)$ como

$$\eta_i = \sum_{j=1}^p x_{ij} \beta_j \quad i = 1, \dots, n \quad (1.2)$$

onde $\beta = (\beta_1, \dots, \beta_p)$ é o *vetor de parâmetros*.

Estas covariáveis se organizam numa matriz que será chamada *matriz do modelo* e denotada, neste trabalho, pela letra X .

A matriz X tem uma linha por cada observação e uma coluna por cada covariável. O elemento x_{ij} , que corresponde à i -ésima linha e à j -ésima coluna da matriz X , é igual ao valor que assume a j -ésima covariável na i -ésima observação.

As covariáveis podem ser *quantitativas* ou *qualitativas*.

As covariáveis quantitativas são aquelas que assumem valores numéricos. Por exemplo, a variável “idade da mulher ao casar” ou “duração da união”, que

⁵Outros nomes encontrados na literatura são: variáveis independentes, variáveis explicativas, preditores, *stimulus*.

podem ser medidas em anos, meses ou dias.

As covariáveis de tipo qualitativo são aquelas que assumem valores não-numéricos que representam a categoria, nível ou grupo ao qual pertence uma observação, dentro de uma classificação para a população. Um exemplo deste tipo de covariável é o tipo de união (formal ou consensual). Pode ser representada pelos valores 1 e 2, respectivamente, mas isto não significa superioridade ou inferioridade de uma forma de união em relação à outra. As covariáveis qualitativas são chamadas de *fatores*.

O preditor linear está determinado pela matriz do modelo e o vetor de parâmetros lineares. Assume-se que os valores da matriz X foram observados sem erros. As componentes do vetor β podem ser conhecidas ou não e, neste último caso, precisam ser estimadas. O método de estimação do vetor de parâmetros para os modelos lineares generalizados será visto na seção 1.3.

Para ilustrar os conceitos anteriores, considere-se um conjunto de n mulheres casadas. Um preditor linear que depende das variáveis "idade ao casar" e "duração da união" pode ser definido como

$$\eta_i = a_i + \beta_1 d_i \quad i = 1, \dots, n$$

onde :

n é o número total de observações

a_i é a idade ao casar da i -ésima mulher

e d_i é a duração do matrimónio da i -ésima mulher

Neste caso, a matriz do modelo e o vetor de parâmetros podem ser identificados como:

$$X = \begin{bmatrix} a_1 & d_1 \\ a_2 & d_2 \\ \vdots & \vdots \\ a_n & d_n \end{bmatrix} \quad e \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$$

com $\beta_0 = 1$ e β_1 desconhecido, a ser estimado. O primeiro termo do preditor linear (dado por a_1, \dots, a_n) será então conhecido⁶. Desta maneira, tem-se que o preditor

⁶Para denominar os termos conhecidos do preditor linear utiliza-se a palavra inglesa *offset*.

linear está definido pelas covariáveis "idade da mulher" e "duração da união", onde esta última covariável está multiplicada por um parâmetro desconhecido que indica a participação que esta tem no preditor linear.

Se as mulheres estudadas são classificadas por algum critério, como área de residência (urbano-ou rural), por exemplo, este fator classificatório pode ser considerado dentro da matriz do modelo de várias maneiras. Uma destas seria agregar mais duas colunas adicionais à matriz X , cada uma representando um tipo de área de residência. Na primeira coluna codifica-se com 1 as mulheres urbanas e com 0 as mulheres rurais, fazendo o contrário para a segunda coluna. Uma outra forma seria trabalhar só com uma coluna adicional na matriz X , onde cada componente assume o valor 1 para as mulheres urbanas e -1 para as mulheres rurais. O uso de variáveis *dummy*⁷ também é muito comum na definição da matriz do modelo em pacotes computacionais. Estas diferentes formas de considerar dentro da matriz do modelo os fatores classificatórios representam diferentes formas de reparametrização do modelo.

1.1.3 Função de Ligação

A função de ligação relaciona o valor esperado da variável resposta (componente aleatória) com o preditor linear (componente sistemática) da seguinte maneira:

$$\mu_i = f^{-1}(\eta_i) \quad i = 1, \dots, n$$

onde $\mu_i = E(Y_i)$, valor esperado da variável Y_i , $i = 1, \dots, n$.

A função f , que deve ser monótona e duas vezes diferenciável, é chamada *função de ligação*.

Quando a relação entre a componente aleatória e sistemática do modelo é dada pela igualdade entre o preditor linear e o parâmetro canônico, i.e., satisfaz a:

$$\eta = \theta$$

⁷A variável *dummy* representa os n níveis do fator por $n - 1$ colunas adicionais à X com valores 0 e 1. Na primeira coluna a observação assume o valor 1 se pertence ao primeiro nível do fator e 0 no caso contrário. Este critério se repete nas $n-2$ colunas restantes. A observação que pertence ao n -ésimo nível terá assumido o valor 0 nas $n - 1$ colunas.

se diz que a função de ligação é *canônica*.

Para a distribuição Poisson, por exemplo, tem-se que $\mu = e^\theta$, então a função logaritmo será uma função de ligação canônica pois

$$\eta = \log(\mu) = \theta$$

As funções de ligação canônicas dão aos modelos lineares generalizados algumas propriedades importantes, entre elas, a existência de estatísticas suficientes de dimensão mínima p para o vetor de parâmetros, dadas por: $S_j = \sum_{i=1}^n \alpha(\phi_i) x_{ij} Y_i$ $j = 1, \dots, p$ e a unicidade dos estimadores de máxima verossimilhança.

A primeira propriedade é facilmente provada mediante o Critério de Fatoração de Neymann-Pearson (Ver apêndice A2), e a segunda será verificada na seção 1.3.

1.2 Análise da Componente Aleatória

A caracterização da distribuição da componente aleatória, como pertencente a uma família exponencial de um parâmetro, permite obter resultados sobre a forma da função de verossimilhança, o valor esperado e a variância.

Função de Log-Verossimilhança

A função de log-verossimilhança, denotada por $\ell(\theta, \phi; \mathbf{y})$, é o logaritmo natural da função de verossimilhança, que será denotada por $L(\theta, \phi; \mathbf{y})$, e que é igual a $\prod_{i=1}^n f(y_i; \theta_i, \phi_i)$ ⁸.

Tem-se então, no caso particular das funções de distribuição pertencentes à família exponencial, que a função de log-verossimilhança está dada por:

$$\ell(\theta, \phi; \mathbf{y}) = \ln(L(\theta, \phi; \mathbf{y})) = \ln\left(\prod_{i=1}^n f(y_i; \theta_i, \phi_i)\right) = \sum_{i=1}^n \{\alpha(\phi_i)[y_i \theta_i - g(\theta_i)] + c(y_i, \phi_i)\} \quad (1.3)$$

⁸Note-se que tanto a função de verossimilhança como a função de log-verossimilhança são funções dos parâmetros. Uma vez que uma amostra é extraída, estas funções representam a plausibilidade dos parâmetros serem os parâmetros verdadeiros da população.

Valor Esperado e Variância:

A média e a variância da variável resposta Y podem ser derivadas das relações a seguir (ver apêndice A3), onde ℓ representa a função de log-verossimilhança $\ell(\theta, \phi; y)$ definida em (1.3).

$$E\left(\frac{\partial \ell}{\partial \theta_i}\right) = 0 \quad (1.4)$$

$$E\left(\frac{\partial^2 \ell}{\partial \theta_i^2}\right) = -E\left(\frac{\partial \ell}{\partial \theta_i}\right)^2 \quad (1.5)$$

para $i = 1, \dots, n$

Derivando a função ℓ com respeito a θ_i , tem-se:

$$\frac{\partial \ell}{\partial \theta_i} = \alpha(\phi_i)(y_i - g'(\theta_i)) \quad (1.6)$$

$$\frac{\partial^2 \ell}{\partial \theta_i^2} = -\alpha(\phi_i)(g''(\theta_i)) \quad (1.7)$$

De (1.4) e (1.6), tem-se:

$$\alpha(\phi_i)E(Y_i) = \alpha(\phi_i)(g'(\theta_i)),$$

logo:

$$\boxed{E(Y_i) = \mu_i = g'(\theta_i)} \quad (1.8)$$

De (1.4), (1.6) e (1.7) tem-se:

$$\begin{aligned} -\alpha(\phi_i)(g''(\theta_i)) &= -\alpha^2(\phi_i)[E(Y_i^2) - 2g'(\theta_i)E(Y_i) + (g'(\theta_i))^2], \quad \text{por (1.8)} \\ &= -\alpha^2(\phi_i)[E(Y_i^2) - 2E^2(Y_i) + E^2(Y_i)] \\ &= -\alpha^2(\phi_i)Var(Y_i) \end{aligned}$$

Logo:

$$\boxed{Var(Y_i) = (g''(\theta_i))/\alpha(\phi_i)} \quad (1.9)$$

Os resultados (1.8) e (1.9) indicam que a variância de Y_i é o quociente de duas funções: $g''(\theta_i)$, dependente do parâmetro canônico θ_i através da função $g'(\theta_i)$, i.e., de μ_i (valor esperado de Y_i), e de uma outra função, $\alpha(\phi_i)$, que depende somente do parâmetro de escala ϕ_i . A função $g''(\theta_i)$ será chamada *função variância* e denotada por V_i . Em outro termos,

$$V_i = g''(\theta_i) \quad \text{e} \quad \text{Var}(Y_i) = V_i / \alpha(\phi_i). \quad (1.10)$$

Como exemplo pode-se considerar o caso da distribuição Poisson. Para esta distribuição $g(\theta) = e^\theta$ e $\alpha(\phi) = 1$, logo:

$$E(Y_i) = g'(\theta_i) = \frac{\partial}{\partial \theta_i}(e^{\theta_i}) = e^{\theta_i},$$

$$\text{Var}(Y_i) = \frac{g''(\theta_i)}{\alpha(\phi_i)} = \frac{\partial^2}{\partial \theta_i^2}(e^{\theta_i}) = e^{\theta_i},$$

e

$$V_i = g''(\theta_i) = e^{\theta_i}.$$

Neste caso, o valor esperado, a variância e a função variância coincidem.

1.3 Estimação do Modelo

Sintetizando o mencionado nas seções anteriores, tem-se que os modelos lineares generalizados podem ser expressos da seguinte maneira:

$$f(\mu) = X\beta,$$

onde

μ : $E(Y)$, valor esperado da variável resposta Y (vetor de dimensão n , onde n é o número de observações)

X : matriz do modelo, de dimensão $n \times p$

β : vetor de parâmetros, de dimensão p

f : função de ligação

Portanto, estimar o modelo é equivalente a estimar as componentes do vetor de parâmetros β . Um dos métodos utilizados com este propósito é o *método da máxima verossimilhança*. Este método consiste em maximizar a função de verossimilhança, ou de forma equivalente, a função de log-verossimilhança. Os valores de β que maximizam esta função serão chamados de estimadores de máxima verossimilhança do vetor de parâmetros β .

No caso especial dos modelos lineares generalizados, o método utilizado para encontrar o valor que maximiza esta função de verossimilhança é equivalente ao método de mínimos quadrados ponderados iterativo⁹.

1.3.1 Estimadores de Máxima Verossimilhança

Nesta seção mostrar-se-á que o estimador de máxima verossimilhança (EMV) do vetor de parâmetros β no preditor linear η , solução das equações de máxima verossimilhança, é equivalente à solução de um procedimento de mínimos quadrados ponderados iterativo com função peso $w = (\frac{\partial \mu}{\partial \eta})^2 / V$ e uma variável dependente modificada $z = \eta + (y - \mu)(\frac{\partial \mu}{\partial \eta})$.

Será mostrado, inicialmente, que as equações de máxima verossimilhança para β_j são dadas por:

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^n \alpha(\phi_i) w_i (y_i - \mu_i) \frac{\partial \eta_i}{\partial \mu_i} x_{ij} = 0 \quad \text{para } j = 1, \dots, p, \text{ onde } w_i = \frac{\left(\frac{\partial \mu_i}{\partial \eta_i}\right)}{V_i} \quad (1.11)$$

Demonstração:

A partir da definição dada em (1.3) tem-se que no caso dos modelos lineares generalizados a função de log-verossimilhança é:

$$\ell = \sum_{i=1}^n \{ \alpha(\phi_i) [y_i \theta_i - g(\theta_i)] + c(y_i, \phi_i) \}. \quad (1.12)$$

⁹Para uma descrição deste método ver apêndice A4.

Pela regra da cadeia tem-se:

$$\frac{\partial \ell}{\partial \beta_j} = \frac{\partial \ell}{\partial \theta_i} \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_h} \frac{\partial \eta_h}{\partial \beta_j}.$$

Também sabe-se, de (1.8) e (1.9) que

$$\mu_i = g'(\theta_i) \quad e \quad V_i = g''(\theta_i)$$

portanto,

$$\frac{\partial \theta_i}{\partial \mu_i} = \frac{1}{V_i}. \quad (1.13)$$

Da equação (1.2) tem-se

$$\eta_h = \sum_{j=1}^p x_{ij} \beta_j,$$

logo,

$$\frac{\partial \eta_h}{\partial \beta_j} = x_{ij} \quad (1.14)$$

Derivando (1.12) com respeito a β_j , tem-se:

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^n \alpha(\phi_i)(y_i - g'(\theta_i)) \frac{\partial \theta_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_h} \frac{\partial \eta_h}{\partial \beta_j}.$$

Substituindo-se na equação anterior os resultados (1.13) e (1.14), tem-se:

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^n \alpha(\phi_i)(y_i - \mu_i) \frac{1}{V_i} \frac{\partial \mu_i}{\partial \eta_h} x_{ij}.$$

Fazendo $w_i = \left(\frac{\partial \mu_i}{\partial \eta_h} \right)^2$ obtem-se

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^n \alpha(\phi_i)(y_i - \mu_i) w_i \frac{\partial \eta_h}{\partial \mu_i} x_{ij}.$$

Tem-se então as equações de máxima verosimilhança:

$$\frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^n \alpha(\phi_i) w_i (y_i - \mu_i) \frac{\partial \eta_h}{\partial \mu_i} x_{ij} = 0 \quad \text{para } j = 1, \dots, p.$$

Mas, estas equações, exceto para o modelo normal ($V_i = 1$, $i = 1, \dots, n$ e função de ligação identidade), são não lineares em β não têm solução algébrica. São necessárias técnicas numéricas para obter o estimador de máxima verossimilhança do vetor de parâmetros β .

Uma dessas técnicas é o método de *score de Fisher*, que foi utilizado por Nelder e Wedderburn (1972) para encontrar as soluções das equações de máxima verossimilhança, e que será explicado a seguir.

Define-se o vetor $u(\beta)$, de dimensão p , cuja j -ésima componente é a derivada da função de log-verossimilhança em relação a β_j , isto é.

$$u_i(\beta) = \frac{\partial \ell}{\partial \beta_i}, \quad i = 1, \dots, p. \quad (1.15)$$

Trata-se agora de encontrar $\hat{\beta}$, tal que $u(\hat{\beta}) = 0$. A solução $\hat{\beta}$ destas equações é a solução das equações de máxima verossimilhança, para os parâmetros do modelo.

Para encontrar a solução deste último sistema utiliza-se o método iterativo de Newton-Raphson¹⁰. Este procedimento consiste em encontrar β^* , solução de:

$$u(\beta) = A(\beta - \beta^*), \quad (1.16)$$

onde A é uma matriz de dimensão $p \times p$, com $A_{rs} = \frac{\partial^2 \ell}{\partial \beta_r \partial \beta_s}$, para $r, s = 1, \dots, p$ e β é suposto conhecido, escolhido como um valor inicial para o processo iterativo.

O método de *score de Fisher* é introduzido quando A_{rs} , para $r, s = 1, \dots, p$, é substituído pelo seu valor esperado $E(A_{rs})$, ou seja, quando $\frac{\partial^2 \ell}{\partial \beta_r \partial \beta_s}$ é substituído por $E\left(\frac{\partial^2 \ell}{\partial \beta_r \partial \beta_s}\right)$.

Uma vez encontrado β^* , solução de (1.16), substitui-se β por β^* e a equação (1.16) é resolvida novamente. Este processo é repetido até a convergência, i.e., quando a diferença entre o valor de β^* na n -ésima iteração e o valor de β^* na $(n + 1)$ -ésima iteração for muito pequena.

Este procedimento corresponde a um processo de estimação através do método de mínimos quadrados ponderados iterativo, quando se trabalha com modelos lineares generalizados, o que é demonstrado a seguir.

¹⁰Ver por exemplo, Burden (1989, p.538).

No caso dos modelos lineares generalizados, onde a função de log-verossimilhança tem a forma particular dada em (1.3), tem-se

$$E \left(\frac{\partial^2 \ell}{\partial \beta_r \partial \beta_s} \right) = - \sum_{i=1}^n \alpha(\phi_i) w_i x_{ir} x_{is}, \quad \text{onde } w_i = \frac{\left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2}{V_i}. \quad (1.17)$$

Demonstração:

Com o objetivo de encontrar uma expressão para $E \left(\frac{\partial^2 \ell}{\partial \beta_r \partial \beta_s} \right)$, procura-se primeiro a forma de $\frac{\partial^2 \ell}{\partial \beta_r \partial \beta_s}$. Sabe-se que

$$\frac{\partial^2 \ell}{\partial \beta_r \partial \beta_s} = \sum_{i=1}^n \frac{\partial^2 \ell_i}{\partial \beta_r \partial \beta_s}, \quad (1.18)$$

onde

$$\ell_i = \alpha(\phi_i) [y_i \theta_i - g(\theta_i)] + c(y_i, \phi_i).$$

Tem-se

$$\frac{\partial^2 \ell_i}{\partial \beta_r \partial \beta_s} = \frac{\partial}{\partial \beta_r} \left(\frac{\partial \ell_i}{\partial \beta_s} \right)$$

e pela regra da cadeia,

$$\frac{\partial^2 \ell_i}{\partial \beta_r \partial \beta_s} = \frac{\partial}{\partial \beta_r} \left(\frac{\partial \ell_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_s} \right)$$

Do resultado dado em (1.14) tem-se:

$$\frac{\partial^2 \ell_i}{\partial \beta_r \partial \beta_s} = \frac{\partial}{\partial \beta_r} \left(\frac{\partial \ell_i}{\partial \eta_i} x_{is} \right)$$

e pela regra da cadeia,

$$\frac{\partial^2 \ell_i}{\partial \beta_r \partial \beta_s} = \frac{\partial}{\partial \eta_i} \left(\frac{\partial \ell_i}{\partial \eta_i} x_{is} \right) \frac{\partial \eta_i}{\partial \beta_r}$$

Utilizando novamente a equação (1.14) tem-se:

$$\frac{\partial^2 \ell_i}{\partial \beta_r \partial \beta_s} = \frac{\partial^2 \ell_i}{\partial \eta_i^2} x_{is} x_{ir} \quad (1.19)$$

Calcula-se agora $\frac{\partial^2 \ell_i}{\partial \eta_i^2}$,

Pela regra da cadeia,

$$\frac{\partial \ell_i}{\partial \eta_i} = \frac{\partial \ell_i}{\partial \theta} \frac{\partial \theta}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i}.$$

Derivando ℓ_i com respeito a θ , e substituindo o resultado (1.13) tem-se:

$$\frac{\partial \ell_i}{\partial \eta_i} = \alpha(\phi_i)(y_i - \mu_i) \frac{1}{V_i} \frac{\partial \mu_i}{\partial \eta_i}.$$

Derivando novamente com respeito a η_i :

$$\begin{aligned} \frac{\partial^2 \ell_i}{\partial \eta_i^2} &= \frac{\partial}{\partial \eta_i} \left(\alpha(\phi_i)(y_i - \mu_i) \frac{1}{V_i} \frac{\partial \mu_i}{\partial \eta_i} \right) \\ &= \alpha(\phi_i) \frac{\partial}{\partial \eta_i} (y_i - \mu_i) \frac{1}{V_i} \frac{\partial \mu_i}{\partial \eta_i} + \alpha(\phi_i)(y_i - \mu_i) \frac{\partial}{\partial \eta_i} \left(\frac{\partial \theta}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \right) \\ &= \alpha(\phi_i) \left[- \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \frac{1}{V_i} + (y_i - \mu_i) \frac{\partial^2 \theta}{\partial \eta_i^2} \right]. \end{aligned} \quad (1.20)$$

Substituindo este último resultado em (1.19) tem-se:

$$\frac{\partial^2 \ell_i}{\partial \beta_r \partial \beta_s} = \alpha(\phi_i) \left[- \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \frac{1}{V_i} + (y_i - \mu_i) \frac{\partial^2 \theta}{\partial \eta_i^2} \right] x_{i,r} x_{i,s}$$

e, substituindo em (1.18),

$$\frac{\partial^2 \ell}{\partial \beta_r \partial \beta_s} = \sum_{i=1}^n \alpha(\phi_i) \left[- \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \frac{1}{V_i} + (y_i - \mu_i) \frac{\partial^2 \theta}{\partial \eta_i^2} \right] x_{i,r} x_{i,s}.$$

Considerando o valor esperado em ambos lados da equação anterior:

$$E \left(\frac{\partial^2 \ell}{\partial \beta_r \partial \beta_s} \right) = E \left[\sum_{i=1}^n \alpha(\phi_i) \left[- \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \frac{1}{V_i} + (y_i - \mu_i) \frac{\partial^2 \theta}{\partial \eta_i^2} \right] x_{i,r} x_{i,s} \right]. \quad (1.21)$$

Como o valor esperado do segundo termo dentro do colchete é zero e o primeiro termo é constante, pois não envolve nenhuma variável aleatória, tem-se:

$$\begin{aligned} E \left(\frac{\partial^2 \ell}{\partial \beta_r \partial \beta_s} \right) &= - \sum_{i=1}^n \alpha(\phi_i) \left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2 \frac{1}{V_i} x_{i,r} x_{i,s} \\ w_i &= \frac{\left(\frac{\partial \mu_i}{\partial \eta_i} \right)^2}{V_i}, \text{ obtem-se:} \\ E \left(\frac{\partial^2 \ell}{\partial \beta_r \partial \beta_s} \right) &= - \sum_{i=1}^n \alpha(\phi_i) w_i x_{i,r} x_{i,s}, \quad r, s = 1, 2, \dots, p, \end{aligned}$$

expressão que queria-se encontrar para $E\left(\frac{\partial^2 \ell}{\partial \beta_r \partial \beta_s}\right)$.

Omitindo-se o fator de dispersão $\alpha(\phi)$ e considerando a aproximação da matriz A pelas esperanças de seus componentes dadas por (1.17), tem-se:

$$A_{rs} = \frac{\partial^2 \ell}{\partial \beta_r \partial \beta_s} \cong - \sum_{i=1}^n w_i x_{ir} x_{is}, \text{ para } r, s = 1, 2, \dots, p, \quad \text{onde } w_i = \frac{\left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2}{V_i}$$

Ao aproximar cada componente da matriz A pelo seu valor esperado, tem-se que o sistema de equações (1.16), $u(\beta) = A(\beta - \beta^*)$, é equivalente ao seguinte sistema

$$K\beta^* = r, \quad (1.22)$$

$$\text{onde } K = -E(A) \quad \text{e} \quad r = \mu(\beta) + K\beta.$$

Lembrando que o vetor $\mu(\beta)$, de dimensão p , definido em (1.15), é igual à expressão abaixo (utilizando a equação (1.11))

$$u_j(\beta) = \frac{\partial \ell}{\partial \beta_j} = \sum_{i=1}^n w_i (y_i - \mu_i) \frac{\partial \eta_i}{\partial \mu_i} x_{ij}, \quad \text{para } j = 1, \dots, p,$$

tem-se que cada componente r_j do vetor r , de dimensão p , assume a forma

$$r_j = \sum_{k=1}^n w_k x_{kj} \left[(y_k - \mu_k) \frac{\partial \eta_k}{\partial \mu_k} + \eta_k \right] \quad \text{para } j = 1, \dots, p$$

Note-se que (1.22) escrita na forma matricial é equivalente a:

$$X^T W X \beta^* = X^T W z$$

onde $z_k = \left[(y_k - \mu_k) \frac{\partial \eta_k}{\partial \mu_k} + \eta_k \right]$, para $k = 1, \dots, n$, W é a matriz $\text{diag}\{w_1, w_2, \dots, w_n\}$ e X é a matriz do modelo. Por este resultado, (1.22) pode ser interpretado como um sistema de mínimos quadrados ponderados com função peso W , com elementos na diagonal $w_i = \left(\frac{\partial \mu_i}{\partial \eta_i}\right)^2 / V_i$, para $i = 1, 2, \dots, n$ e elementos fora da diagonal nulos, ou seja, $w_{ij} = 0$ para $i, j = 1, 2, \dots, n$ e $i \neq j$, e uma variável dependente modificada z com $z = \eta + (y - \mu) / \left(\frac{\partial \mu}{\partial \eta}\right)^{11}$. Portanto, a solução das equações de máxima verossimilhança para um modelo linear generalizado equivale a calcular, repetidamente, uma regressão linear ponderada de z sobre X , com uma função peso W que se modifica durante o processo iterativo.

¹¹Ver apêndice A4

1.3.2 Considerações Específicas sobre os Estimadores de Máxima Verossimilhança

No procedimento apresentado na seção anterior surgem situações particulares que devem ser levadas em consideração ao estimar os parâmetros dos modelos lineares generalizados quando se utiliza o método de máxima verossimilhança. Como considerações finais sobre o procedimento de estimação de máxima verossimilhança se apresentam a seguir algumas destas situações.

1. Nos modelos lineares generalizados com funções de ligação canônicas, as derivadas de segunda ordem da função de log-verossimilhança e os seus valores esperados coincidem, portanto a técnica de score de Fisher e o método de Newton-Raphson se reduzem ao mesmo algoritmo. Isto se deve ao fato de $w_i(\frac{\partial \mu_i}{\partial \eta_i})$ ser constante e por isso o segundo termo da expressão à direita da equação (1.21) fica exatamente igual a zero.
2. A unicidade dos estimadores de máxima verossimilhança para os modelos com função de ligação canônica, que foi afirmada anteriormente, pode ser verificada a partir da equação (1.11). Para este tipo de modelo $V = w = \frac{\partial \mu}{\partial \eta}$, e assim (1.11) se reduz a: $\sum_{i=1}^n \alpha(\phi_i) y_i x_{ij} = \sum_{i=1}^n \alpha(\phi_i) \hat{\mu}_i x_{ij}$.
3. Como cada observação y_i representa uma estimativa da média μ_i , o meio natural de inicializar o processo iterativo é tomar como primeira aproximação de μ_i o valor de y_i .
4. Deve-se verificar que a função de ligação esteja bem definida para todos os possíveis valores da variável resposta y a serem observados. Se isto não ocorre, a solução será substituir a variável resposta y por $y + c$, escolhendo c de maneira tal que o valor esperado da função de ligação avaliada em $y + c$ seja tão próximo quanto possível da função de ligação avaliada em μ .

Expandindo $f(y + c)$ em séries de Taylor, e negligenciando os termos de ordem c^2 , temos que c deve ser igual à seguinte expressão (Ver apêndice A5):

$$\frac{-V f''(\mu)}{2\alpha(\phi) f'(\mu)}$$

No caso do modelo log- Poisson, que será apresentado posteriormente, o valor da constante c é $\frac{1}{2}$ (Ver apêndice A3).

5. Se X não é de posto completo¹², não existe solução única, mas as funções $\hat{\eta}$ e $\hat{\mu}$ são estimáveis¹³. Se X é de posto completo, o sistema pode divergir; se convergir, é necessário verificar se converge produzindo um máximo global.

Caso $\ell(\beta)$ seja estritamente côncava como função de β , se existe um máximo, este é único.

Também é preciso verificar se o valor $\hat{\beta}$ que maximiza a função de verossimilhança é finito e pertence ao interior do espaço de parâmetros especificado.

As condições necessárias para a existência e unicidade de máximo no interior do espaço paramétricos e a natureza finita dos estimadores dos parâmetros dos modelos lineares generalizados são discutidas por Wedderburn (1976) que apresenta uma tabela para diversas funções de distribuição e de ligação. Para a distribuição Poisson com função de ligação $\eta = \log(\mu)$, está garantida a existência de um máximo finito e a unicidade deste.

1.4 Avaliação do Ajuste do Modelo

Uma vez estimados os parâmetros do modelo é importante obter uma medida da adequabilidade do modelo para descrever o conjunto de observações. Vários métodos são utilizados para medir a qualidade do ajuste (muitas vezes chamada *bondade de ajuste*), baseados tanto na construção de estatísticas que medem a diferença entre o ajuste dado por um modelo que descreva perfeitamente os dados e o modelo que se está estudando, ou estatísticas da discrepância entre os valores observados e os valores ajustados pelo modelo, entre outros. Os métodos que envolvem a análise dos

¹²O *posto* de uma matriz está dado pelo número de linhas (ou colunas) linearmente independentes. Uma matriz é chamada de *posto completo* se todas suas linhas (ou colunas) são linearmente independentes.

¹³Neste caso, diz-se que uma função de β é *estimável* se é invariante quanto à escolha de $\hat{\beta}$, solução do método de estimação de máxima verossimilhança.

resíduos são muito úteis para obter uma medida do ajuste do modelo e verificar se as características necessárias para a implementação do modelo (como a distribuição da variável resposta, o tipo de função de ligação e as covariáveis utilizadas) se adequam ao conjunto de dados.

Nesta seção se apresentam alguns dos métodos mais comumente utilizados para medir a bondade de ajuste de um modelo linear generalizado.

1.4.1 Função Desvio

Uma forma de medir o bondade de ajuste de um modelo é comparar o valor da log-verossimilhança do modelo em estudo com o máximo atingível da log-verossimilhança. O valor máximo que pode ser atingido pela função de log-verossimilhança corresponde a um modelo *completo* ou *saturado*. O modelo completo para os modelos lineares generalizados está definido da seguinte forma: (Dobson, 1990)

1. O modelo completo é um modelo linear generalizado com a mesma distribuição para a variável resposta que o modelo sob estudo.
2. O modelo completo tem a mesma função de ligação entre o valor esperado da variável resposta e o preditor linear que o modelo sob estudo.
3. O número de parâmetros no modelo completo é igual a n , número total de observações.

A função de log-verossimilhança do modelo completo será denotada $\ell(\hat{\beta}_{m\acute{a}x}, y)$ enquanto a função de log-verossimilhança para o modelo em estudo será denotada por $\ell(\hat{\beta}, y)$.

Nelder e Wedderburn (1972) propõem utilizar o dobro da diferença entre a log-verossimilhança do modelo completo e a log-verossimilhança do modelo sob estudo, definido a *função desvio*¹⁴, como:

$$D = 2 \left(\ell(\hat{\beta}_{m\acute{a}x}, y) - \ell(\hat{\beta}, y) \right) / \alpha(\phi) \quad (1.23)$$

¹⁴*Deviance*, em inglês.

A função desvio padronizada está definida como:

$$D^* = 2 \left(\ell(\hat{\beta}_{m\hat{\alpha}x}, y) - \ell(\hat{\beta}, y) \right) \quad (1.24)$$

Explicitamente, utilizando a expressão para a função de log-verossimilhança dada em (1.3) tem-se:

$$D^* = 2 \left[\sum_{i=1}^n \{ \alpha(\phi_i) [y_i \tilde{\theta}_i - g(\tilde{\theta}_i)] + c(y_i, \phi_i) \} - \sum_{i=1}^n \{ \alpha(\phi_i) [y_i \hat{\theta}_i - g(\hat{\theta}_i)] + c(y_i, \phi_i) \} \right],$$

onde $\tilde{\theta} = \theta(\hat{\beta}_{m\hat{\alpha}x})$ e $\hat{\theta} = \theta(\hat{\beta})$.

Simplificando a equação anterior, tem-se que a função desvio padronizada para os modelos lineares generalizados está dada por

$$D^* = 2 \sum_{i=1}^n \alpha(\phi_i) [y_i (\tilde{\theta}_i - \hat{\theta}_i) + g(\tilde{\theta}_i) - g(\hat{\theta}_i)]. \quad (1.25)$$

Cada componente da função desvio mede a diferença entre os logaritmos da função de verossimilhança observada e a função de verossimilhança ajustada, para a observação correspondente. A soma destas componentes mede a discrepância total entre as duas funções de log-verossimilhanças (observada e ajustada).

A função desvio é sempre maior ou igual a zero e, cada vez que uma nova covariável é agregada à componente sistemática, a função desvio decresce até se tornar zero para o modelo completo.

A cada modelo é associada uma quantidade p , chamada *graus de liberdade*, que corresponde ao posto da matriz X , ou equivalentemente, ao número de parâmetros linearmente independentes a serem estimados. Para uma amostra de n observações independentes, a função desvio do modelo fica com $n - p$ graus de liberdade.

A função desvio padronizada definida em (1.24) pode ser reescrita como:

$$D = 2 \left[\left(\ell(\hat{\beta}_{m\hat{\alpha}x}, y) - \ell(\beta_{m\hat{\alpha}x}, y) \right) - \left(\ell(\hat{\beta}, y) - \ell(\beta, y) \right) + \left(\ell(\beta_{m\hat{\alpha}x}, y) - \ell(\beta, y) \right) \right] \quad (1.26)$$

Os vetores de parâmetros para o modelo completo (com n parâmetros), e para o modelo estudado (com p parâmetros), estão representados por $\beta_{m\hat{\alpha}x}$ e β , respectivamente, e os seus estimadores por $\hat{\beta}_{m\hat{\alpha}x}$ e $\hat{\beta}$.

O primeiro e segundo termo dentro do colchete tem distribuição χ_n^2 e χ_p^2 , respectivamente (ver apêndice A4). O terceiro termo é positivo e se aproxima de zero quando a descrição do conjunto de dados fornecida pelo modelo com p parâmetros é quase tão exata como aquela fornecida pelo modelo completo.

Portanto, se o modelo em estudo fornece um "bom" ajuste, ou seja, representa os dados quase como o modelo completo, a função desvio terá uma distribuição χ_{n-p}^2 . Se, por outro lado, o modelo estudado não fornece um bom ajuste, é "ruim", o último termo de (1.26) será maior que zero e a função desvio terá um valor maior que aquele projetado pela distribuição χ_{n-p}^2 .

Já que $E(\chi_{n-p}^2) = n - p$, se considera que o modelo tem um bom ajuste se D se aproxima a $(n - p)$, isto é, se a função desvio é aproximadamente igual aos graus de liberdade.

Christensen (1990) observa que se $\alpha(\phi_i)$ não fosse conhecido, a função desvio padronizada D^* , não seria uma estatística, pois o seu valor dependeria de parâmetros desconhecidos. Neste caso, pode-se trabalhar com a estatística $D^* = D/\alpha(\phi_i)$, que independe de ϕ .

1.4.2 Estatística de Pearson Generalizada

Uma outra medida utilizada para avaliar o ajuste do modelo é a *estatística de Pearson generalizada* definida por:

$$X^2 = \sum_{i=1}^n \frac{\alpha(\phi_i)(y_i - \hat{\mu}_i)^2}{\hat{V}_i}$$

onde \hat{V}_i é a função de variância estimada, definida em forma análoga à definição dada na seção 2.2, como $\hat{V}_i = g''(\hat{\theta}_i)$.

Esta estatística tem uma distribuição assintótica χ_{n-p}^2 .

Nas famílias exponenciais de um parâmetro, $\alpha(\phi_i)$ é constante para todo valor de $i = 1, 2, \dots, n$, e sem perder generalidade pode-se assumir $\alpha(\phi_i) = 1$ para $i = 1, 2, \dots, n$. Neste caso, Christensen (1990) mostra que se o modelo fosse correto,

$$\frac{X^2}{n-p} \cong 1.$$

Portanto, se $X^2/(n-p)$ é substancialmente maior que 1, tem-se uma indicação da falta de ajuste do modelo.

1.4.3 Resíduos

As estatísticas de bondade de ajuste fornecem uma medida da adequabilidade do modelo. Porém, a investigação de aspectos específicos do modelo é facilitada pelo uso dos resíduos. Através do estudo exploratório dos resíduos pode-se obter informação da adequabilidade do modelo a respeito da validade das hipóteses do mesmo, como por exemplo, a escolha da função de variância, da função de ligação e dos termos no preditor linear. Também pode-se detectar a presença de valores atípicos. Os resíduos expressam a discrepância entre uma observação e seu valor ajustado pelo modelo.

Apresenta-se a seguir os tipos de resíduos mais utilizados na teoria dos modelos lineares generalizados.

1. Resíduo de Pearson

A estatística de Pearson generalizada mencionada anteriormente pode ser decomposta como

$$X^2 = \sum_{i=1}^n r_{p_i}^2$$

onde cada componente r_{p_i} é o *resíduo de Pearson*, e está dada por

$$r_{p_i} = \frac{\sqrt{\alpha(\phi_i)} \cdot (y_i - \hat{\mu}_i)}{\sqrt{\hat{V}_i}} \quad (1.27)$$

Uma restrição para este tipo de resíduo é que geralmente a sua distribuição é bastante assimétrica para os modelos não-normais.

2. Resíduo de Ascombe

Ascombe propôs utilizar a função $N(y)$, no lugar de y , onde $N(\cdot)$ é escolhida de tal maneira que a distribuição de $N(y)$ seja o mais próxima possível da distribuição normal. Para as funções de verossimilhança que ocorrem nos modelos lineares generalizados, a função $N(\cdot)$ é dada por:

$$N(\mu) = \int \frac{\partial \mu}{V^{\frac{1}{2}}(\mu)} \quad (1.28)$$

Como $N'(\mu) \left(\frac{V}{\phi}\right)^{\frac{1}{2}}$ é a aproximação de primeira ordem para o desvio padrão de $N(y)$, o resíduo de Ascombe, visando a normalização da distribuição e estabilização da variância, é expresso por:

$$r_{A_i} = \frac{\sqrt{\alpha(\phi_i)} [N(y_i) - N(\hat{\mu}_i)]}{N'(\hat{\mu}_i) \sqrt{\hat{V}_i}} \quad (1.29)$$

3. Resíduo definido como componente da função desvio

Se a função desvio é usada como uma medida de discrepância do ajuste de um modelo linear generalizado, então cada observação contribui com uma quantidade

d_i , de tal forma que $D = \sum_{i=1}^n d_i$, i.e.,

$$d_i = 2 \left[\alpha(\phi_i) \{y_i(\tilde{\theta}_i - \hat{\theta}_i) + g(\tilde{\theta}_i) - g(\hat{\theta}_i)\} \right] \quad (1.30)$$

com $\tilde{\theta}_i$ e $\hat{\theta}_i$ definidos em (1.25).

O valor d_i representa a distância da observação y_i ao seu valor ajustado $\hat{\mu}_i$ medida na escala da log-verossimilhança.

Define-se então este tipo de resíduo como:

$$r_{D_i} = \text{sinal}(y_i - \hat{\mu}_i) \sqrt{d_i} \quad (1.31)$$

onde a função sinal está dada por

$$\text{sinal}(x) = \begin{cases} 1 & \text{se } x \geq 0 \\ 0 & \text{se } x = 0 \\ -1 & \text{se } x < 0 \end{cases}$$

Em geral, os gráficos dos resíduos versus algumas variáveis de interesse, como os valores ajustados, a variável resposta, as covariáveis, ou o gráfico dos resíduos padronizados num papel de probabilidade normal são de vital importância para detectar se as condições supostas pelo modelo não são cumpridas.

1.5 Modelos Log-Lineares

Os modelos log-lineares são casos particulares dos modelos lineares generalizados, onde a função de ligação é do tipo logarítmico. Nestes modelos a dependência de $\mu_i = E(Y_i)$ no vetor de covariáveis x_i se assume como função multiplicativa e geralmente se escreve como:

$$\log(\mu_i) = \eta_i = \sum_{j=1}^p x_{ij} \beta_j \quad i = 1, 2, \dots, n \quad (1.32)$$

Geralmente, além desta relação logarítmica entre o preditor linear e a variável resposta, também se assume que:

$$\text{Var}(Y_i) = \sigma^2 E(Y_i) \quad i = 1, 2, \dots, n \quad (1.33)$$

onde σ^2 , o parâmetro de dispersão, é constante sobre todas as observações.

Wedderburn (1974) explica como os parâmetros dos modelos lineares generalizados podem ser estimados utilizando a função de *quasi-verossimilhança*, que não precisa do conhecimento da distribuição das observações mas somente da relação que existe entre o valor esperado e a variância das observações. Ele mostra também que para as distribuições pertencentes à família exponencial de um parâmetro, as funções de verossimilhança e quasi-verossimilhança coincidem.

1.5.1 Modelo Log-Poisson

Dentro dos modelos log-lineares se encontra o modelo log-Poisson, onde a variável aleatória Y tem distribuição Poisson com parâmetro μ ,

$$Y_i \sim \text{Poisson}(\mu_i) \quad i = 1, 2, \dots, n$$

ou seja,

$$P(Y_i = y_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}, \quad \text{para } y_i \text{ inteiro não negativo} \quad (1.34)$$

Caraterização do Modelo Log-Poisson

Mostrar-se-á a seguir como o modelo log-Poisson é incluído na classe dos modelos lineares generalizados.

Transformando de forma conveniente a equação (1.34) tem-se que a função de probabilidade de Y_i é dada por

$$P(Y_i = y_i) = \exp[-\mu_i + y_i \ln \mu_i - \ln(y_i!)]$$

Identificando:

$$\alpha(\phi_i) = 1$$

$$\theta_i = \ln \mu$$

$$g(\theta_i) = \mu$$

$$c(y_i, \phi_i) = \ln(y_i!)$$

para $i = 1, 2, \dots, n$,

tem-se que a distribuição de Y_i pertence à família de distribuições dada em (1.1).

Note-se que, na realidade, a distribuição de Y depende só do parâmetro θ . O parâmetro de escala ϕ não interfere, pois é constante e igual a 1.

A função de ligação entre o valor esperado da variável resposta Y_i , μ_i , e o preditor linear η_i , é a função logaritmo natural::

$$\eta_i = \ln(\mu_i) \quad i = 1, 2, \dots, n,$$

onde

$$\eta_i = \sum_{j=1}^p x_{ij}\beta_j \quad i = 1, 2, \dots, n,$$

com X matriz do modelo.

A função de log-verossimilhança para o modelo log-Poisson está dada por

$$\ell(\theta, \phi; \mathbf{y}) = \sum_{i=1}^n [y_i \ln(\mu_i) - \mu_i + \ln(y_i!)] .$$

Neste caso, como $E(Y_i) = Var(Y_i) = \mu_i$, tem-se que $\sigma^2 = 1$, para satisfazer a condição (1.33).

Função desvio

Da definição de função desvio dada em (1.23) tem-se que:

$$\begin{aligned} D &= 2 \left(\ell(\hat{\beta}_{m\acute{a}x}; \mathbf{y}) - \ell(\hat{\beta}; \mathbf{y}) \right) \\ &= 2 \sum_{i=1}^n \left[y_i \ln \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) \right] \\ &= 2 \sum_{i=1}^n y_i \ln \left(\frac{y_i}{\hat{\mu}_i} \right) - 2 \sum_{i=1}^n (y_i - \hat{\mu}_i) . \end{aligned}$$

No modelo log-Poisson a segunda somatória é zero se $\mathbf{1} \in X$, i.e., se a matriz X das covariáveis tem uma coluna onde todos os seus elementos são iguais à unidade.¹⁵

Neste caso, o desvio é igual à razão de máxima verossimilhança

$$S_p = 2 \sum_{i=1}^n y_i \ln \left(\frac{y_i}{\mu_i} \right) .$$

¹⁵Ver, por exemplo, Nelder e Wedderburn (1972)-Apêndice.

Estatística de Pearson Generalizada

A estatística de Pearson Generalizada para o modelo log-Poisson se reduz a

$$X^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}$$

pois $\alpha(\phi_i) = 1$ e $\hat{V}_i = \hat{\mu}_i$.

Resíduos

No modelo log-Poisson os resíduos considerados anteriormente assumem as seguintes formas particulares:

i) Resíduo de Pearson:

$$r_{P_i} = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{\mu}_i}}$$

ii) Resíduo de Ascombe:

Para a distribuição de Poisson tem-se

$$A(\mu) = \int \frac{\partial \mu}{\mu^{\frac{1}{3}}} = \frac{3}{2} \mu^{\frac{2}{3}}$$

e

$$A'(\mu) \sqrt{V(\mu)} = \frac{\partial}{\partial \mu} \left(\frac{3}{2} \mu^{\frac{2}{3}} \right) \mu^{\frac{1}{2}} = \mu^{\frac{1}{6}}$$

Então, da definição dada em (1.29), temos que para o modelo log-Poisson o resíduo de Ascombe toma a forma:

$$r_{A_i} = \frac{\frac{3}{2} (y_i^{\frac{2}{3}} - \hat{\mu}_i^{\frac{2}{3}})}{\hat{\mu}_i^{\frac{1}{6}}}$$

iii) Resíduo da Componente do Desvio:

$$r_{D_i} = \text{sign}(y_i - \hat{\mu}_i) \sqrt{2 \left[y_i \ln \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) \right]}$$

1.6 Comentários

Neste capítulo apresentou-se os modelos lineares generalizados de uma maneira geral, definindo as suas características, métodos de estimação dos parâmetros e algumas medidas do ajuste do modelo estimado. A classe dos modelos lineares generalizados abrange um grande número de tipos de variáveis resposta. Além disso, o número de combinações possíveis entre o tipo de função de distribuição da variável resposta e a função de ligação existente entre a esperança da variável e o conjunto de covariáveis é muito grande. Na literatura estatística existe uma grande quantidade de trabalhos que abordam os modelos lineares generalizados, mas sob um ponto de vista particular discutindo casos específicos, como por exemplo, análise de regressão, tabelas de contingência, análise de sobrevivência, etc.

Na síntese dos modelos lineares generalizados apresentada neste capítulo foi dada maior atenção ao caso do modelo log-Poisson. É através deste modelo que vai ser analisada a fecundidade marital no Peru, entre os anos 1977 e 1992.

No capítulo seguinte será apresentada uma síntese dos modelos que tradicionalmente têm sido utilizados para analisar a fecundidade em populações humanas, para posteriormente introduzir o modelo log-Poisson nesta análise.

Capítulo 2

PADRÕES DE FECUNDIDADE MARITAL

A expressão *fecundidade* é usada para indicar o desempenho reprodutivo efetivo de uma mulher ou de um grupo de mulheres que já completaram o período reprodutivo. Por período reprodutivo entende-se aquele que vai desde a idade da menarca até a idade da menopausa. Muito embora estas idades, tanto a do início como a do final do período, variem de mulher para mulher, convencionou-se considerar como período reprodutivo aquele compreendido entre os 15 e 49 anos (Berquó, 1980).

A fecundidade, como processo responsável pela reprodução biológica da população humana, constitui um aspecto essencial na determinação do crescimento e do perfil demográfico de uma população. Nas últimas décadas, devido à rápida expansão da população mundial, a fecundidade e os seus determinantes têm-se convertido em temas urgentes de investigação. As tentativas para regular o crescimento da população têm-se concentrado basicamente na redução dos níveis de fecundidade (Mauldin, 1984).

Isto motivou o interesse de avançar nesta linha de pesquisa, considerando de forma particular a evolução dos níveis de fecundidade de mulheres casadas, ou mais genericamente, de mulheres em união, também denominada *fecundidade marital*.

Esta análise está baseada nos modelos de fecundidade clássicos, que serão apresentados e discutidos em detalhe ao longo deste capítulo. Inicialmente, a fecun-

didade marital tem sido estudada através dos modelos propostos por Coale (1971), Coale e Trussell (1974) e posteriormente por Page (1977), que incorpora aos modelos anteriores baseados na idade da mulher uma nova dimensão, a duração das uniões.

Posteriormente, será encontrada uma forma de estimação do modelo de interesse, que será utilizada para a análise específica da fecundidade marital no Peru no período 1977-1992.

2.1 Modelo de Nupcialidade de Coale

A origem dos modelos da análise da fecundidade se encontra no trabalho de Coale (1971) dedicado inicialmente a estudar os padrões de nupcialidade. No final deste trabalho Coale lança as primeiras idéias sobre os modelos de fecundidade marital.

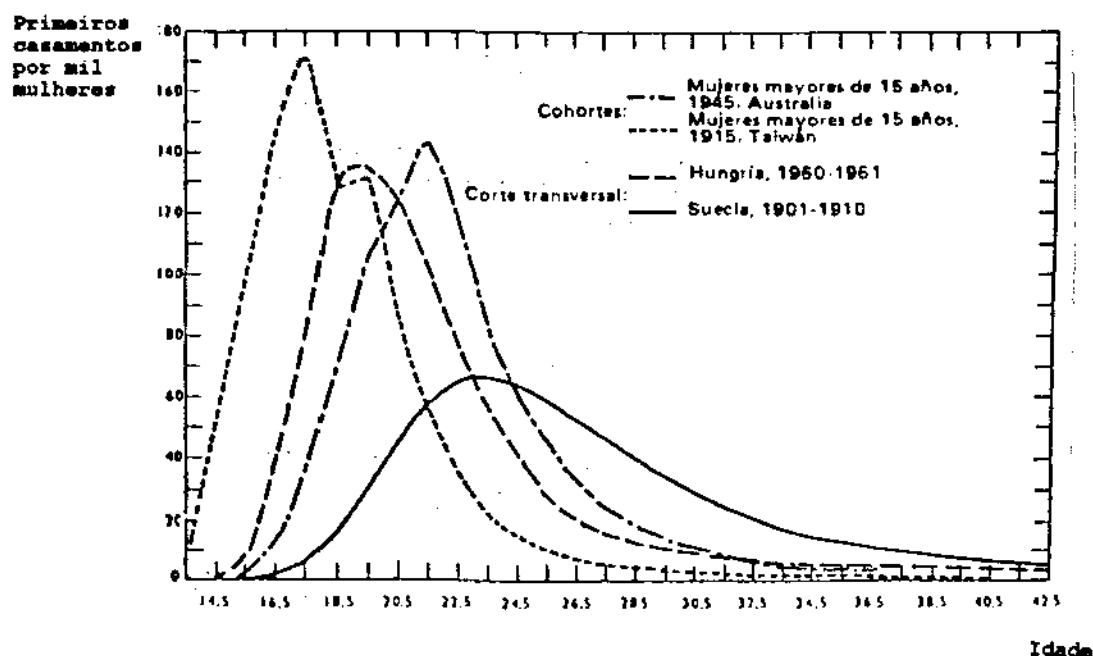
O primeiro modelo de nupcialidade por idade foi proposto por Coale (1971). O autor, juntamente com Ivan Lakos e Etienne van der Walle, ao investigar o declínio da fecundidade marital na Europa, identificaram a existência de um padrão de nupcialidade por idade comum às populações com características sociais e demográficas muito diversas.

Na Figura 2.1, pode-se observar a distribuição da freqüência dos primeiros casamentos¹ para quatro conjuntos de dados muito diferentes entre si. A coorte² de mulheres que tinham 15 anos em 1915 em Taiwan, cuja proporção de mulheres solteiras maiores de 50 anos é 1%; a coorte Australiana que tinha 15 anos em 1944, cujo padrão de fecundidade foi afetado pela situação pós-guerra e a imigração; a população da Suécia entre 1901-1910, com um padrão de nupcialidade tardia e uma proporção de mulheres solteiras maiores de 50 anos de 25%; e a população de Hungria

¹Neste caso, freqüência de casamentos é definida por Coale(1971) como o número de primeiros casamentos num intervalo de idade dividido pelo número de mulheres nesse intervalo, independentemente de seu estado conjugal.

²Uma coorte é um conjunto de indivíduos (ou estatísticas de eventos ocorridos nesse grupo) que têm um fator temporal comum. Exemplos: o grupo de mulheres brasileiras nascidas em 1940 (coorte de nascimento), ou o grupo de mulheres peruanas casadas no ano 1962 (coorte de matrimônio).

Figura 2.1: Freqüência de primeiros casamentos por idade, para quatro populações selecionadas.



Fonte: Coale (1971), Figura 6, p.198.

em 1960-1961 que apresenta um padrão de nupcialidade precoce. As curvas apresentadas nesta figura mostram as diferenças mencionadas no que se refere aos padrões de nupcialidade.

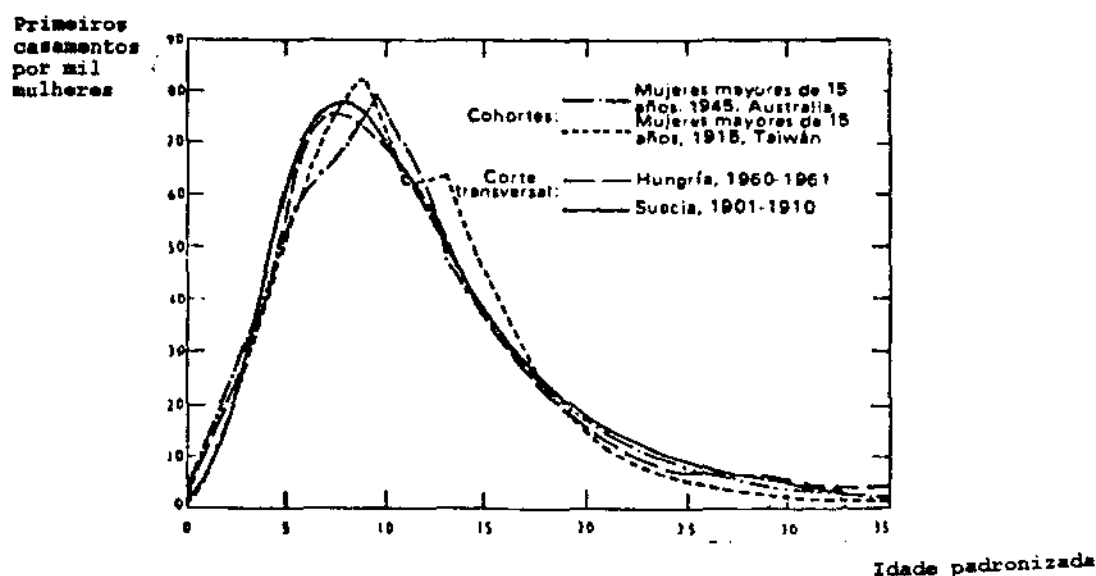
Na Figura 2.2 ilustram-se as distribuições de frequência de primeiros casamentos apresentadas anteriormente, após a padronização proposta por Coale (1971) e que será discutida com detalhe a seguir.

2.1.1 Função Densidade Padrão de Primeiros Casamentos

Escolhendo um ponto de início comum, e modificando a escala horizontal e vertical, as distribuições apresentadas na Figura 2.1, estas distribuições assumem uma forma típica, como foi observado na Figura 2.2.

Baseando-se neste fato, Coale propõe construir uma curva que represente, da melhor maneira possível, a forma típica dos padrões de nupcialidade por idade

Figura 2.2: Frequência de primeiros casamentos em populações selecionadas, com escala e origem ajustados.



Fonte: Coale (1971), Figura 5, p.197.

das diferentes populações, que possa ser utilizada como base para vários tipos de estimações e cálculos.

O modelo de nupcialidade por idade desenvolvido por Coale está baseado nas frequências de casamentos registrados na Suécia entre 1865 e 1869. As demais populações terão uma distribuição de frequências de casamentos³ por idade similar a esta, diferindo somente na origem, área total sob a curva e na escala horizontal.

Considere-se a distribuição empírica de casamentos por idade, $F(x)$, ou seja, a proporção de mulheres de idade x alguma vez unidas, excluindo aquelas que nunca se casarão⁴, segundo os dados da Suécia selecionados por Coale. Com o objetivo de achar uma função de distribuição de probabilidade, a distribuição empírica $F(x)$ foi plotada num papel logarítmico, mas não foi encontrada nenhuma relação linear. No

³Entende-se por casamento o casamento civil e/ou religioso, ou a união consensual, dependendo da análise a ser realizada.

⁴Uma coorte ou população de mulheres pode-se dividir em dois grupos: aquelas que vão se unir (se sobreviver até uma idade avançada) e o grupo de mulheres que nunca vão se unir. O primeiro grupo é chamado de mulheres "elegíveis".

entanto, se ao invés de trabalhar com $F(x)$ diretamente, utilizar a função risco⁵ $r(x)$ definida por

$$r(x) = \frac{F'(x)}{1 - F(x)}, \quad (2.1)$$

encontra-se que esta função é muito bem ajustada por uma função dupla-exponencial dada por

$$r(x) = 0.174e^{-4.411e^{-0.309x}}. \quad (2.2)$$

Esta última expressão é escolhida por Coale para representar a função risco padrão de casamentos.

O ajuste da função dada em (2.2) aos valores derivados dos dados da Suécia (1865-1869) pode ser observado na Figura 2.3. Infelizmente, quando (2.2) é substituída em (2.1), não se encontra nenhuma forma algébrica para $F(x)$.

Coale e McNeil (1972) utilizam um método alternativo para solucionar este problema, assumindo que $F(x)$ é a convolução⁶ de um número (possivelmente infinito) de componentes simples. Neste caso, como a função risco tem uma assíntota constante (r_0), a convolução contém pelo menos uma componente com distribuição exponencial, com parâmetro r_0 . Uma vez removida esta componente, obtém-se como função de distribuição resultante $F_1(x)$ (Ver apêndice B1) :

$$F_1(x) = F(x) + \frac{F'(x)}{r_0} \quad (2.3)$$

$F_1(x)$ pode ser obtida diferenciando numericamente a distribuição empírica $F(x)$ e fazendo $r = 0.174$ (valor que toma $r_0(x)$ quando $x \rightarrow \infty$).

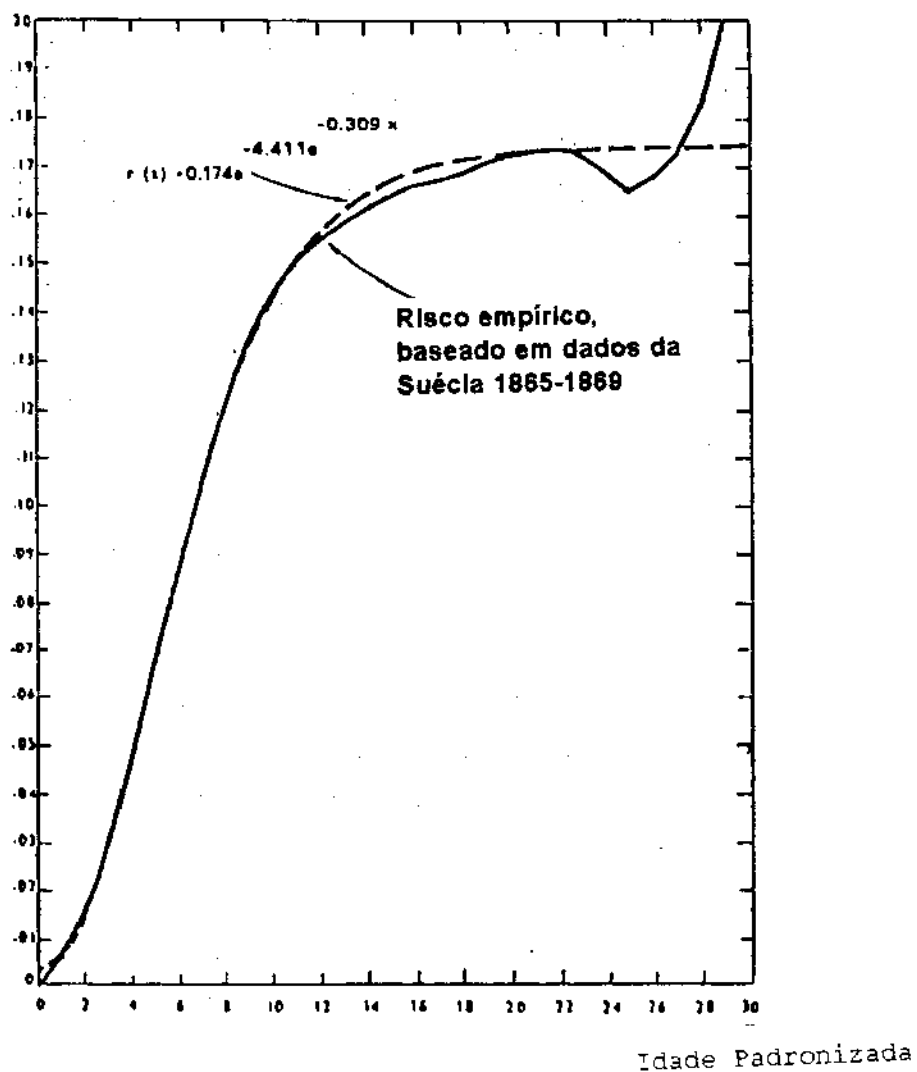
A função risco $r_1(x)$, correspondente a $F_1(x)$, tem uma assíntota constante ($r_1 = 0.483$) o que evidencia que $F_1(x)$ é a convolução de alguma função de distribuição $F_2(x)$ e uma distribuição exponencial com parâmetro r_1 . Ao ser removida a distribuição exponencial, utilizando de uma forma análoga a equação (2.3), pode-se obter a distribuição resultante $F_2(x)$.

⁵A função risco $r(x)$ representa a probabilidade de uma mulher solteira passar a condição de casada na idade x . Quando $F(x)$ é a distribuição empírica, $r(x)$ é chamado "risco empírico".

⁶A convolução de duas funções de distribuição $F_X(x)$ e $G_Y(y)$ está dada pela função de distribuição $H_Z(z) = \int_{-\infty}^z F_X(z) \delta G_Y(z-t)$.

Figura 2.3: Risco empírico, baseado na nupcialidade da Suécia (1865-1869), e ajustado por uma função dupla-exponencial.

Risco anual de pasar
condição de solteira
a casada



Repetindo sucessivamente este procedimento obtém-se a função de distribuição resultante $F_n(x)$, quando a n -ésima componente exponencial for retirada:

$$F_n(x) = F_{n-1}(x) + \frac{F'_{n-1}(x)}{r_{n-1}} \quad (2.4)$$

onde $r_{i-1} = 0.174 + (i-1)0.309$ para $i = 1, 2, \dots, n$

Note-se que quando a n -ésima componente é removida a média da distribuição resultante fica diminuída em $\frac{1}{r_{n-1}}$, e variância é diminuída em $\frac{1}{r_{n-1}^2}$. Como $\sum \frac{1}{r_n}$ é divergente (porém $\sum \frac{1}{r_n^2}$ converge), este procedimento não pode ser repetido indefinidamente sem reduzir a média da distribuição resultante a $-\infty$.

Este problema é facilmente solucionado adicionando um termo constante apropriado ao argumento de $F_n(x)$, cada vez que uma componente exponencial é removida. Desta maneira, a média da distribuição resultante mantém-se constante enquanto a sua variância é reduzida.

A seguir mostrar-se-á que a função de distribuição correspondente à função risco dada em (2.2) pode ser representada por uma convolução infinita de distribuições exponenciais (com médias corrigidas) mais um termo constante.

Considere-se a variável aleatória

$$X_n = a + \sum_{j=1}^n \left(Z_j - \frac{1}{\alpha + (j-1)\lambda} \right) \quad (2.5)$$

onde Z_j são variáveis aleatórias independentes com distribuição exponencial com média $\frac{1}{\alpha + (j-1)\lambda}$, para $j = 1, \dots, n$.

Coale e McNeil (1972) demonstram que a função de densidade de X_n é:

$$g(x) = \frac{\lambda}{\Gamma(\frac{\alpha}{\lambda})} \exp(-\alpha(x - \mu) - \exp(-\lambda(x - \mu))) \quad (2.6)$$

onde

$$\mu = a + \frac{\Psi(\frac{\alpha}{\lambda})}{\lambda}$$

Γ : função gamma

Ψ : função digamma = Γ'/Γ

a : valor esperado de X_n

Quando $\alpha = 0.174$, $\lambda = 0.288$, $a = 11.36^7$, a função de distribuição de freqüências dada em (2.6) e aquela derivada por (2.2) diferem por menos de 0.001, em toda a faixa etária. Assim, estas duas distribuições de freqüências de casamentos por idade, obtidas de maneiras diferentes (uma ajustando a função risco empírica e a outra sob o suposto dos componentes exponenciais para a função de distribuição empírica) ajustam os dados quase com a mesma precisão. Este fato permite aproximar a distribuição padrão de freqüências de casamentos por $g(x)$, dada em (2.6), com os valores de α , λ e a especificados acima. Tem-se então que a função de densidade padrão dos casamentos é:

$$g_s(x) = 0.1946 \exp(-0.174(x - 6.06) - \exp(-0.288(x - 6.06))) \quad (2.7)$$

O ajuste de $g(x)$ para os dados de freqüências de casamentos da Suécia (1865-1869) é mostrado na Figura 2.4.

Note-se que quando $n \rightarrow \infty$ (2.5) pode-se escrever como:

$$X_\infty = a + \sum_{j=1}^{\infty} \left(Z_j - \frac{1}{\alpha + (j-1)\lambda} \right) = \sum_{j=1}^m Z_j + Y_m$$

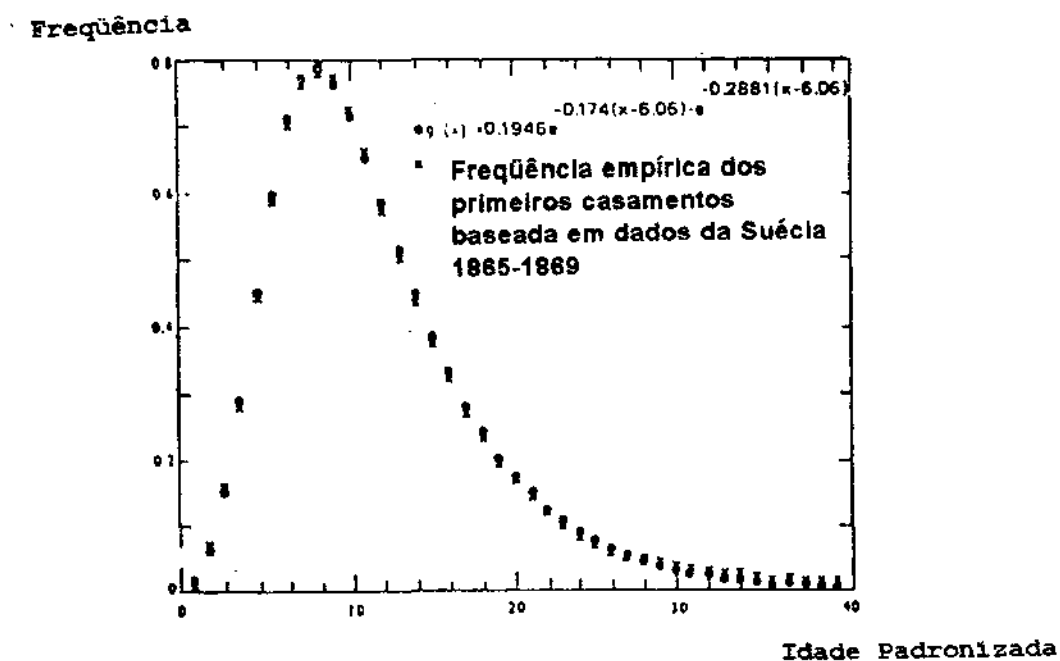
com

$$Y_m = a - \sum_{j=1}^n \frac{1}{\alpha + (j-1)\lambda} + \sum_{j=m+1}^{\infty} \left(Z_j - \frac{1}{\alpha + (j-1)\lambda} \right)$$

Coale & McNeil (1972) observam que a variável aleatória Y_m é aproximadamente normal para $m \geq 3$, portanto, $g_s(x)$ é aproximadamente a função de densidade da convolução de uma distribuição normal e três distribuições exponenciais, que são interpretadas da seguinte maneira:

⁷Estes valores foram obtidos da seguinte maneira: 11.36 é a média da distribuição "empírica" de nupcialidade; 0.174 é o valor assintótico da função risco calculada empiricamente; 0.288 é o valor de X que assegura a concordância de $g(x)$ com a curva padrão na vizinhança da moda.

Figura 2.4: Comparação das freqüências dos primeiros casamentos de Suécia (1865-1869) e a função de densidade padrão $g(x)$.



Fonte: Coale e McNeil (1972), Figura A, p.745.

-A variável aleatória com distribuição normal representa a idade de entrada no mercado matrimonial.

-As variáveis aleatórias com distribuição exponencial representam três fases iniciadas na idade em que a mulher está apta para a união, até a união propriamente dita. Estas fases estão relacionadas com o tempo até encontrar um parceiro apropriado, ficar noiva e casar.

2.1.2 Padronização das Distribuições de Frequências de Primeiros Casamentos

A partir da função de densidade padrão de primeiros casamentos, pode-se expressar a experiência dos casamentos de uma outra coorte (ou população), mediante três parâmetros:

i) **Origem da curva** (a_0) que representa a menor idade na qual um número significativo de uniões ocorrem. Depende das características socio-econômico-culturais da população em estudo.

iii) **Escala horizontal** (k) que representa a velocidade com que ocorrem os casamentos, uma vez que o processo se inicia, relativa à distribuição padrão.

ii) **Escala vertical** (C) que corresponde à proporção de mulheres casadas no final do processo. Assume-se que a partir de uma certa idade o número de casamentos é aproximadamente zero e, as mulheres que ficam solteiras até essa idade já não se casarão mais.

Suponha-se que se está analisando uma coorte de mulheres sujeita a um padrão de nupcialidade da forma padrão, com origem a_0 , proporção de mulheres casadas no final do processo C , e uma escala horizontal k . Sejam também, x_s a idade na escala padrão, $g_s(x_s)$ a função de densidade padrão dos casamentos dada em (2.7) para a idade x_s , e $G_s(x_s)$ a função de distribuição acumulada dos casamentos na idade x_s .

A partir das definições dadas acima, tem-se:

- distância da idade x é k vezes x_s , i.e.:

$$x - a_0 = kx_s$$

- proporção de mulheres casadas na idade x :

$$G(x) = CG_s \left(\frac{x - a_0}{k} \right) \quad (2.8)$$

- frequência de casamentos à idade x :

$$g(x) = G'(x) = \frac{C}{k} g_s \left(\frac{x - a_0}{k} \right) \quad (2.9)$$

Destas relações tem-se que, uma vez conhecidas a_0 , k e C , pode-se calcular a função de densidade de casamentos por idade, para qualquer coorte que esteja caracterizada por estes três parâmetros.

2.1.3 Reparametrização da Função de Densidade Padrão de Primeiros Casamentos

Os parâmetros de locação e de escala (a_0 , k) do modelo de Coale, não são de fácil interpretação e sua determinação é bastante arbitrária. Parâmetros de maior interesse e de mais fácil estimação são a média e o desvio padrão da idade à primeira união.

Rodriguez e Trussel (1980) propõem reparametrizar a função de densidade padrão dada por Coale, de tal forma que o modelo seja ajustado através dos parâmetros μ (média) e σ (desvio padrão) da idade ao casar ao invés de a_0 e k .

Procura-se então, uma nova função de densidade padrão para a idade ao casar, com média zero e variância igual à unidade.

Se a função de densidade da variável aleatória X_s é dada por (2.7), o valor esperado, $E(X_s)$ é igual à 11.36 e a variância, $Var(X_s)$, é igual a 43.34.

Considerando-se

$$X = kX_s + a_0$$

tem-se:

$$E(X) = a_0 + 11.36k$$

$$Var(X) = 43.34k^2$$

Resolvendo-se o sistema de equações para $E(X) = 0$ e $Var(X) = 1$ resulta $a_0 = -1.726$ e $k = 0.152$.

Substituindo-se estes resultados em (2.9) obtém-se $g_0(x)$, a função de densidade da idade ao casar para uma população com $a_0 = -1.726$ e $k = 0.152$, onde:

$$g_0(x) = \frac{C}{0.152} g_s \left(\frac{x + 1.726}{0.152} \right)$$

Assumindo-se $C = 1$, ou seja, todas as mulheres da população se casam até o fim do processo e, substituindo na função de densidade padrão de casamentos por idade proposta por Coale e dada em (2.7), obtém-se:

$$g_0(x) = 1.12811 \exp(-1.145(x + 0.805) - \exp(-1.896(x + 0.805)))$$

A função $g_0(x)$ será a nova função de densidade padrão para a idade ao casar, com a propriedade de ter média 0 e variância 1. A população na qual está baseada esta nova função de densidade padrão, para a idade à primeira união, é uma população hipotética, que difere da população considerada por Coale para obter a função de densidade padrão de casamentos (2.7), nos parâmetros a_0 e k , que assumem os valores -1.726 e 0.152 respectivamente.

Para obter a função de densidade de uma população com idade média ao casar μ e desvio padrão σ faz-se uma transformação análoga à que foi realizada anteriormente.

Para uma população com idade média ao casar μ e desvio padrão σ , a função de densidade da idade ao casar assume a seguinte expressão:

$$g(x) = \frac{C}{\sigma} g_0 \left(\frac{x - \mu}{\sigma} \right)$$

onde C corresponde à proporção de mulheres não solteiras ao final do processo.

Durante o desenvolvimento do modelo de nupcialidade por idade, Coale constata que tal modelo, combinado com uma estrutura de fecundidade marital por idade, determinaria um modelo de fecundidade por idade. Esta idéia é apresentada nas seções seguintes.

2.2 Modelo de Fecundidade

2.2.1 Modelo de Fecundidade de Coale e Trussell

O suposto básico no modelo de fecundidade por idade proposto por Coale e Trussell (1974) é que fecundidade específica na idade a , $f(a)$, é o produto da proporção de mulheres de idade a que vivem em união, $G(a)$, e a fecundidade dessas mulheres que vivem em união, $r(a)$. Numa população hipotética onde os nascimentos só ocorrem dentro do casamento e na qual não existe dissolução das uniões, ou as viúvas ou separadas voltam a se unir imediatamente, $G(a)$ é a proporção de mulheres alguma vez unidas e $r(a)$ a fecundidade marital. Na seção anterior foi mostrado que $G(a)$ segue uma forma padrão que, entre uma população e outra, difere somente na idade do início das uniões conjugais e no ritmo ao qual se atinge a proporção final de mulheres alguma vez unidas. A variação na estrutura de $r(a)$ é também limitada como é mostrado por Coale (1971), a partir da conceito de fecundidade natural elaborado por Henry (1961).

Henry (1961) identifica a existência de um padrão característico de fecundidade marital em populações onde o controle voluntário dos nascimentos é nulo ou quase nulo. Ele define o **controle voluntário** como uma conduta que afeta a fecundidade e que é modificada conforme a parturição⁸ aumenta e a **ausência de controle ou fecundidade natural**, como a conduta que pode (ou não) afetar a fecundidade, mas que sempre é a mesma independente do número de filhos tidos⁹. Estes conceitos

⁸Número de filhos tidos.

⁹A amamentação prolongada, abstinência ou pouca frequência de relações sexuais, até o uso de anticoncepcionais podem-se incluir neste tipo de conduta, sempre e quando não sejam usados com o fim de limitar o número de filhos.

precisam ser muito bem compreendidos pois, daqui em diante, serão parte essencial das idéias apresentadas no resto deste trabalho.

Coale (1971) mostra que o padrão de fecundidade marital por idade de uma população qualquer, ou segue uma fecundidade natural, (se não é praticado um controle voluntário dos nascimentos), ou se afasta desta segundo um padrão específico, que aumenta com a idade. Este afastamento específico é dado por:

$$\frac{r(a)}{n(a)} = M \exp(m \cdot v(a)) \quad (2.10)$$

onde:

$r(a)$: taxa de fecundidade marital na idade a da população estudada

$n(a)$: taxa de fecundidade marital na idade a da população com fecundidade natural

M : parâmetro de escala que expressa o nível de fecundidade marital ao redor dos 20 anos em termos da fecundidade natural.

$v(a)$: expressa o padrão típico da redução da fecundidade na idade a , por contraste à fecundidade natural. Em outras palavras, representa a tendência das mulheres de idade a de praticar contracepção ou aborto com o fim de reduzir o nível de fecundidade natural.

m : expressa o grau de controle da fecundidade na população estudada. ($m = 0$ quando $r(a) = n(a)$, para todo valor de a ; $m > 0$ implica um maior declínio da fecundidade por idade que o estabelecido pelo padrão de fecundidade natural; $m < 0$ implica um menor declínio da fecundidade marital da população observada com respeito à fecundidade natural; valores de m maiores de 1 evidenciam um considerável controle da fecundidade).

As séries $n(a)$ e $v(a)$ são assumidas invariantes e foram calculadas empiricamente por Coale e Trussell (1974) da seguinte maneira:

- Foram estimados seis valores para $n(a)$, um para cada grupo etário de tamanho cinco anos (desde 20-24 até 45-49 anos), calculando a média aritmética das séries consideradas como de fecundidade natural por Henry (1961). As séries de Henry começam na idade de 20 anos, pois concepções premaritais têm um efeito,

considerável e irregular, na fecundidade marital das idades menores de 20. Das séries consideradas por Henry como de fecundidade natural foram selecionadas dez, omitindo aquelas que foram baseadas em pesquisas de pouca confiabilidade.

- Foram obtidos seis valores para $v(a)$, um para cada grupo etário a partir dos 20-24 anos, como anteriormente, selecionando 43 séries de fecundidade marital registradas no Anuário Demográfico das Nações Unidas (1965), dos países onde os dados são de boa qualidade.

Para obter $v_i(a)$ (para a idade a no i -ésimo grupo etário), se faz $m = 1$ e utiliza-se (2.10):

$$v_i(a) = \ln \left(\frac{r_i(a)}{M \cdot n_i(a)} \right) \quad (2.11)$$

O parâmetro de escala M , que representa o nível de fecundidade marital para o grupo etário 20-24, foi escolhido de tal forma que $v_i(a)$ fosse zero para o grupo etário 20-24. A média aritmética dos 43 valores obtidos para cada grupo etário foi definida como $v(a)$ para esse grupo etário.

Os valores resultantes dos procedimentos acima explicados são:

	grupo etário					
	20-24	25-29	30-34	35-39	40-44	45-49
$n(a)$	0.460	0.431	0.396	0.321	0.167	0.024
$v(a)$	0.000	-0.279	-0.667	-1.042	-1.414	-1.671

Os valores de $n(a)$ e $v(a)$ para idades simples são obtidos por interpolação. Os valores da série $n(a)$ para idades de 12 a 20 anos estão baseados na informação biomédica de que uma capacidade reprodutiva plena é atingida alguns anos após da menarca¹⁰, e que a idade média da menarca varia entre os 12 e 16 anos.

Do dito acima tem-se que a estrutura da fecundidade marital para uma população tem uma variabilidade limitada, cujo afastamento da fecundidade natural

¹⁰Primeira menstruação.

depende somente dos parâmetros de escala (M), representando o nível de fecundidade marital da população, e um parâmetro (m) que representa o grau de controle da fecundidade nessa população. A estrutura de fecundidade marital $r(a)$, é combinada com a estrutura de nupcialidade, $G(a)$, para obter estruturas de fecundidade. Esta idéia é elaborada por Coale e Trussell (1974) para obter a série de estruturas de fecundidade humana.

Tabelas modelo de fecundidade de Coale e Trussell

Considerando os modelos propostos por Coale (1971) de nupcialidade e fecundidade marital, Coale e Trussell (1974) construíram um conjunto de tabelas modelo que representam as diferentes estruturas de fecundidade por idade encontradas na população humana. Estas tabelas permitem analisar a fecundidade utilizando somente três parâmetros: a idade de início das uniões conjugais (a_0), o ritmo de aumento das uniões (k), e o grau de controle da fecundidade (m). Os autores fornecem também um programa computacional para calcular a estrutura da fecundidade para valores específicos dos parâmetros a_0 , k e m .

A fecundidade para cada idade simples foi calculada como o produto de um número que representa a quantidade de mulheres que vivem em união e outro número que representa a fecundidade específica por idade para aquelas mulheres que vivem em união. A partir das combinações destas duas séries, são construídas tabelas que expressam todas as estruturas de fecundidade por idade que podem ser achadas na raça humana. Esta afirmação baseia-se na regularidade do padrão de nupcialidade por idade e da variação de fecundidade marital com a idade e do padrão de afastamento da fecundidade marital respeito da fecundidade natural, segundo a idade, também sugeridos por Coale(1971).

A taxa de fecundidade por idade é o resultado do produto da proporção de mulheres alguma vez unidas nessa idade, pela taxa de fecundidade das mesmas, ambas referidas à idade a , (Coale e Trussell, 1974), ou seja:

$$f(a) = G(a)r(a) \quad (2.12)$$

onde:

$f(a)$: taxa específica de fecundidade para a idade a .

$G(a)$: proporção de mulheres não solteiras na idade a .

$r(a)$: taxa de específica fecundidade marital para a idade a .

É importante notar que (2.12) aplica-se sob o suposto de que não existem nascimentos fora da união e não há dissolução da união antes do fim do período reprodutivo. Como estas condições são quase impossíveis de serem satisfeitas na prática, Coale e Trusell sugerem modificar levemente os parâmetros que determinam as funções $G(a)$ e $r(a)$ de tal forma que representem a conduta real da fecundidade na população estudada¹¹.

A representação da fecundidade pela expressão (2.12) torna possível o cálculo de padrões de fecundidade com a especificação de três parâmetros: a_0 e k para determinar $G(a)$, e m para determinar $r(a)$. Note-se que os parâmetros de escala C e M , presentes nas equações (2.8) e (2.10) que definem $G(a)$ e $r(a)$, respectivamente, não são utilizados porque as tabelas construídas indicam o padrão de fecundidade por idade e não o nível da mesma. Porém, os níveis de fecundidade específica podem ser obtidos multiplicando cada taxa específica dada pelo modelo, pela taxa de fecundidade total (TFT)¹² da população. Isto é possível porque a soma das fecundidades específicas das tabelas modelo é igual a 1.0.

Considerando a equação (2.10), omitindo o parâmetro de escala, tem-se que a taxa de fecundidade marital para a idade a é dada por

$$r(a) = n(a) \exp(m \cdot v(a))$$

Substituindo este resultado na equação (2.12), tem-se que a fecundidade para a idade a pode ser expressa como:

$$f(a) = G(a)n(a) \exp(m \cdot v(a))$$

¹¹As concepções pre-nupciais podem ser incorporadas com um valor menor de a_0 ; os nascimentos por parte das mulheres solteiras nas idades mais avançadas podem ser assimilados por uma maior fecundidade nas mulheres unidas (aumentando $r(a)$); o efeito do término de uma união por um maior grau de controle da fecundidade (modificando m).

¹²Número de filhos nascidos vivos que teriam, em média, as mulheres durante a sua vida reprodutiva, se as taxas de fecundidade específicas por idade se mantivessem constantes e se nenhuma mulher morresse antes de finalizar o seu período reprodutivo.

onde $G(a) = \int_0^a g(x) \partial x$, com $g(x)$ dada em (2.9), omitindo também o fator de escala C .

O padrão de fecundidade por idade apresentado nesta seção, assim como o modelo de nupcialidade apresentado na seção anterior, são muito úteis para fins de estimação e análise¹³. Porém, como é mencionado em Coale (1979), existem limitações nas tabelas de fecundidade e nupcialidade elaboradas. Na seguinte seção se apresenta o modelo de fecundidade marital formulado por Page (1977) como uma tentativa de aperfeiçoar o modelo pionero de Coale.

2.2.2 Modelo de Fecundidade Marital Page

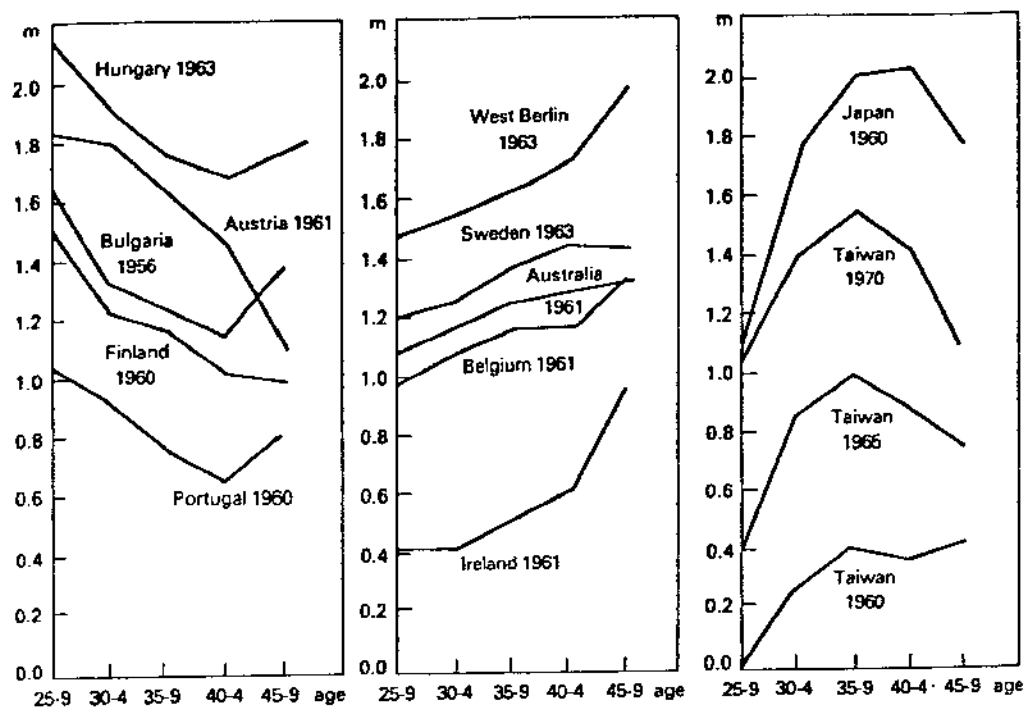
O modelo de fecundidade marital de Coale é interessante pois decompõe a estrutura da fecundidade marital por idade em duas estruturas baseadas na idade $n(a)$ e $v(a)$, (ver equação (2.10)) com significado demográfico. Outro fato interessante é a sua simplicidade: requer somente um parâmetro (m , intensidade média do controle da fecundidade na população), para descrever e comparar os padrões de fecundidade marital (Page(1977)).

Mas, pelo fato desta simplicidade, o modelo resulta numa aproximação um pouco grosseira do padrão real de fecundidade da população. Na prática, existem exceções com relação ao padrão de divergência entre a estrutura de fecundidade marital e a estrutura de fecundidade natural ($r(a)/n(a)$) dada em (2.10). Isto pode ser observado nas figuras 2.5 e 2.6. Se o modelo de Coale estivesse absolutamente correto, o valor de m não deveria variar nos grupos etários da população. Porém, as Figuras 2.5 e 2.6 mostram variações importantes e sistemáticas do valor de m . Parte da falta de ajuste observada, pode-se dever ao uso de uma estrutura de fecundidade natural que não é necessariamente a mesma para todas as populações, mas a causa fundamental da falta de ajuste é sua simplificação extrema.(Page(1977)).

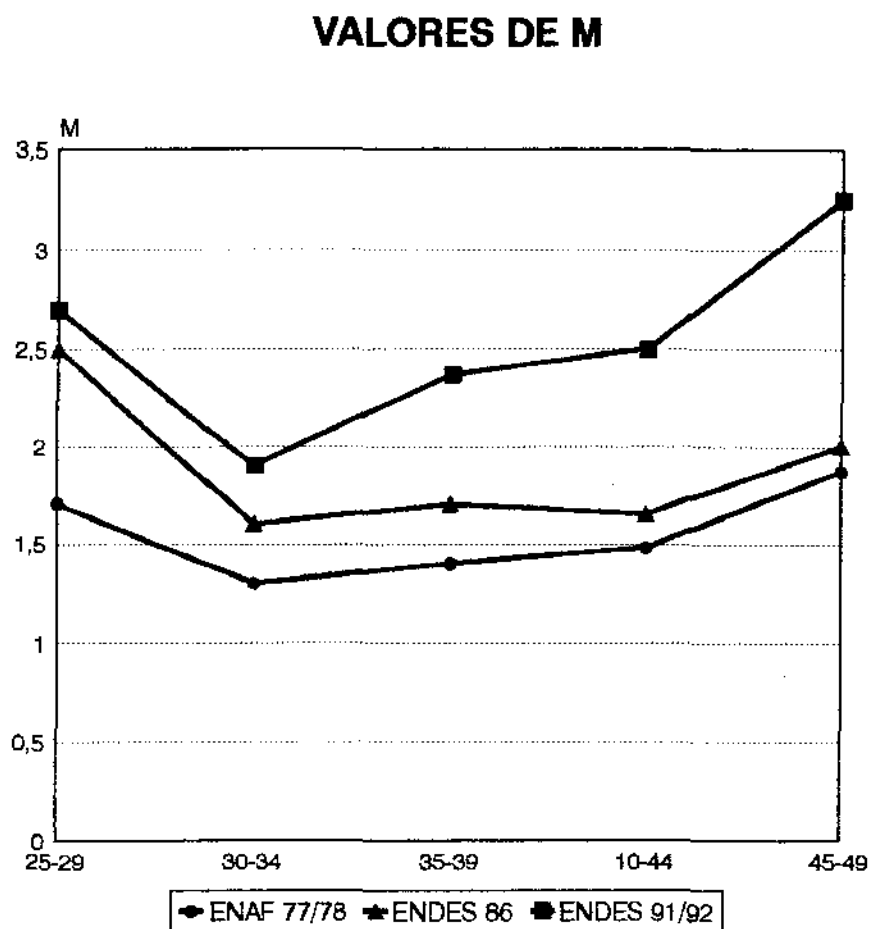
A idade é uma variável determinante para o padrão de fecundidade natural, mas é inadequada para explicar o padrão de controle da fecundidade. O controle da

¹³Exemplos ilustrativos destes modelos para a situação brasileira são os trabalhos de Altmann e Wong (1980), Altmann e Wong (1983), Wong (1983), Oliveira (1991).

Figura 2.5: Valores de m , por idade, para populações selecionadas.



Fonte: Page (1977), Figura 1, p.87.

Figura 2.6: Valores de m , por idade. Peru, anos 1977, 1986, 1992.

Fonte: Dados ENAF 77/78, ENDES 86, ENDES 91/92.

fecundidade depende da experiência prévia em fecundidade, ou seja, do número de filhos tidos. A idade não pode fornecer a mesma relação da experiência em fecundidade para todas as populações, pois a idade à qual as mulheres começam a ter filhos varia muito de uma população para outra. Para populações onde os nascimentos ocorrem dentro das uniões, a duração da união é uma variável explicativa mais direta que a idade para detectar padrões de controle da fecundidade. Deve-se observar que a duração da união por si só não pode descrever a fecundidade marital. A idade da mulher também deve ser levada em consideração (Page,1977).

Por estes motivos, Page (1977) propõe um modelo mais geral de fecundidade marital por idade à primeira união e duração dessa união¹⁴. A autora deriva este modelo partindo de um estudo preliminar que revela que a diferença proporcional em fecundidade entre um grupo etário e o seguinte é a mesma para os grupos com o mesmo tempo de duração da união. Em outras palavras, a diferença nas taxas de fecundidade entre mulheres de duas durações de união é proporcionalmente a mesma para todos os grupos etários.

Page formula então o seguinte modelo:

$$f(a, d, t) = T(t) \cdot A(a, t) \cdot D(d, t) \quad (2.13)$$

onde:

$f(a, d, t)$: taxa de fecundidade na idade a e duração da primeira união d para o período t .

$T(t)$: nível geral de fecundidade no período t sobre todas as idades e durações.

$A(a, t)$: padrão de fecundidade por idade no período t . É compartilhado por todas as coortes de uniões.

$D(d, t)$: padrão de fecundidade por duração da união no período t . É compartilhado por todos os grupos etários.

Para estimar estas funções, Page ajusta o modelo a uma série de dados registrados na Suécia, Inglaterra e Austrália, utilizando o método baseado nas médias

¹⁴Note-se que a idade da mulher está indiretamente presente no modelo pois é igual à soma destas duas variáveis.

trimmed, desenvolvido por Tuckey (1977), obtendo um excelente resultado no que se refere ao ajuste.

Analisando as funções $A(a, t)$ e $D(a, t)$ estimadas, a autora encontra dois resultados importantes que vão servir para simplificar o modelo, sem perder a qualidade do ajuste do mesmo:

- $A(a, t)$ é quase constante no tempo e segue um padrão semelhante ao padrão de fecundidade natural de Henry.
- $D(d, t)$ é uma função monótona da duração d , para cada período t , e pode ser bem aproximada por uma função exponencial.

Utilizando estes resultados, o modelo dado em 2.13 é simplificado e reescrito como:

$$f(a, d) = \theta n(a) \exp(\beta d) \quad (2.14)$$

onde:

$f(a, d)$: taxa específica de fecundidade marital à idade a .

θ : parâmetro que representa o nível de fecundidade marital.

$n(a)$: taxa de fecundidade natural à idade a .

β : parâmetro que representa o grau de afastamento da fecundidade marital com respeito à fecundidade natural.

Tomando logaritmo em ambos os lados da equação 2.14:

$$\ln \left(\frac{f(a, d)}{n(a)} \right) = \alpha + \beta d \quad (2.15)$$

onde $\alpha = \ln \theta$.

Daqui, tem-se que o logaritmo da razão entre a fecundidade marital e a fecundidade natural, em idade e duração determinadas, é uma função linear do tempo transcorrido desde a primeira união.¹⁵

¹⁵Uma aplicação deste modelo, sob o enfoque Rodriguez e Cleland, assim como o modelo de Coale e Trussell, para o estudo da fecundidade marital no México, pode ser encontrado em Moreno(1990).

Rodriguez e Cleland (1988) mostram que no modelo de Page pode-se separar os padrões de fecundidade por idade e por duração, que implicitamente participam no modelo.

- Padrão de fecundidade por idade

$$f(a) = n(a) \exp(\alpha + \beta d_c(a))$$

onde $d_c(a)$ é um valor de duração dependente da idade tal que $\exp(\beta d_c(a))$ seja o grau de controle na idade a para todas as durações.

- Padrão de fecundidade por duração da primeira união

$$f(d) = \bar{n}(d) \exp(\alpha + \beta d)$$

onde $\bar{n}(d)$ é a fecundidade natural média para mulheres com duração d .

2.3 Comentários

Fazendo uma síntese dos conceitos apresentados neste capítulo, tem-se que o modelo de nupcialidade e de fecundidade de Coale (1971), ambos baseados na idade da mulher, constituem os primeiros esforços realizados com o objetivo de modelar a fecundidade, utilizando variáveis explicatórias com significado inerente ao processo.

O modelo de Coale e Trussell (1974) combina os dois modelos anteriores, fornecendo um modelo de fecundidade, cuja principal vantagem é a simplicidade, pois emprega somente três parâmetros. Já o modelo de fecundidade marital de Page (1977), resulta num modelo mais completo, pois incorpora indiretamente o efeito da nupcialidade e da experiência prévia em fecundidade, ao agregar a duração da união como variável explicatória. A desvantagem deste modelo é sua difícil utilização, pois precisa de dados com uma estrutura mais complexa, com tabulações por idade da mulher e duração da união. Apesar dessas dificuldades, para os casos em que se dispõe

desta informação poderia ser interessante tentar combinar o modelo de fecundidade marital de Page com o padrão de nupcialidade de Coale para obter um novo modelo de análise da fecundidade.

No próximo capítulo será discutido o modelo de Page sob um ponto de vista estatístico, utilizando os conceitos de modelos lineares generalizados apresentados no primeiro capítulo, que permitirá estimar o modelo de Page de uma forma mais geral. Isto vai colocar em evidência um ponto de junção entre a metodologia estatística e a metodologia demográfica, na implementação de um modelo de análise da fecundidade marital, onde os parâmetros estimados terão um sentido demográfico, além de propriedades estatísticas que fornecem medidas de ajuste do modelo.

Capítulo 3

MODELO LOG-POISSON PARA A FECUNDIDADE MARITAL

Neste capítulo se apresenta uma aproximação log-linear do modelo de fecundidade marital de Page, discutido no capítulo anterior, que corresponde a um modelo linear generalizado, especificamente ao modelo log-Poisson apresentado no capítulo 1.

O suposto básico que permite chegar a esta aproximação é considerar o número de nascimentos das mulheres unidas como uma variável aleatória com distribuição Poisson cuja média depende da taxa de fecundidade marital.

Esta aproximação permite a realização da estimação estatística dos parâmetros e medir a bondade de ajuste do modelo. Porém, os parâmetros estimados têm uma interpretação demográfica que será explicada neste capítulo.

3.1 Aproximação Log-Linear

Rodriguez e Cleland (1988) analisam o modelo de fecundidade de Page (1977) e propõem uma aproximação log-linear do mesmo. Com este objetivo, os autores definem:

- $B(a, d)$ como o número de nascimentos observados nas mulheres de idade a e duração da união d ;

- $T(a, d)$ como o tempo de exposição à gravidez¹ observado nas mulheres de idade a e duração da união d .

Os autores assumem que as contagens de nascimentos $B(a, d)$ são variáveis aleatórias independentes com distribuição Poisson. Seu valor esperado é igual ao produto do tempo de exposição $T(a, d)$ e uma taxa de fecundidade marital teórica $f(a, d)$. Ou seja:

$$B(a, d) \sim \text{Poisson}(\mu_{a,d})$$

com

$$\mu_{a,d} = E(B(a, d)) = T(a, d) \cdot f(a, d) \quad (3.1)$$

A escolha da distribuição de Poisson pode ser justificada em termos de supostos naturais relativos à natureza estocástica das contagens de nascimentos. Porém, o aspecto essencial desta escolha é que a distribuição de Poisson capta o resultado empírico de maior variância da contagem dos nascimentos para maiores níveis de fecundidade. (Rodriguez e Cleland, 1988).

Assume-se então que a taxa de fecundidade teórica necessária em (3.1), $f(a, d)$, segue a forma proposta no modelo simplificado de Page, dado em (??). Isto é,

$$f(a, d) = \theta n(a) \exp(\beta d)$$

onde:

θ representa o nível de fecundidade marital da população em estudo

$n(a)$ é a taxa de fecundidade natural para a idade a

β é o parâmetro que representa o grau de afastamento da fecundidade marital da população, com respeito da fecundidade natural.

Portanto, a equação (3.1) pode ser escrita como:

$$E(B(a, d)) = T(a, d) \theta n(a) \exp(\beta d) \quad (3.2)$$

¹O tempo de exposição à gravidez pode ser considerado como o tempo em que uma mulher vive em união sexual.

Considerando-se o logaritmo em ambos os lados da equação (3.2), tem-se:

$$\ln(E(B(a, d))) = \ln(T(a, d)) + \ln(n(a)) + \alpha + \beta d \quad (3.3)$$

onde $\alpha = \ln(\theta)$.

Este modelo pertence à classe dos modelos lineares generalizados definidos por Nelder e Wedderburn, pois corresponde a caracterização dada na seção 1.1, com:

i) A componente aleatória dada por $B(a, d)$ com distribuição de $B(a, d)$ pertencente à família exponencial (distribuição Poisson).

ii) A componente sistemática, o preditor linear, está dada por:

$$\eta_{a,d} = \ln(T(a, d)) + \ln(n(a)) + \alpha + \beta d$$

Este preditor linear tem duas componentes: um termo conhecido (offset) formado por $\ln(T(a, d)) + \ln(n(a))$ e uma regressão linear simples da duração com intercepto α e inclinação β .

iii) A função de ligação é de tipo logarítmico entre o valor esperado de $B(a, d)$ e o preditor linear.

Já que as variáveis aleatórias $B(a, d)$ tem função de distribuição de Poisson e função de ligação de tipo logarítmico, temos que (3.3) é um modelo log-Poisson generalizado (Ver seção 1.5.1).

Uma implicação importante deste fato é que estimadores de máxima verossimilhança para os parâmetros do modelo estão disponíveis, além de estimadores dos erros padrões e a estatística de bondade de ajuste χ^2 de razão de log-verossimilhança.

3.1.1 Bondade de Ajuste do Modelo

A medida de bondade do ajuste definida na seção 1.4.1 corresponde neste caso a

$$D = 2 \sum B(a, d) \ln \left[\frac{B(a, d)}{\widehat{B(a, d)}} \right] \quad (3.4)$$

onde:

$B(a, d)$: corresponde aos valores observados

$B(\widehat{a}, d)$: corresponde aos valores ajustados

Como o tempo de exposição à gravidez $T(a, d)$ é conhecido , tem-se que:

$$B(\widehat{a}, d) = T(a, d) \cdot f(\widehat{a}, d)$$

Portanto, (3.4) pode ser escrita como

$$D = 2 \sum T(a, d) f(a, d) \ln \left[\frac{f(a, d)}{f(\widehat{a}, d)} \right]$$

Representando $T(a, d)$ como uma fração do tempo total de exposição T , somado sobre todas as caselas de idade e duração, define-se $w(a, d) = \frac{T(a, d)}{T}$, então:

$$D = 2T \sum w(a, d) f(a, d) \ln \left[\frac{f(a, d)}{f(\widehat{a}, d)} \right] \quad (3.5)$$

onde $f(a, d)$ e $f(\widehat{a}, d)$ representam, as taxas de fecundidade observadas e ajustadas pelo modelo, respetivamente.

Como resultado temos que a medida de bondade de ajuste D é o produto do tempo total de exposição e uma medida de discrepância entre as taxas observadas e ajustadas de fecundidade, ponderada pelo tempo de exposição para cada categoria de idade e duração de união. Esta medida de bondade de ajuste é, então, sensível ao tamanho da amostra e à falta de ajuste em si.

Rodriguez e Cleland (1988) utilizam este novo modelo para estimar a estrutura da fecundidade marital em 38 países, ao redor do ano 1977, obtendo um bom ajuste em todos estes. Porém, as faltas de ajuste encontradas foram devidas, na sua maioria, a erros nas medidas das taxas de fecundidade específica, à situação particular da fecundidade nos primeiros anos de união e a presença de algumas observações atípicas nas amostras.

3.2 Interpretação dos Parâmetros

Os parâmetros α e β que são estimados ao ajustar o modelo ao conjunto de dados, têm uma interpretação do ponto de vista do fenômeno demográfico analisado.

Este fato é mostrado por Rodriguez e Cleland (1988) ao estudar as correlações existentes entre estes parâmetros e as condutas de controle da fecundidade, nos países considerados. Estes autores interpretam os parâmetros α e β como componentes de espaçamento e limitação da reprodução.

A seguir apresenta, com mais detalhe, a interpretação que Rodriguez e Cleland fazem destes parâmetros.

3.2.1 Parâmetro α

O parâmetro α reflete o nível de fecundidade natural da população em estudo. Está em escala logarítmica ($\alpha = \ln \theta$) e assume geralmente valores negativos.

Este parâmetro é redefinido por Rodriguez e Cleland (1988) como uma porcentagem da redução do nível observado da fecundidade marital em relação a um máximo biológico, introduzindo o índice I_a , definido como:

$$I_a = 100 \left(1 - \frac{11.85}{15.3} e^\alpha \right)$$

Após expressar o parâmetro α na escala original (fazendo $\theta = e^\alpha$), se multiplica este valor por 11.85, soma das taxas específicas da fecundidade natural dadas por Coale e Trussell ($n(a)$), e dividido por 15.3, valor máximo da fecundidade natural determinada por Bongaarts².

Conforme o nível de fecundidade diminui o índice I_a aumenta. Este índice não pode exceder a 100, mesmo em populações com um nível de fecundidade marital muito baixo, mas pode assumir valores negativos em populações cuja fecundidade marital exceda 15.3 filhos por mulher.

Os fatores que podem afetar o parâmetro α são a duração do período de infertilidade post-parto (determinado principalmente pelo período de amamentação, abstinência post-parto e amenorréia), a freqüência das relações sexuais, a mortalidade intrauterina espontânea e a duração do período fértil. (Rodriguez e Cleland, 1988)

²Bongaart (1978) supõe que a taxa de fecundidade total de 15.3 filhos por mulher é o máximo para qualquer população. Esta convenção é utilizada por Rodriguez e Cleland (1988) para permitir a comparabilidade do Índice I_a em populações com uma elevada fecundidade marital.

3.2.2 Parâmetro β

O parâmetro β reflete o grau de controle da fecundidade marital. Também está em escala logarítmica. Geralmente assume valores muito próximos a zero, que indicam que não há controle, ou valores negativos não muito distintos de zero em populações onde existe um controle da fecundidade dentro da união.

Re-expressando β numa escala mais conveniente, Rodriguez e Cleland (1988) definem o índice I_β da seguinte maneira:

$$I_\beta = 100 (1 - e^{10\beta})$$

O parâmetro β é multiplicado por 10 (valor arbitrário) para obter uma medida de controle para 10 anos de união, e logo é exponenciado para retornar à escala original de fecundidade. Desta maneira, o índice I_β representaria a queda nos níveis de fecundidade após 10 anos de união. Conforme o controle da fecundidade se incrementa, o índice I_β aumenta. Este índice não pode exceder a 100 ainda em populações com elevados níveis de controle, mais pode assumir valores negativos se a fecundidade marital decresce muito lentamente com a duração da união.

O parâmetro β (ou de forma equivalente, o índice I_β) está relacionado com o uso dos anticoncepcionais e o aborto com o propósito de limitar o número de filhos (Rodriguez e Cleland,1988).

3.3 Comentários

A aproximação log-linear que Rodriguez e Cleland (1988) propõem para o modelo de fecundidade marital de Page (1977), representa um novo modelo para a análise da fecundidade marital. Moreno (1990) refere-se a este modelo como o modelo de Rodriguez-Cleland. Neste modelo a estimação dos parâmetros só depende da distribuição das mulheres por idade e duração da união. Uma outra propriedade deste modelo é que permite a estimação da taxa de fecundidade marital total através de uma aproximação que depende somente da duração da união, usando a expressão:

$$TFMT(\alpha, \beta, d) = \int_0^{50-m} n(m+y) \exp(\alpha + \beta y) \partial y$$

onde m representa a idade média ao casar na amostra das mulheres alguma vez unidas, que se uniram nos 20 anos anteriores à entrevista, ou seja, mulheres com no máximo 20 anos de união.

O modelo de Rodriguez-Cleland será utilizado no próximo capítulo para analisar a fecundidade marital no Peru, entre os anos 1977-1992. Os dados utilizados correspondem a três levantamentos realizados no Peru a nível nacional. Os níveis e tendências da fecundidade das mulheres peruanas que vivem em união serão estimados através do modelo apresentado neste capítulo.

Capítulo 4

ANÁLISE DA FECUNDIDADE MARITAL NO PERU ATRAVÉS DO MODELO LOG-POISSON

Os modelos lineares generalizados (apresentados no capítulo 1) e a sua relação com a função analítica que modela o comportamento reprodutivo das mulheres que vivem em união (apresentado no capítulo 2), são considerados neste capítulo na análise da fecundidade marital no Peru aqui apresentada. Nesta análise serão estudados os padrões e mudanças da fecundidade marital ao longo do período 1977-1992 e seus diferenciais entre grupos com características geográficas e socioeconômicas diversas.

Para esta análise foram utilizados os dados levantados em três pesquisas realizadas no Peru, a nível nacional, nos anos 1977, 1986 e 1992.

Com o objetivo de apresentar os principais fatores que influem no comportamento reprodutivo e as características que estes assumem no Peru, se apresenta nas próximas seções um breve resumo do contexto peruano, assim como dos fatores socioeconômicos que estariam afetando a fecundidade marital.

4.1 Peru: Principais Aspectos de Interesse

Apresenta-se, nesta seção uma síntese da situação peruana no que se refere àqueles fatores que teriam uma influência significativa nas mudanças que experimenta o comportamento reprodutivo. A heterogeneidade estrutural, que ao nível econômico e social caracteriza os países da América Latina, determina um comportamento também diferencial em termos de fecundidade. Os fatores sociais que teriam uma grande importância teórica na orientação do comportamento reprodutivo são mencionados por Guadalupe (1988) como: o nível de educação da mulher e a sua participação na atividade econômica e a eficiência e abrangência dos serviços de saúde e segurança social.

Os fatores que serão levados em consideração nesta análise específica da fecundidade são: a área de residência da mulher (urbano ou rural), sua região de residência (Lima Metropolitana, resto da Costa, Serra ou Selva), seu nível de educação (medido em três grupos de anos de estudos), e o tipo de união (formal ou consensual). Acredita-se que estas variáveis estão fortemente relacionadas com os fatores mencionados por Guadalupe como relevantes no comportamento reprodutivo. Os níveis de participação da mulher no mercado de trabalho assim como os serviços de saúde e segurança social, variam segundo o nível de urbanização e a região de residência, portanto, estas variáveis de alguma forma refletem esses fatores. O tipo de união será considerado como fator relevante tomando como base o estudo realizado para o Brasil por Lazo (1994), onde se observam diferenças significativas no comportamento reprodutivo segundo o tipo de união da mulher.

Apresenta-se a seguir uma breve síntese das variáveis que serão incorporadas no modelo de análise.

4.1.1 Regiões Geográficas

O Peru está dividido em três grandes regiões geográficas (ou naturais) que o atravessam longitudinalmente, de norte a sul: Costa, Serra e Selva. (Ver mapa do apêndice C). Estas regiões têm características muito diferentes no que diz respeito às

condições de vida e níveis de desenvolvimento econômico e social.

A região **Costa**, situada entre o Oceano Pacífico e a Cordilheira dos Andes, ocupa 11% do território nacional e nela reside 52% da população total. Nesta região localiza-se a capital, **Lima**, que é a cidade com maior densidade populacional do país (nela habitam 28% da população total) e constitui o maior centro econômico cultural e político do Peru. A grande corrente migratória à capital originou um crescimento desmesurado desta nos últimos anos, sobretudo na periferia da cidade, zona urbano-marginal, com sérios problemas de saneamento básico (falta de água encanada, eletricidade, esgoto, etc.).

A região **Serra**, localizada na Cordilheira dos Andes, está constituída por 26% do território nacional e nela habita 37% da população total. Sua população reside entre 2000 e 3500 metros acima do nível do mar, nas localidades onde existem as melhores condições para o desenvolvimento agrícola. É uma região preponderantemente rural.

A região **Selva** é a menos povoada do Peru, pois representa 63% do território nacional e compreende somente 11% da população peruana. Os seus habitantes residem em povoados dispersos ao longo de rios, freqüentemente isolados e de difícil acesso. Esta região apresenta os mais baixos níveis de desenvolvimento econômico e social.

4.1.2 Área de Residência

É de vital importância considerar a divisão do território nacional em áreas urbana e rural. As profundas desigualdades nos níveis de vida e nas oportunidades de tipo econômico, social e cultural, que cada uma oferece a sua população, condiciona o comportamento reprodutivo nestas áreas. Neste sentido, a população urbana tem maior acesso aos serviços de educação e saúde que a população rural, melhores condições de moradia e nutrição e, em geral, conta com uma maior quantidade de serviços básicos e com um maior número de fontes e canais de informação. Pode-se concluir, portanto, que as condições de vida na área urbana são melhores que na área rural.

Neste trabalho se considera como população urbana a que reside em centros de pelo menos 2000 habitantes. A evolução da distribuição da população urbano e rural, nos últimos anos, pode ser observada na Tabela 4.1.

Tabela 4.1: Distribuição urbano-rural da população. Peru 1960-1990.

Ano	% urbano	% rural	total
1960	46.3	53.7	100
1970	57.4	42.6	100
1975	61.4	38.6	100
1980	64.5	35.5	100
1985	67.2	32.8	100
1990	70.2	29.8	100

Fonte: Anuário de América Latina e Caribe (1990)

É importante destacar que o grande incremento da população urbana se dá na década de 60, quando a população urbana cresce 11.1%. Este ritmo continua alto na década seguinte, com um incremento de 7.1% entre 1970 e 1980 e na última década já se observa um incremento menor (de 5.7% na década de 80). mesmo assim, a população urbana representa mais do 70% em 1990.

O processo de urbanização e industrialização, em geral, está associado a uma mudança nos valores e normas do comportamento reprodutivo, o que traz como consequência uma diminuição da fecundidade nas grandes cidades. Porém, deve-se notar que nas grandes cidades do Peru, como em muitos países da América Latina, a taxa de fecundidade continua sendo elevada devido a população que mora no setor urbano-marginal, constituída na sua maioria por migrantes da área rural, continua por um curto período com o comportamento reprodutivo anterior.(Guadalupe 1988).

A diferença no comportamento reprodutivo nas áreas urbano e rural está determinada, segundo o enfoque de Leibenstein (1967), pelos diferentes custos e utilidades que os filhos proporcionam segundo o grupo social ao qual pertence a família. Na zona rural as utilidades dos filhos são altas (mão de obra, fonte de segurança na velhice dos pais) e os custos que estes representam são baixos (gastos em educação, vestuário e alimentação são mínimos). Por outro lado, nas sociedades urbanas as util-

idades dos filhos são baixas (pois estes colaboram pouco ou nada com a renda familiar, enquanto a fonte de segurança na velhice é fornecida pelos planos de segurança social), representando custos elevados pois as despesas básicas com educação, vestuário e alimentação aumentam a medida que o nível de vida e modernização cresce.

Numerosos estudos assumem a existência de uma relação inversa entre o nível de fecundidade e o nível de renda. Sob esta suposição, uma distribuição desigual da renda traz como consequência níveis de fecundidade diferenciais, sendo os grupos com um baixo nível de renda aqueles que têm um elevado nível de fecundidade. Lesevic (1987) e Guadalupe (1988) fazem uma revisão deste tipo de enfoque para o Peru.

As crises sucessivas que experimentou a economia peruana na década dos oitenta, aumentaram a pobreza geral do país e a distribuição regressiva da renda. Para ter um idéia do nível de pobreza¹ da população peruana e sua distribuição nas áreas urbano e rural, ver a Tabela 4.2. A porcentagem de pobres mantém-se quase constante no período 1970-1991 (aproximadamente a metade da população peruana se encontra em estado de pobreza). Porém, cabe destacar, que enquanto a porcentagem de pobres se manteve constante na área rural (ao redor de 65%), esta porcentagem foi aumentando na área urbana, quase se duplicando no período de 20 anos, entre 1970 e 1991, que aumenta de 28% em 1970 a aproximadamente 50% em 1991.

Reconsiderando o mencionado anteriormente tem-se que as utilidades e os custos que cada filho representa para a família dependem da situação econômica e participação no processo de produção desta. Portanto, o tamanho de família que um grupo considera como ideal está associado ao nível econômico que este grupo possui. Assim por exemplo, os altos níveis de fecundidade que registra a área rural pode ser explicados como uma resposta às condições de extrema pobreza em que está submersa esta população. Um alto nível de fecundidade assegura, em certa medida, a sobrevivência do grupo familiar que possui as mais altas taxas de mortalidade, mantendo assim um tamanho de família que é considerado como ideal (Guadalupe 1988).

¹Magnitude da pobreza: Porcentagem de famílias cuja renda é inferior ao dobro do custo de uma cesta básica de alimentos.

Tabela 4.2: Evolução dos domicílios em situação de pobreza e indigência, segundo área de residência. Peru 1970-1991

	% POBRES		
	Indigentes	Não Cobrem o Básico	Total
1970			
Total País	25	25	50
Urbano	8	20	28
Rural	39	29	68
1980			
Total País	21	25	46
Urbano	12	23	35
Rural	37	18	55
1986			
Total País	25	27	52
Urbano	16	29	45
Rural	39	21	60
1991			
Total País	22	32	54
Lima Metropolitana	10	39	49
Costa urbana	22	33	55
Serra urbana	19	29	48
Serra rural	47	21	68

Fonte: Instituto de la Mujer -Espanha e FLACSO (1994)

Ao se tratar do tamanho ideal de família surge um outro aspecto de importância: de que maneira as famílias são assistidas pelos sistemas de planejamento familiar para que o número ideal de filhos coincida com o número real? A resposta a esta interrogação varia segundo a área de residência da família. No Peru, segundo o estudo realizado por Westoff, Moreno e Goldman (1989), analisando os dados do levantamento do ano de 1986², mais da 30% dos nascimentos foram registrados como não desejados. Também pode-se verificar que a diferença entre a taxa de fecundidade

²Este levantamento é o mesmo que será utilizado nesta análise: ENDES 1986, que será detalhado nas seções posteriores.

Tabela 4.3: Taxa de fecundidade total observada e desejada, segundo área. Peru, 1986-1992

	TFT observada	TFT desejada	diferença
1986			
Total País	4.5	2.9	1.6
Urbano	3.3	2.3	1.0
Rural	6.9	4.1	2.8
1992			
Total País	3.5	2.0	1.5
Urbano	2.8	1.7	1.1
Rural	6.2	3.0	3.2

Fonte: Westoff, Moreno, Goldman (1989) e ENDES 91/92.

total observada e a taxa de fecundidade total desejada³ é mais do dobro na área rural do que na área urbana, e esta diferença não tem diminuído significativamente nos últimos anos (ver Tabela 4.3).

4.1.3 Educação

O nível de educação é um dos fatores sociais mais amplamente estudados devido a o seu impacto no comportamento reprodutivo. O nível de educação atingido pela mulher tem provado ser uma das variáveis mais consistentes e fortemente relacionadas com a fecundidade. (Weinberger, Lloyd e Blanc, 1989)

No que se refere à influência da educação dos pais e, em particular, da mulher sobre os níveis de fecundidade tem-se que a educação afeta inversamente à fecundidade através da nupcialidade mais tardia e da orientação do comportamento reprodutivo das mulheres em união estável.

O maior nível de educação dos pais permite considerar o número de filhos como uma variável de decisão, de maneira que controlarão mais a sua fecundidade e

³A taxa total de fecundidade pode ser dividida em duas componentes: "desejada" e "não desejada". A fecundidade "não desejada" corresponde a todos aqueles nascimentos que ocorreram após o último nascimento desejado. Em teoria, estes nascimentos não teriam ocorrido se as mulheres tivessem acesso à métodos eficientes de controle de fecundidade. (Westoff, Moreno e Goldman, 1989).

terão menos filhos que outros pais com menor nível de educação, pois os primeiros desejarão fazer um maior gasto por filho para brindar-lhes um melhor nível de qualificação. As normas culturais existentes sobre o tamanho ideal da família podem ser alteradas com o aumento do nível de educação, pois as aspirações por bens materiais e qualidade educacional dos filhos aumenta. Além disso, os pais com maior nível de educação têm maior acesso a informação acerca dos métodos de controle da natalidade e são mais eficientes na sua utilização (Becker, 1960).

O nível de educação também tem um papel importante na diminuição da fecundidade das mulheres unidas, através da participação da mulher no mercado de trabalho. Quanto maior o nível de educação da mulher, maior é sua participação na atividade econômica e maiores são as oportunidades de concorrer a ocupações bem remuneradas e com maior *status*. É aqui que a fecundidade se vê modificada, pois o custo de oportunidade de um novo nascimento é elevado, devido à incompatibilidade dos papéis de mãe e de trabalhadora (Gualdalupe 1988). Note-se, porém, que isto varia dependendo do grupo de mulheres que se está considerando. O emprego feminino tradicional, com reduzidos níveis de qualificação e remuneração, não tem necessariamente efeitos sobre o nível de fecundidade. Por exemplo, o trabalho das mulheres camponesas é compatível com uma alta fecundidade (Lesevic, 1987).

Neste trabalho, o nível de educação da mulher será medido através da variável *anos de estudo*. Esta variável foi agrupada de tal forma a representar três níveis de educação: de 0 a 2 anos de estudo (grupo das mulheres analfabetas e com educação mínima), de 3 a 8 anos de estudo (grupo de mulheres com educação básica), e um terceiro grupo com as mulheres que têm 9 ou mais anos de estudo, ou seja, mulheres que atingiram pelo menos o segundo ano de educação secundária⁴. Esta divisão foi feita com o objetivo de ter um número razoável de indivíduos dentro de cada grupo e assim minimizar os erros de ajuste do modelo devido ao tamanho insuficiente de observações por nível ao realizar a análise. Acredita-se que esta classificação produz grupos homogêneos e bem diferenciados entre si, no que a nível de educação se refere.

⁴No sistema educativo peruano, 6 anos correspondem à educação primária seguidos de 5 anos de educação secundária.

4.2 Os Dados

Nesta seção são tratados aspectos específicos referentes ao conjunto de dados utilizados na análise de fecundidade marital no Peru. Detalha-se as características das fontes dos dados assim como o método de seleção da população alvo e das variáveis de interesse a partir do conjunto de dados original.

4.2.1 Fontes

Os dados utilizados nesta análise provém, como foi mencionado anteriormente, de três levantamentos nacionais sobre fecundidade e saúde materno-infantil realizados no Peru nos anos 1977/78, 1986 e 1991/92.

O levantamento do ano 1977/78 é a Pesquisa Nacional de Fecundidade do Peru (ENAF- Encuesta Nacional de Fecundidad del Perú) realizada dentro do projeto de Pesquisa Mundial de Fecundidade (WFS- World Fertility Survey). Os levantamentos dos anos 1986 e 1991/92 foram realizados sob o marco da Pesquisa Demográfica e de Saúde (DHS - Demographic and Health Survey) que, como a WFS, se realizou a nível mundial . No Peru estes levantamentos foram denominados Pesquisas Demográficas e de Saúde Familiar (ENDES-Encuesta Demográfica y de Salud Familiar).

Um dos objetivos principais destas pesquisas era determinar os níveis e tendências da fecundidade e nupcialidade e a prevalência do uso de métodos anti-concepcionais. É por isto que estas pesquisas são dirigidas a mulheres em idade reprodutiva, isto é, mulheres com idades entre 15 e 49 anos. Porém, estas pesquisas também fornecem informações sobre as características gerais da população, como: distribuição por sexo e idade, distribuição regional da população, níveis de instrução, entre outros aspectos.

A amostra ENAF 77/78 está constituída exclusivamente de mulheres alguma vez unidas, diferente do resto dos países de América Latina, onde a WFS usou uma amostra de mulheres em idade reprodutiva, independente do estado conjugal. Porém, as mulheres solteiras e aquelas de 50 anos e mais, dos domicílios amostrados, responderam sobre a sua história de fecundidade, mas restrita às datas de nascimentos de

filhos tidos nascidos vivos.

Nas ENDES, tanto na de 1986 como na de 1991/92, a população alvo está constituída por mulheres entre 15 e 49 anos, independente do estado conjugal. O tamanho das amostras destas pesquisas foram 5640, 4999 e 15882 mulheres, para a ENAF 77/78, a ENDES 86 e a ENDES 91/92, respectivamente.

Nas três pesquisas o processo de amostragem foi de multi-etapas e estratificado, com frações de amostragem variáveis. Para ENDES 91/92, por exemplo, o desenho de amostragem foi o seguinte:

- para cada região administrativa (13), a fração de amostragem foi obtida como resultado de ponderar a representatividade da região como domínio de estudo e a procura de estimadores confiáveis a nível nacional.

- a alocação ideal, segundo o primeiro critério anterior, seria atribuir $N_A/13$ entrevistas a cada região, onde N_A é o tamanho total da amostra. Esta distribuição fazia aumentar enormemente o desvio padrão dos estimadores a nível nacional.

- para solucionar o problema do aumento do erro dos estimadores nacionais, determinou-se que o número ideal de entrevistas seria proporcional ao tamanho populacional da região, ou seja, o tamanho da amostra para a i -ésima região seria $n_i = N_i/N_T$, onde N_i é a população da i -ésima região e N_T é a população total. Este procedimento atribuía, por exemplo, 170 entrevistas à região de Ucayali, cuja população projetada para a data do levantamento era de somente 54.916 habitantes, enquanto atribuía 5184 entrevistas à região de Lima, cuja população projetada era de 1.673.749 habitantes.

- como uma forma de equacionar estes dois critérios, foi proposta uma distribuição que permitia fazer inferências a nível regional e nacional, mantendo os erros dos estimadores nacionais dentro de níveis razoáveis.

A seleção da amostra foi de duas etapas nas áreas metropolitanas e de três etapas nas outras áreas. Nas áreas metropolitanas o número de conglomerados resultou da divisão do número total de entrevistas por 15 e, nas outras áreas, o número de entrevistas foi dividido por 20. Nas áreas não metropolitanas, além do número de conglomerados, também foi escolhido o número de distritos (unidades primárias de amostragem) como o número de conglomerados dividido em 3, pois de cada distrito

seriam selecionados 3 conglomerados, e logo, dentro de cada conglomerado, teria lugar a seleção dos domicílios. Ao todo foram selecionados 910 conglomerados, 514 nas áreas metropolitanas e 3 conglomerados de cada um dos 132 distritos selecionados nas áreas não metropolitanas, totalizando os 396 conglomerados restantes.

A técnica de amostragem utilizada na seleção de distritos, conglomerados e domicílios foi sistemática, com início aleatório.

Disponibilidade e Estrutura do Conjunto de Dados

Os três conjuntos de dados estão disponíveis em disquetes, ocupando uma memória de 4.6 MB, 5.8 MB e 30.7 MB para a ENAF 77/78, ENDES 86, e ENDES 91/92, respectivamente.

Para o uso em micro-computador do tipo PC, a estrutura dos arquivos é do tipo **retangular** ou **hierárquico**. Na estrutura retangular, cada mulher entrevistada está representada por um conjunto de registros, cada um dos quais fornece informação sobre um tópico do questionário⁵. O número de registros por mulher é fixo. Na estrutura hierárquica, o número de registros por mulher é variável, pois neste tipo de estrutura os registros existem somente para aquelas informações respondidas no questionário.

Por exemplo, uma mulher com 3 filhos terá somente 3 registros correspondentes à seção 22 (história de nascimentos), enquanto que uma mulher com 8 filhos terá 8 registros correspondentes a essa mesma seção. Nos arquivos rectangulares, todas as mulheres têm 20 registros atribuídos à seção de história de nascimentos, que foi considerado como o número máximo de filhos. No exemplo anterior, a primeira mulher teria 3 registros com os dados dos seus filhos e logo 17 registros em branco, e a segunda mulher teria 8 registros com os dados seguidos de 12 registros em branco.

Os arquivos que têm uma estrutura hierárquica ocupam muito menos espaço que aqueles que possuem uma estrutura retangular, porém requerem um processo um pouco mais complexo para o acesso às informações.

Cada seção do questionário está identificada por um número de dois dígitos,

⁵A seção 51 das ENDES 86 e ENDES 91/92, por exemplo, diz respeito a nupcialidade. Todas as informações referentes a este item estão contidas no registro correspondente a essa seção.

especificado no início de cada registro. Se comparadas as três pesquisas, os números das seções não coincidem necessariamente, mas as diferenças existentes entre as duas ENDES são mínimas.

Os dados utilizados na análise da fecundidade marital no Peru estavam apresentados num arquivo do tipo hierárquico. A informação foi extraída utilizando o sistema SAS (Statistical Analysis System), mediante programas que estão incluídos no apêndice B.

4.2.2 Processamento

Do conjunto de dados originais, especificado na seção anterior, foi selecionado um certo número de variáveis com o objetivo de obter um arquivo final de dados, adequado para a implementação do modelo.

A seguir especificam-se os procedimentos seguidos.

Definição da População Alvo

A população de interesse está constituída por todas aquelas mulheres que no momento da entrevista estavam unidas. Esta união pode ser do tipo formal ou consensual, mas a análise será restrita às primeiras uniões⁶. Desta maneira, pode-se estudar a fecundidade marital, isto é, a fecundidade das mulheres dentro de uma união.

Para selecionar este subconjunto da amostra, foram consideradas as variáveis *estado conjugal* e *o número de uniões* da mulher. Além disso, para controlar a qualidade da informação, foram retiradas da amostra as mulheres que tinham tido filhos antes de iniciar sua primeira união (aproximadamente 6 % das mulheres em primeira união).

Estas restrições diminuíram o tamanho das amostras para 4792, 2608 e 8113 mulheres nas ENAF 77/78, ENDES 86 e ENDES 91/92, respectivamente.

⁶As ENDES não incluem histórias de uniões, por isso não podem ser identificados os nascimentos por união, quando a mulher tem mais de um casamento. Porém, o número de mulheres com mais de uma união é relativamente pequeno (aproximadamente 8 % do total de mulheres unidas no momento da entrevista).

Seleção de Variáveis

As variáveis selecionadas diretamente do arquivo original foram :

1. Data da entrevista
2. Data de nascimento
3. Região de residência
4. Área de residência
5. Anos de estudo⁷.
6. Número de filhos tidos nos 5 anos anteriores à entrevista⁸.
7. Situação conjugal atual.
8. Data da primeira união

A partir deste grupo de variáveis foram calculadas as principais variáveis do modelo:

9. Tempo de exposição à gravidez
10. Idade atual
11. Duração da união

O tempo de exposição à gravidez, medido em meses, foi considerado como o tempo em que a mulher esteve unida nos 5 anos anteriores à entrevista⁹. Isto é, o tempo transcorrido desde a primeira união ou 60 meses, o que for menor. A idade atual se obteve como a diferença entre a data da entrevista e a data de nascimento e, a duração da união como a diferença entre a data da entrevista e a data da união¹⁰

⁷Não estava de forma direta na ENDES 91/92. Teve que ser calculada através das variáveis "nível de estudo atingido" e "número de anos de estudo no máximo nível atingido".

⁸Teve que ser calculado através das datas de nascimento dos filhos e a data da entrevista.

⁹Foi utilizado o período de 5 anos com o fim de obter informação sobre os níveis de fecundidade recentes, mas também para minimizar erros amostrais.

¹⁰Todas as datas estão expressas no código CMC (Century Month Code), número de meses transcorridos desde o começo do século.

Montagem do Arquivo Final

Cada registro do arquivo final está formado pelo conjunto de mulheres pertencentes a uma idade e duração de união determinada, assim como a um nível do fator classificatório. O número de filhos e o tempo de exposição à gravidez daquelas mulheres que compartilham a mesma idade, duração da primeira união e nível do fator classificatório são acumulados e constituem as variáveis *número de filhos* e *tempo de exposição* do arquivo final. Em síntese, cada registro do arquivo final terá cinco variáveis: nível do fator classificatório (tipo de união ou área de residência ou região de residência ou nível de instrução), idade (em anos), duração da união (em anos), tempo de exposição (em meses), e número de filhos das mulheres. Além destas cinco variáveis, foi considerada uma outra: o número de mulheres pertencentes a cada categoria, que porém não é necessária para o ajuste de modelo, mas de utilidade para fins de análises exploratórias¹¹

Este arquivo final é transformado em código ASCII e, desta maneira, fica pronto para o uso nos pacotes GLIM ou STATA, que são utilizados para justar o modelo.

4.3 Aplicação do Modelo

Para estimar os parâmetros do modelo log-Poisson foi utilizado o “software” GLIM (Generalised Linear Interactive Modelling), desenvolvido na sua primeira versão por Baker, Clarke e Nelder, no ano 1978.

O número de filhos é declarado como a variável independente, com distribuição Poisson. Declara-se também a função de ligação do tipo logarítmico. O logaritmo do tempo de exposição e o logaritmo do nível de fecundidade natural são somados para constituir o “offset”. O fator classificatório é definido como fator e a duração da união é declarada como variável preditora. No apêndice B se apresenta a lista de comandos utilizados para ajustar o modelo.

¹¹Para utilizar os procedimentos mencionados anteriormente, foi utilizado o sistema SAS (Statistical Analysis System). Os programas criados para este objetivo estão apresentados no apêndice B.

Como saída fornecida pelo programa tem-se a função de desvio, os graus de liberdade, os estimadores com os seus desvios padrões, os valores observados, os resíduos, os gráficos solicitados, etc.

De forma equivalente pode-se utilizar o “software” STATA. A lista de comandos para a definição do modelo com o uso deste pacote também é apresentada no apêndice B.

4.4 Resultados

Nesta seção se apresentam os resultados obtidos ao ajustar o modelo log-Poisson ao conjunto de dados especificados na seção anterior. Os parâmetros estimados e reexpressos como os índices de espaçamento dos nascimentos (I_α) e de limitação de nascimentos (I_β)¹², a bondade de ajuste do modelo (medida através do quociente desvio/graus de liberdade), assim como as taxas de fecundidade marital total estimadas $T\widehat{FMT}$, são apresentados nas tabelas 4.4, 4.5, 4.6, 4.7 e 4.8.

A partir destes resultados se realiza uma análise comparativa entre os distintos grupos de estudo, durante o período 1977-1992.

A medida resumo que se utiliza para detectar diferenciais e tendências do padrão de fecundidade é a $T\widehat{FMT}$ ¹³, apresentada nas tabelas mencionadas anteriormente. Os índices de espaçamento (I_α) e de limitação (I_β), são também utilizados para comparar os diferentes tipos de comportamento reprodutivo encontrado.

4.4.1 Resultados Gerais

O resultado mais geral indica que, no período de 15 anos estudado, a fecundidade marital no Peru reduziu-se em mais de 2 filhos por mulher (ver Tabela 4.4). A taxa de fecundidade marital total estimada ($T\widehat{FMT}$) para 1977/78 (9.6 filhos por

¹²Em inglês: *spacing* e *limiting index*, respectivamente.

¹³A $T\widehat{FMT}$ foi calculada acumulando as taxas específicas de fecundidade marital ajustadas. Para a idade a , a taxa específica ajustada é igual ao número de nascimentos ajustados correspondentes às mulheres de idade a , divididos entre o tempo de exposição das mulheres de idade a .

mulher) é reduzida a 7.3 filhos por mulher em 1992¹⁴. Isto significa que o Peru se situa, neste período, dentro de um processo de transição da fecundidade¹⁵. Esta transição é consequência clara do maior controle da fecundidade marital exercido pela população, cujos índices de limitação (I_f), que representam o controle da fecundidade após 10 anos de união, quase se triplicam no período (de 17.9 em 1977/78 a 58.8 em 1992). Por outro lado, os índices de espaçamento (I_o) permanecem quase constantes

Tabela 4.4: Índices de espaçamento, de limitação, bondade de ajuste, e $TF\widehat{M}T$, por tipo de união. Peru 1977-1992

		I_o	I_f	DESVIO	G.L.	$TF\widehat{M}T$
TOTAL	ENAF 77/78	20.8	17.9	571	714	9.6
	ENDES 86	25.9	31.6	517	600	8.0
	ENDES 91/92	5.7	58.8	676	795	7.3
FORMAL	ENAF 77/78	18.9	20.8	564	712	9.6
	ENDES 86	32.3	28.9	504	598	7.5
	ENDES 91/92	7.2	60.8	634	793	7.0
CONSENSUAL	ENAF 77/78	26.1	8.2	564	712	9.9
	ENDES 86	19.7	29.7	504	598	8.7
	ENDES 91/92	8.2	51.5	634	793	7.8

na primeira metade do período analisado, mas logo caem atingindo o valor de 5.7 em 1992. Este resultado estaria refletindo as gravidezes precoces, que em termos da $TF\widehat{M}T$ são compensadas com um maior controle ao longo da união (I_f igual a 58.8 em 1992) (ver Tabela 4.4).

Em relação à bondade de ajuste do modelo tem-se que o desvio dividido pelos graus de liberdade (D/ G.L.) variam entre 0.80 e 0.86, o que evidencia um bom ajuste do modelo aos dados¹⁶

¹⁴Rodriguez e Aravena (1991) utilizando o mesmo procedimento encontram $TF\widehat{M}T$ que diferem em nível, mas não na tendência (6.7 e 5.1 para 1977 e 1986, respectivamente.). Situação análoga se encontra nos índices I_o e I_f . As discordâncias podem-se dever à diferenças na montagem do arquivo final de dados. Estes resultados serão encaminhados aos autores citados, para discussão.

¹⁵No apêndice D se inclui um resumo da evolução da fecundidade no Peru a nível global, considerando todas as mulheres entre 15 e 49 anos, independente do estado conjugal.

¹⁶Lembrar que se o ajuste fosse perfeito este quociente seria 1.

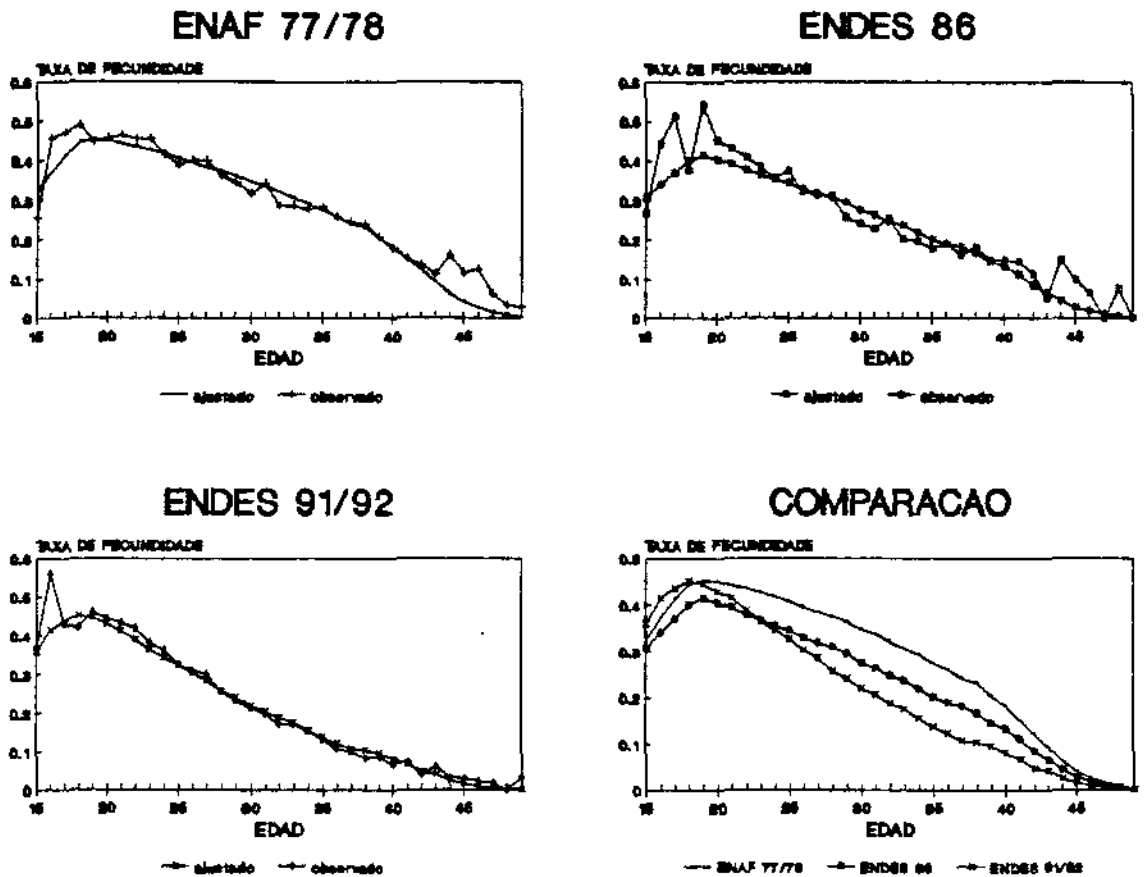


Figura 4.1: Taxas de fecundidade marital por idade. Peru 1977-1992.

Com o objetivo de se visualizar como o padrão de fecundidade marital mudou durante o período de análise, apresenta-se o Gráfico 4.1, correspondente aos valores observados e ajustados das taxas de fecundidade específicas por idade das mulheres da amostra. Este gráfico ilustra uma redução geral na fecundidade por idade no período 1977/1986. Já no período seguinte se faz evidente o aumento da fecundidade marital precoce assim como o aumento do controle da fecundidade tardia. Resultado análogo de observa ao considerar a fecundidade marital por união.

4.4.2 Efeito do Tipo de União

Um primeiro resultado, que pode ser observado na Tabela 4.4, é a existência de maiores taxas de fecundidade das uniões consensuais em relação as uniões formais. Ao nível nacional, as diferenças relativas entre as $T\widehat{FMT}$ das mulheres em união consensual e união formal, é de 3.1%, 16% e 11.4% para os anos 1977, 1986 e 1992, respectivamente. Note-se que na análise da fecundidade marital por tipo de união realizada para o Brasil no período 1976-1984, Lazo (1994) mostrou que as taxas de fecundidade marital das uniões formais também são menores que das uniões consensuais.

É interessante observar o comportamento diferenciado da queda da fecundidade segundo os dois tipos de união. Enquanto nas uniões formais há uma grande queda das taxas de fecundidade no período 1977-1986 (numa redução de 2.1 filhos por mulher), nas uniões consensuais esta mesma redução leva um tempo 50% vezes maior (de 1977 a 1992). Ou seja, o processo de transição da fecundidade se observa primeiro nas mulheres em união formal e logo a seguir, nas mulheres em união consensual.

No que diz respeito aos índices de limitação (I_{β}), se observa que, em geral, são mais altos nas uniões formais que nas uniões consensuais. (Ver Tabela 4.4). Este fato se reflete nas menores taxas de fecundidade marital para esse tipo de união, como foi mencionado acima.

Outro fato importante de ser destacado é o grande diferencial existente entre os índices de limitação (I_{β}) das uniões formais e consensuais em 1977/78, (20.8 e 8.9, respectivamente) quase desaparecem em 1986 e aumentam levemente até atingir uma

diferença de 18% em 1992. (Ver Tabela 4.4).

O índice de espaçamento (I_α), para ambos tipos de união, tem um padrão análogo ao observado para o Peru como todo, níveis altos e próximos no período 1977-1986 e um nível muito menor para o ano 1992.

Em relação ao ajuste do modelo, o quociente desvio/graus de liberdade (D/G.L.), continua sendo próximo de 1 (varia entre 0.79 e 0.84), portanto pode-se concluir que o ajuste do modelo é satisfatório.

4.4.3 Efeito da Área de Residência

Entre as duas áreas, urbano e rural, existem marcadas diferenças nos níveis de fecundidade. (Ver Tabela 4.5). No período analisado esta diferença variou entre 1.2 entre as mulheres em união consensual no ano 1992 ($T\widehat{FMT}$ de 7.3 na área urbana e 8.5 na área rural) e chega a 3.5 entre as mulheres urbanas e rurais em uniões formais no ano 1986 (com $T\widehat{FMT}$ de 6.2 e 9.7, respectivamente).

Outra diferença importante se observa entre os valores do índice de limitação (I_β) da área urbana e rural. Na área urbana estes índices são muito maiores que na área rural, em qualquer período e para qualquer tipo de união. Ao nível geral estas diferenças são muito maiores no período 1977-1986 (22.5 e 49.2, respectivamente), mas o grande aumento do índice I_β no último período para a área rural reduz esta diferença a 8.8 em 1992 (I_β igual a 62.2 e 53.4 para as áreas urbana e rural, respectivamente). Isto indica que o controle da fecundidade, com o fim de limitar o número de filhos aumentou muito na última década para o setor rural.

No período 1977-1986 o índice de espaçamento (I_α) foi sempre maior na área rural que na área urbana, com exceção das uniões consensuais, mantendo um nível relativamente próximo no período 1977-1986. No período 1986-1992 este índice decresce a patamares bem mais baixos, reflexo de uma fecundidade precoce, já citada por outros trabalhos como fenômeno característico dos anos 90. (Instituto de la Mujer-Espanha e FLASCO, 1994).

Considerando o tipo de união observa-se maiores diferenças na $T\widehat{FMT}$ dentro da área urbana. As uniões consensuais têm uma fecundidade relativa 4.6%, 25% e

Tabela 4.5: Índices de espaçamento, de limitação, bondade de ajuste, e $TF\widehat{MT}$, por área de residência e tipo de união. Peru 1977-1992

URBANO		I_α	I_β	DESVIO	G.L.	$TF\widehat{MT}$
TOTAL	ENAF 77/78	18.4	28.1	570	721	9.0
	ENDES 86	17.0	53.9	513	627	7.1
	ENDES 91/92	9.7	62.2	593	788	6.8
FORMAL	ENAF 77/78	16.1	31.2	487	674	8.7
	ENDES 86	17.3	56.3	464	557	6.2
	ENDES 91/92	11.7	63.8	519	723	6.3
CONSENSUAL	ENAF 77/78	25.6	16.0	338	458	9.1
	ENDES 86	22.9	38.2	252	339	7.8
	ENDES 91/92	11.3	54.9	412	585	7.3
RURAL						
TOTAL	ENAF 77/78	20.7	5.6	570	721	10.8
	ENDES 86	27.8	4.7	513	627	10.0
	ENDES 91/92	-3.4	53.4	592	788	8.2
FORMAL	ENAF 77/78	18.2	7.7	487	674	10.9
	ENDES 86	37.4	-9.0	464	557	9.7
	ENDES 91/92	-7.3	56.2	519	720	8.1
CONSENSUAL	ENAF 77/78	24.6	1.8	338	458	10.7
	ENDES 86	11.6	25.9	256	339	9.6
	ENDES 91/92	2.4	47.4	412	585	8.5

15% maiores que as uniões formais para os anos 1977, 1986 e 1992, respectivamente, enquanto que na área rural essas diferenças são de apenas -1.8%, -1.0% e 4.9% para os mesmos anos (ver Tabela 4.5). Por outro lado, tem-se que a partir do ano 1986 as uniões consensuais tendem a se aproximar às uniões formais no que se refere aos níveis da $TF\widehat{MT}$.

Na classificação por áreas de residência o ajuste do modelo decresce em relação às análises anteriores, mas ainda atinge um bom nível. O quociente desvio/graus de liberdade (D / G.L.) varia entre 0.7 e 0.83.

4.4.4 Efeito da Região de Residência

Considerando o efeito da região de residência, pode-se observar uma situação análoga à observada na análise do efeito urbano-rural. (Ver Tabela 4.6). Existe uma grande diferença entre os níveis de fecundidade marital na região Lima Metropolitana e nas regiões Selva e Serra, ficando a região Resto da Costa numa situação intermediária.

Na região Lima Metropolitana, a partir do ano 1986, a taxas de fecundidade marital se mantêm relativamente baixas e estáveis (ao redor de 6.3 filhos por mulher), e os índices de limitação (I_f) são relativamente altos e constantes (ao redor de 65). Este fato pode ser consequência da enorme centralização existente no Peru, onde Lima, seguida de algumas outras cidades da Costa, possuem as melhores condições de vida do país.

Nas regiões Serra e Selva se observam as maiores taxas de fecundidade marital do Peru no ano 1986, mas por outro lado, é nestas regiões onde o processo de transição da fecundidade se acelera no período analisado. Cabe destacar a redução da $T\widehat{FMT}$ na região da Selva no período 1986-1992 em 1.7 filhos por mulher, independente, inclusive, do tipo de união. Já a velocidade da transição da região Serra se dá mais lentamente, igualando em ritmo à região do Resto da Costa. Em ambas regiões há uma redução nas $T\widehat{FMT}$ de quase um filho por mulher, atingindo uma $T\widehat{FMT}$ de 8.0 e 7.0 filhos por mulher nas regiões Serra e Resto da Costa, respectivamente. (Ver Tabela 4.6).

A queda dos índices de espaçamento (I_a) é marcante no período 1986-1992 para todas as regiões (com exceção de Lima Metropolitana), o que levaria a concluir que na diminuição das $T\widehat{FMT}$ as condutas dirigidas a limitar os nascimentos (relacionadas com o índice I_f) estão assumindo mais intensidade que aquelas de espaçar os nascimentos (relacionadas com o índice I_a). As maiores quedas se observam na região Serra, onde o índice de espaçamento chega a diminuir em 93%. Na região Lima só se observam índices de espaçamento (I_a) relativamente próximos no período 1986-1992.

No ano 1977, as regiões consideradas pela ENAF 77/78 foram as chamadas "regiões de planejamento" e não as regiões geográficas ou naturais, que foram citadas

Tabela 4.6: Índices de espaçamento, de limitação, bondade de ajuste, e $TFMT$, por região de residência e tipo de união. Peru 1977-1992

LIMA METROP.		I_α	I_β	DESVIO	G.L.	$TFMT$
TOTAL	ENAF 77/78	13.7	40.7	1052	1413	8.0
	ENDES 86	11.8	64.4	797	967	6.3
	ENDES 91/92	14.6	66.9	1040	1407	6.3
FORMAL	ENAF 77/78	10.5	45.5	904	1239	8.0
	ENDES 86	15.6	64.6	668	807	4.9
	ENDES 91/92	13.7	67.8	926	1276	5.9
CONSENSUAL	ENAF 77/78	28.8	11.2	483	663	8.5
	ENDES 86	8.8	58.6	306	424	6.9
	ENDES 91/92	18.2	62.3	694	880	6.3
RESTO COSTA						
TOTAL	ENDES 86	20.7	38.5	797	967	8.0
	ENDES 91/92	3.2	64.2	1040	1407	7.0
FORMAL	ENDES 86	17.2	43.0	668	807	7.5
	ENDES 91/92	3.9	66.7	926	1276	6.8
CONSENSUAL	ENDES 86	27.6	25.3	306	424	7.4
	ENDES 91/92	6.3	56.2	594	880	7.6
SERRA						
TOTAL	ENDES 86	32.2	9.9	797	967	8.9
	ENDES 91/92	-2.2	57.0	1040	1407	8.0
FORMAL	ENDES 86	42.1	-1.1	668	807	7.9
	ENDES 91/92	-4.9	59.4	926	1276	7.4
CONSENSUAL	ENDES 86	16.2	25.9	306	424	9.1
	ENDES 91/92	6.2	47.5	594	880	8.0
SELVA						
TOTAL	ENDES 86	18.6	21.8	797	967	9.0
	ENDES 91/92	9.1	53.4	1040	1407	7.3
FORMAL	ENDES 86	26.6	19.9	668	807	8.4
	ENDES 91/92	13.4	54.9	926	1276	6.8
CONSENSUAL	ENDES 86	13.0	19.4	306	424	9.1
	ENDES 91/92	9.0	48.9	594	880	7.6

Tabela 4.7: Índices de espaçamento, de limitação, bondade de ajuste, e $TF\widehat{MT}$, por região de residência e tipo de união. Peru 1977.

		I_{α}	I_{β}	DESVIO	G.L.	$TF\widehat{MT}$
	TOTAL	20.5	15.7	1052	1413	9.9
CENTRO	FORMAL	16.6	18.8	904	1239	9.8
	CONSENSUAL	29.2	7.3	483	663	9.4
	TOTAL	19.3	9.2	1052	1413	10.7
NORTE	FORMAL	22.2	6.38	904	1239	10.2
	CONSENSUAL	10.9	17.9	483	663	10.7
	TOTAL	23.8	9.3	1052	1413	10.1
SUR	FORMAL	26.3	6.9	904	1239	9.6
	CONSENSUAL	18.0	15.2	483	663	9.9
	TOTAL	23.4	12.9	1052	1413	9.5
ORIENTE	FORMAL	10.7	27.7	904	1239	9.3
	CONSENSUAL	33.2	-2.5	483	663	9.0

anteriormente. Este sistema de divisão gera cinco regiões: Lima Metropolitana, Norte, Centro, Sul e Oriente. (Ver mapa, apêndice B). Cabe destacar que o desenvolvimento socio-econômico das regiões Sul e Centro é baixo, mas a região Sul se destaca pelos centros urbanos que possui (Ortiz, 1987).

As regiões consideradas pela ENAF 77/78 mostram, em geral, taxas de fecundidade marital altas e índices de limitação (I_{β}) baixos, em todas as regiões. (Ver Tabela 4.7). Por outro lado, os índices de espaçamento (I_{α}) no ano 1977/78 são relativamente altos, inclusive considerando-se o tipo de união, a área de residência, e o nível de instrução. Estes resultados mostram uma grande homogeneidade nacional, a qual se rompe nos anos 80 com o início de uma transição de fecundidade no Peru, diferencial por regiões.

Em Lima Metropolitana existe uma grande diferença entre as $TF\widehat{MT}$ das uniões consensuais e das uniões formais no ano 1986 (a $TF\widehat{MT}$ correspondente às uniões consensuais supera em 40% à $TF\widehat{MT}$ das uniões formais). Analisando o período 1986 - 1992 para o resto das regiões de estudo se observa que as $TF\widehat{MT}$

das uniões consensuais são maiores que as $TF\widehat{MT}$ das uniões formais (entre 8% e 15%), com uma única exceção registrada na região Resto da Costa no ano 1986, onde a $TF\widehat{MT}$ das uniões consensuais é 1.3% menor que das uniões formais.

No referente ao ajuste do modelo, tem-se que o quociente desvio/graus de liberdade ($D / G.L.$) ainda pertence a níveis razoáveis, variando entre 0.72 e 0.83.

4.4.5 Efeito da Educação

Os resultados correspondentes ao efeito do nível de instrução corroboram o padrão geral mencionado anteriormente: as mulheres com maior nível de instrução têm menores taxas de fecundidade que aquelas com menor educação (ver Tabela 4.8).

A diferença nas taxas de fecundidade marital entre as mulheres que têm até 2 anos de estudo e aquelas que têm 9 ou mais, chega até 3 filhos por mulher, no ano 1986, sendo este o maior diferencial encontrado entre os grupos de instrução estudados.

Também pode-se observar que quando o nível de educação da mulher aumenta, o controle voluntário da fecundidade (I_{β}) também aumenta, e atinge níveis entre 60 e 70. (Ver Tabela 4.8).

Observando-se o comportamento do índice de espaçamento (I_{α}), encontra-se um resultado interessante: o comportamento deste índice é muito diferente no grupo de mulheres com mais de 9 anos de estudo que nos outros grupos de instrução. Para as mulheres com mais instrução o índice de espaçamento aumenta ao longo de todo o período analisado. Isto não ocorre com nenhum dos outros grupos estudados, onde índice I_{α} decresce notavelmente no período, e seu maior decréscimo se observa entre 1986 e 1992.

Isto indicaria que o fenômeno de aumento das gravidezes precoces dos anos 90, não atingiria às mulheres de maior nível de instrução. Este seria um fenômeno mais relacionado às regiões de menor desenvolvimento nas quais se concentra uma maior proporção de população de baixo nível de instrução.

Estes resultados evidenciam que a transição da fecundidade teria-se iniciado nos grupos de maior instrução que, no começo do período analisado, já apresentava

Tabela 4.8: Índices de espaçamento, de limitação, bondade de ajuste, e $T\widehat{FMT}$, por educação e tipo de união. Peru 1977-1992.

EDUCAÇÃO 0-2		I_a	I_f	DESVIO	G.L.	$T\widehat{FMT}$
TOTAL	ENAF 77/78	22.8	5.1	690	949	10.6
	ENDES 86	25.8	9.1	596	790	9.8
	ENDES 91/92	-21.9	58.7	802	1070	8.5
FORMAL	ENAF 77/78	17.1	9.7	629	883	10.5
	ENDES 86	34.1	0.1	526	674	9.2
	ENDES 91/92	-36.8	63.4	701	923	7.3
CONSENSUAL	ENAF 77/78	30.5	-3.0	348	493	10.3
	ENDES 86	14.9	22.3	291	379	9.6
	ENDES 91/92	-6.3	50.6	523	733	8.7
EDUCAÇÃO 3-8						
TOTAL	ENAF 77/78	12.5	32.0	690	949	9.2
	ENDES 86	13.5	42.7	596	790	8.0
	ENDES 91/92	-4.8	59.0	802	1070	7.6
FORMAL	ENAF 77/78	11.0	34.2	629	883	9.1
	ENDES 86	12.1	44.2	526	674	8.0
	ENDES 91/92	-6.9	61.3	701	923	7.4
CONSENSUAL	ENAF 77/78	16.8	24.5	348	493	9.2
	ENDES 86	16.4	38.6	291	379	8.2
	ENDES 91/92	-0.5	54.4	523	733	8.0
EDUCAÇÃO 9 +						
TOTAL	ENAF 77/78	3.0	57.3	690	949	8.6
	ENDES 86	6.2	66.5	596	790	6.8
	ENDES 91/92	11.2	65.5	802	1070	6.9
FORMAL	ENAF 77/78	1.2	57.5	629	883	7.8
	ENDES 86	7.9	68.1	526	674	6.0
	ENDES 91/92	10.5	66.8	701	923	6.8
CONSENSUAL	ENAF 77/78	11.8	61.8	348	493	6.2
	ENDES 86	12.8	48.4	281	379	7.4
	ENDES 91/92	16.3	57.9	523	733	7.2

menores $T\widehat{FMT}$. Este processo continua até meados da década de 80, momento em que a $T\widehat{FMT}$ se estabiliza ao redor de 6.8 filhos por mulher. No extremo oposto se encontra a fecundidade das mulheres que têm menos de 2 anos de educação, cuja transição teria começado somente nos anos 80, diminuindo a taxa de fecundidade marital em 1.3 filhos no período 1986-1992. O grupo com 3 a 8 anos de estudo se encontra, aparentemente, no meio do processo de transição, com uma redução da $T\widehat{FMT}$ de 0.8 filhos no período 1977-1986 e de 0.4 filhos no período 1986-1992. Isto corresponde a uma $T\widehat{FMT}$ de 7.6 filhos por mulher em 1992, um filho a mais dos apresentados pelas mulheres com 9 ou mais anos de estudo e um filho a menos das mulheres com menos de 2 anos de estudo, no ano 1992. (Ver Tabel 4.8).

Analisando o tipo de união dentro por nível de educação, observa-se uma queda da $T\widehat{FMT}$ dentro das uniões formais para o grupo de menor educação (redução de 43.8% no período 1977/1992), mas a queda observada nas uniões consensuais é consideravelmente menor (1.8% no mesmo período). Por outro lado, no nível de educação intermediário, tanto nas uniões formais como nas uniões consensuais, se evidencia uma redução das $T\widehat{FMT}$ mais próximos (22% e 15% no período 1977/1992, respectivamente). Já no grupo de maior educação se observam comportamentos opostos nos dois tipos de união. Nas uniões formais a $T\widehat{FMT}$ decresce em 14% enquanto nas uniões consensuais a $T\widehat{FMT}$ cresce em 16%.

Para finalizar, pode-se observar que o ajuste deste modelo foi satisfatório, assim como nas análises anteriores. O quociente desvio/graus de liberdade ($D / G.L.$) assume valores entre 0.70 e 0.78.

4.5 Comentários

Nas seções anteriores foi realizada uma análise do efeito dos fatores socioeconômicos na fecundidade marital no Peru. Estes efeitos foram discutidos individualmente mas certamente os fatores considerados estão fortemente relacionados entre si.

Assim, as áreas urbanas apresentam uma grande proporção de mulheres com

maior nível de educação, ao contrário do que ocorre na área rural, onde a porcentagem de mulheres com 9 ou mais anos de instrução é muito menor. No ano 1986, 50% das mulheres urbanas entrevistadas tinham 9 ou mais anos de estudo, enquanto só 1% das mulheres rurais pertencia a este nível de instrução¹⁷.

Cabe destacar que a estrutura do tipo de união independe da área de residência analisada. A proporção de mulheres em união formal é de 70% e o restante está em união consensual, tanto para a área urbana como para a área rural.¹⁸

Outra interação importante entre os fatores analisados é a relação entre regiões e grau de urbanização. Lima, por ser área metropolitana, tem 100% de área urbana. O Resto da Costa tem 74% de área urbana e a Serra e Selva tem 33% e 39% de área urbana, respectivamente.¹⁹

Este tipo de interações se evidenciam nos resultados apresentados, quando se observam as semelhanças entre os padrões dos índices de espaçamento, limitação, assim como nas taxas de fecundidade marital na região Lima, na área urbana e no grupo de maior nível de educação. Nestes grupos a transição da fecundidade começou mais cedo que nos outros grupos analisados, onde ainda os níveis de fecundidade marital estão em processo de transição e continuam em patamares elevados.

¹⁷Fonte: ENDES 86.

¹⁸Fonte: ENDES 86.

¹⁹Fonte: ENDES 86.

Capítulo 5

CONSIDERAÇÕES FINAIS

O objetivo principal desta dissertação foi apresentar a utilidade de um tópico da metodologia estatística pouco explorado, como são os modelos lineares generalizados, para analisar um importante fenômeno demográfico como é a fecundidade marital de uma população.

Foi difícil para uma estatística, submersa em números e teorias analíticas se introduzir no mundo das ciências sociais, da mesma forma que para um cientista social deve ser difícil se familiarizar com cálculos e teorias facilmente compreendidas por aqueles que trabalham nas ciências exatas. Pode ser difícil, mas vale a pena o esforço, pois na junção destas duas áreas de trabalho surgem resultados que têm a característica de serem “confiáveis” na metodologia utilizada e que podem ser interpretados no mesmo contexto ao qual pertence o fenômeno em estudo.

Acredita-se que este fato foi percebido no modelo final apresentado, o modelo de Rodriguez e Cleland (1988), onde a metodologia dos modelos lineares generalizados foi utilizada para analisar a fecundidade marital. Os parâmetros estimados, que possuem “credibilidade” estatística, dada pela bondade de ajuste do modelo, são interpretados como aspectos relacionados como o nível de fecundidade natural e o grau de controle desta.

Por último, pode-se mencionar a importância da disponibilidade dos dados utilizados. São necessárias pesquisas e levantamentos de dados bem planejados, de grande porte e de boa qualidade que forneçam esta informação. Este trabalho so-

mente foi possível utilizando os levantamentos da World Fertility Survey -WFS e da Demographic and Health Survey-DHS, já mencionados. Esta é uma necessidade primordial nos países em vias de desenvolvimento, onde a compreensão da realidade do país pode levar a conseguir soluções adequadas para problemas que possuem.

A autora desta tese ficaria muito satisfeita se contribuiu de alguma forma a mostrar a importância que tem relacionar a ciência estatística com a ciência demográfica. Acredita-se que é nesta interação com outras áreas da ciência que a estatística assume a sua verdadeira dimensão, como ciência ao serviço da humanidade

Apêndice A

Teoria Estatística

A1) Descrição de uma Variável Aleatória

Definição 1: Variável Aleatória Discreta

Diz-se que a variável aleatória X é *discreta* se assume valores dentro de um conjunto enumerável $\aleph = \{x_1, x_2, \dots\}$. A distribuição de probabilidades de X é dada pelos números

$$P_X(x_j) = P(X = x_j)$$

que representam a probabilidade da variável aleatória X assumir o valor x_j . A função $P_X(x_j)$ definida em \aleph , é chamada a *função de probabilidade* de X e satisfaz

$$\sum_{j=1}^{\infty} P_X(x_j) = P_X(\aleph) = 1 \quad \text{e} \quad P_X(x_j) \geq 0, \quad \forall j.$$

Definição 2: Variável Aleatória Contínua

Diz-se que a variável aleatória X é *contínua* se assume valores pertencentes a um intervalo I subconjunto dos números reais (R). A função de densidade de X é dada por uma função $f_X(x)$, não negativa e integrável em R , tal que

$$P_X(A) = \int_A f_X(x) \partial x$$

representa a probabilidade da variável aleatória X assumir valores pertencentes ao

intervalo A . A função $f_X(x)$ definida em R é chamada a *função de densidade* de X e satisfaz

$$\int_{-\infty}^{+\infty} f_X(x) dx = P_X(R) = 1.$$

Definição 3: Função de Distribuição

A *função de distribuição de probabilidade acumulada* da variável aleatória X em Ω (o conjunto de valores possíveis de X , chamado *espaço amostral*), denotada por $F_X(x)$, é definida como

$$F_X(x) = P_X(X \leq x), \quad x \in R.$$

É importante destacar que existe uma relação biunívoca entre a função de distribuição acumulada da variável aleatória X e a sua distribuição de probabilidades (se X for discreta) ou função de densidade (se X for contínua).

A2) Estatísticas suficientes para os parâmetros dos modelos lineares generalizados com funções de ligação canônicas.

Teorema 1: Critério de Fatoração de Neymann-Pearson

Seja (X_1, X_2, \dots, X_n) um vetor aleatório com função de probabilidade $p_\theta(x_1, x_2, \dots, x_n) = P_\theta(X_1 = x_1, X_2 = x_2, \dots, X_n = x_n)$ ou função de densidade $f_\theta(x_1, x_2, \dots, x_n)$ onde θ (vetor de parâmetros) $\in \Theta$ (espaço paramétrico), e a forma de p_θ ou f_θ é conhecida, exceto para o valor numérico específico de θ .

A estatística $T = T(X_1, X_2, \dots, X_n)$ é suficiente para θ se e somente se p_θ ou f_θ podem ser fatoradas da seguinte forma

$$p_\theta(x_1, x_2, \dots, x_n) = h(x_1, x_2, \dots, x_n) g_\theta(T(x_1, x_2, \dots, x_n))$$

ou

$$f_\theta(x_1, x_2, \dots, x_n) = h(x_1, x_2, \dots, x_n) g_\theta(T(x_1, x_2, \dots, x_n))$$

onde a função h depende do vetor de observações (x_1, x_2, \dots, x_n) e não de θ , e g_θ depende de θ e (x_1, x_2, \dots, x_n) somente através de $T(x_1, x_2, \dots, x_n)$.

A demonstração deste teorema pode ser encontrada em Rohatghi(1984, p. 645), por exemplo.

Nos modelos lineares generalizados, definidos no capítulo 1, com função de ligação canônica, as estatísticas $S_j = \sum_{i=1}^n \alpha(\phi_i) x_{ij} Y_i$, para $j = 1, \dots, p$, são suficientes e de dimensão p para o vetor de parâmetros β .

Sejam $s_j = \sum_{i=1}^n \alpha(\phi_i) x_{ij} y_i$ realizações da variável aleatória $S_j = \sum_{i=1}^n \alpha(\phi_i) x_{ij} Y_i$, para $j = 1, \dots, p$.

A função de log-verossimilhança $\ell(\theta, \phi; y)$, definida em (1.4) pode ser escrita como $\ell(\theta, \phi; y) = L_1 + L_2$, sendo L_1 uma função de β e (y_1, y_2, \dots, y_n) somente através de (s_1, s_2, \dots, s_n) ; e L_2 que depende somente de (y_1, y_2, \dots, y_n) .

Demonstração:

$$\begin{aligned} \ell(\theta, \phi; y) &= \sum_i \{ \alpha(\phi_i) [y_i \theta_i - g(\theta_i)] + c(y_i, \phi_i) \} \\ &= \sum_i \alpha(\phi_i) y_i \theta_i - \sum_i \alpha(\phi_i) g(\theta_i) + \sum_i c(y_i, \phi_i) \end{aligned}$$

Mas como a função de ligação é canônica, i.e., $\theta_i = \eta_i$, tem-se:

$$\begin{aligned} \ell(\theta, \phi; y) &= \sum_i \alpha(\phi_i) y_i \eta_i - \sum_i \alpha(\phi_i) g(\eta_i) + \sum_i c(y_i, \phi_i) \\ &= \sum_i \alpha(\phi_i) y_i \sum_j x_{ij} \beta_j - \sum_i \alpha(\phi_i) g(\sum_j x_{ij} \beta_j) + \sum_i c(y_i, \phi_i) \\ &= \sum_j \left(\sum_i \alpha(\phi_i) y_i x_{ij} \right) \beta_j - \sum_i \alpha(\phi_i) g(\sum_j x_{ij} \beta_j) + \sum_i c(y_i, \phi_i) \\ &= \underbrace{\sum_j s_j \beta_j - \sum_i \alpha(\phi_i) g(\sum_j x_{ij} \beta_j)}_{L_1} + \underbrace{\sum_i c(y_i, \phi_i)}_{L_2} \end{aligned}$$

Do exposto acima e pelo teorema do Critério de Fatoração de Neymann-Pearson, tem-se que as estatísticas S_j são suficientes para os parâmetros β_j , $j = 1, \dots, p$.

A3) Resultados importantes sobre o valor esperado das derivadas da função de log-verossimilhança das distribuições pertencentes a família exponencial

As distribuições $f(y, \theta)$, com $\theta \in \Theta$, pertencentes à família exponencial satisfazem as condições de regularidade:

- i) $\frac{\partial}{\partial \theta} \log(f(y, \theta)) \quad \forall y, \theta$
- ii) $\frac{\partial}{\partial \theta} \int \dots \int \prod_{i=1}^n f(y, \theta) \partial y_1 \dots \partial y_n = \int \dots \int \frac{\partial}{\partial \theta} \prod_{i=1}^n f(y, \theta) \partial y_1 \dots \partial y_n$
- iii) $0 < E_{\theta} \left[\frac{\partial}{\partial \theta} \log(f(y, \theta)) \right]^2 < \infty \quad \forall \theta \in \Theta$

Com estes resultados tem-se:

$$1) \quad E \left(\frac{\partial \ell}{\partial \theta} \right) = 0$$

Prova:

$$\begin{aligned} E \left(\frac{\partial \ell}{\partial \theta} \right) &= E \left(\frac{\partial \log(f(y, \theta))}{\partial \theta} \right) = \int \frac{\partial \log(f(y, \theta))}{\partial \theta} f(y, \theta) \partial y = \int \frac{\frac{\partial}{\partial \theta} f(y, \theta)}{f(y, \theta)} f(y, \theta) \partial y \\ &= \frac{\partial}{\partial \theta} \int f(y, \theta) \partial y = \frac{\partial}{\partial \theta} (1) = 0 \end{aligned}$$

□

$$2) \quad E \left(\frac{\partial^2 \ell}{\partial \theta^2} \right) = -E \left(\frac{\partial \ell}{\partial \theta} \right)^2$$

Prova:

$$\begin{aligned} E \left(\frac{\partial^2 \ell}{\partial \theta^2} \right) &= E \left(\frac{\partial}{\partial \theta} \left(\frac{\partial \ell}{\partial \theta} \right) \right) = E \left(\frac{\partial}{\partial \theta} \left(\frac{\partial \log(f(y, \theta))}{\partial \theta} \right) \right) = E \left(\frac{\partial}{\partial \theta} \left(\frac{\frac{\partial}{\partial \theta} f(y, \theta)}{f(y, \theta)} \right) \right) \\ &= E \left[\frac{\left(\frac{\partial^2}{\partial \theta^2} f(y, \theta) \right) f(y, \theta) - \left(\frac{\partial}{\partial \theta} f(y, \theta) \right) \left(\frac{\partial}{\partial \theta} f(y, \theta) \right)}{(f(y, \theta))^2} \right] \end{aligned}$$

$$\begin{aligned}
&= E \left[\frac{\frac{\partial^2}{\partial \theta^2} f(y, \theta)}{f(y, \theta)} \right] - E \left[\frac{\frac{\partial}{\partial \theta} f(y, \theta)}{f(y, \theta)} \right]^2 \\
&= \int \frac{\frac{\partial^2}{\partial \theta^2} f(y, \theta)}{f(y, \theta)} f(y, \theta) dy - E \left[\frac{\partial \log(f(y, \theta))}{\partial \theta} \right]^2 \\
&= \frac{\partial^2}{\partial \theta^2} \int f(y, \theta) dy - E \left[\frac{\partial \log(f(y, \theta))}{\partial \theta} \right]^2 \\
&= \frac{\partial^2}{\partial \theta^2} (1) - E \left(\frac{\partial \ell}{\partial \theta} \right)^2 = -E \left(\frac{\partial \ell}{\partial \theta} \right)^2
\end{aligned}$$

□

A4) Mínimos Quadrados Ponderados

Método de Mínimos Quadrados Ordinários

O método de mínimos quadrados é utilizado para ajustar um modelo a um conjunto de dados e se baseia no princípio de minimizar a soma dos quadrados dos desvios dos valores observados aos valores ajustados pelo modelo..

Suponha-se o caso de uma regressão linear simples, ou seja, que se está trabalhando com o conjunto de dados $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$. Supondo uma relação linear entre y e x , pode-se propor o seguinte modelo de regressão

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad \text{para } i = 1, 2, \dots, n,$$

onde se assume $E(\epsilon) = 0$, $Var(\epsilon) = \sigma^2 I$, com $\epsilon = (\epsilon_1, \epsilon_2, \dots, \epsilon_n)$, $0 = (0, \dots, 0)$ vetores de dimensão $p \times 1$, e I matriz identidade de dimensão $n \times n$.

A soma de quadrados dos erros considerados pelo modelo é:

$$S = \sum_{i=1}^n \epsilon_i^2 = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2.$$

Os valores de β_0 e β_1 que minimizam S são dados por

$$\hat{\beta}_0 = \frac{\sum x_i y_i - n \bar{x} \bar{y}}{\sum x_i^2 - n \bar{x}^2}$$

e

$$\hat{\beta}_1 = \bar{y} - \beta_0 \bar{x}$$

Em geral, pode-se propor o modelo $y = X\beta + \epsilon$, $E(\epsilon) = 0$ e $Var(\epsilon) = \sigma^2 I$. Se a matriz do modelo X é de posto completo, isto é, se nenhuma de suas colunas é combinação linear das outras, pode ser mostrado que o vetor de parâmetros estimado pelo método de mínimos quadrados é dado por

$$\hat{\beta} = (X^T X)^{-1} X^T y,$$

solução do sistema de equações

$$X^T X \beta = X^T y,$$

onde X^T representa a transposta da matriz X e y o vetor de observações da variável resposta Y .

Mínimos Quadrados Ponderados

No caso anterior assume-se que para o vetor de erros ϵ , $E(\epsilon) = 0$ e $Var(\epsilon) = I\sigma^2$. Porém, a condição de variância constante para todas as componentes ϵ_i de ϵ , não é sempre satisfeita. Em algumas situações, as observações não provem de distribuições com a mesma variância e portanto a matriz de variância-covariância dos erros não será mais $\sigma^2 I$, mas sim tomará a forma de uma matriz não singular diagonal. Se as observações são correlacionadas a matriz de variância-covariância terá elementos diferentes de zero fora da diagonal.

Denotando a matriz de variância-covariância como W , pode se mostrar que o estimador do vetor de parâmetros está dado por:

$$\hat{\beta} = (X^T W^{-1} X)^{-1} X^T W^{-1} y$$

solução do sistema de equações

$$X^T W^{-1} X \beta = X^T W^{-1} y$$

Maiores detalhes sobre estes métodos, podem ser encontrados em Draper e Smith(1981), por exemplo.

A5) Cálculo do valor de c a ser somado às observações, no caso da função de ligação não estar definida para todos os valores possíveis dos dados.

Expressando $f(y + c)$ em série de Taylor ao redor de μ , tem-se:

$$f(y + c) \cong f(\mu) + (y + c - \mu)f'(\mu) + \frac{(y + c - \mu)^2 f''(\mu)}{2}.$$

Desenvolvendo esta expressão obtém-se:

$$\begin{aligned} f(y + c) \cong f(\mu) + yf'(\mu) + cf'(\mu) - \mu f'(\mu) + \frac{y^2 f''(\mu)}{2} + ycf''(\mu) + \frac{c^2 f''(\mu)}{2} \\ - y\mu f''(\mu) - c\mu f''(\mu) + \frac{\mu^2 f''(\mu)}{2}. \end{aligned}$$

Omitindo-se os termos com c^2 e tomando o valor esperado em ambos os lados, lembrando que $E(y) = \mu$ e $E(y^2) = Var(y) + \mu^2$, tem-se:

$$\begin{aligned} E(f(y + c)) = f(\mu) + \mu f'(\mu) + cf'(\mu) - \mu f'(\mu) + \frac{Var(y)f''(\mu)}{2} + \frac{\mu^2 f''(\mu)}{2} \\ + \mu cf''(\mu) - \mu^2 f''(\mu) - c\mu f''(\mu) + \frac{\mu^2 f''(\mu)}{2}. \end{aligned}$$

Então,

$$E(f(y + c)) = f(\mu) + cf'(\mu) + \frac{Var(y)f''(\mu)}{2}. \quad (A.1)$$

Precisa-se que $E(f(y + c))$ seja igual à $f(\mu)$, então, substituindo em (A.1), tem-se:

$$f(\mu) = f(\mu) + cf'(\mu) + \frac{Var(y)f''(\mu)}{2},$$

portanto,

$$c = -\frac{Var(y)f''(\mu)}{2f'(\mu)}. \quad (A.2)$$

No modelo log-Poisson $f(\mu) = \log(\mu)$, então $f'(\mu) = \frac{1}{\mu}$ e $f''(\mu) = -\frac{1}{\mu^2}$. Além disso, $Var(y) = \mu$, substituindo estes resultados em (A.2), obtem-se

$$c = \frac{-\mu \left(-\frac{1}{\mu^2}\right)}{2 \left(\frac{1}{\mu}\right)} = \frac{1}{2}.$$

□

A6 Distribuição de $\ell(\hat{\beta}, y) - \ell(\beta, y)$.

Suponha-se que o modelo de interesse envolve m parâmetros, denotados pelo vetor de parâmetros β .

Expandindo $\ell(\beta, y)$, função de verossimilhança do modelo, em séries de Taylor ao redor do estimador de máxima verossimilhança $\hat{\beta}$ de β :

$$\ell(\beta, y) \cong \ell(\hat{\beta}, y) + (\beta - \hat{\beta})^T U(\hat{\beta}) + \frac{1}{2}(\beta - \hat{\beta})^T H(\hat{\beta})(\beta - \hat{\beta}) \quad (A.3)$$

onde $U(\hat{\beta})$ é o vetor de derivadas $\frac{\partial \ell}{\partial \beta_i}$ avaliado em $\hat{\beta}$ e $H(\hat{\beta})$ matriz de segundas derivadas $\frac{\partial^2 \ell}{\partial \beta_i \partial \beta_j}$ avaliada em $\hat{\beta}$.

Como $\hat{\beta}$ é o estimador de máxima verossimilhança de β , $U(\hat{\beta}) = 0$.

Além disso, para grandes amostras pode-se aproximar $-H(\hat{\beta})$ por K , onde $K_{ij} = E \left(-\frac{\partial^2 \ell}{\partial \beta_i \partial \beta_j} \right)$.

Do dito acima, tem-se que (A.3) pode ser escrita como:

$$\ell(\hat{\beta}, y) - \ell(\beta, y) = \frac{1}{2}(\beta - \hat{\beta})^T K(\beta - \hat{\beta})$$

É sabido que, para grandes amostras, $(\beta - \hat{\beta}) \sim \text{Normal}(0, K^{-1})$, ou, equivalentemente, $(\beta - \hat{\beta})^T K(\beta - \hat{\beta})$, às vezes chamada estatística de Wald, tem uma distribuição X_m^2 .

Daqui, tem-se que

$$2 \left(\ell(\hat{\beta}, y) - \ell(\beta, y) \right) \sim X_m^2$$

B1) Função de distribuição resultante ao remover a componente exponencial convolucionada.

Seja $F(x)$ convolução de $F_1(x)$ e a distribuição exponencial (r).

Da definição de convolução,

$$\begin{aligned} F(x) &= \int_0^x F_1(t) r e^{-r(x-t)} dt \\ &= r e^{-rx} \int_0^x F_1(t) e^{rt} dt \end{aligned}$$

Derivando ambos lados com respeito a x :

$$\begin{aligned} F'(x) &= -r^2 e^{-rx} \int_0^x F_1(t) e^{rt} dt + r e^{-rx} F_1(x) e^{rx} \\ &= -r \int_0^x F_1(t) r e^{-r(x-t)} dt + r F_1(x) \\ &= -r F(x) + r F_1(x) \end{aligned}$$

Daqui,

$$F_1(x) = F(x) + \frac{F'(x)}{r}$$

□

Apêndice B

Aspectos Computacionais

Programas SAS para extrair a informação desejada do conjunto de dados originais

Os programas SAS que foram utilizados para extrair informação dos arquivos originais são apresentados a seguir (Programas 1, 2 e 3), para cada uma das pesquisas analisadas. A estrutura dos programas é a mesma, mas existem diferenças nos nomes que assumem as variáveis e as seções, assim como na locação destas. Os nomes das variáveis, seções, o seu significado e locação, além dos valores possíveis das respostas e como estas são identificadas, estão presentes nos arquivos com extensão .MAP. Os dados se encontram nos arquivos com extensão .DAT.

Programa SAS de seleção de variáveis e população alvo

A partir desta primeira seleção de variáveis, procede o cálculo das variáveis de interesse e a seleção da população alvo a ser estudada. Como exemplo do processo seguido, se apresenta o programa SAS utilizado para a pesquisa do ano 1992 (Programa 4), sendo os programas para as outras pesquisas, análogos, diferenciando-se somente no nomes das variáveis.

Programa GLIM para aplicação do modelo

Para estimar os parâmetros do modelo log-Poisson foi utilizado o pacote GLIM (Programa5), utilizando a lista de comandos apresentada, após definir a data como c : fator classificatório, i : idade, d : duração, te : tempo de exposição (em

meses), f : número de filhos.

Programa STATA para aplicação do modelo

O pacote STATA pode também ser usado para estimar os parâmetros do modelo. Ao utilizar este pacote é preciso definir a fecundidade natural de Henry para cada mulher, segundo a sua idade, e introduzir esta variável junto com as variáveis principais do modelo. O arquivo que contém a informação será um pouco diferente daquele utilizado pelo pacote GLIM. A lista de comandos utilizados para definir o modelo é apresentada no Programa 6.

Programa 1

Para a ENAF 77/78:

```
data wfsa;
infile "c:\pesr02.dat";
input secas 8-9 @;
retain i caseid v006 v007 v008;
retain v101 v103 v115 v109;
retain bdat;
retain v701 v702 v704 v705;
retain s004 s005;
retain nasfil1-nasfil6;
array vetor{6} nasfil1-nasfil6;

if secas = 00 then do;
  input caseid 1-7 v006 40-43 v007 44-47 v008 48-51;
  i = 0;
  do j=1 to 6;
    vetor{j} = 0;
  end;
end;

if secas = 10 then
input v101 10-11 v103 14-15 v115 40-43 v109 26-27;

if secas= 21 then do;
  input bdat 11-14 @;
  if ( i < 6) & (bdat >= v007-60) then do;
    i = i+1;
    input x 11-14;
    vetor{i} = x;
  end;
end;

if secas = 70 then
input v701 10-11 v702 12-13 v704 16-17 v705 18- 19;

if secas = 96 then do;
  input s004 22-23 @;
  if (s004 = 1) or (s004 = 2) then do;
    input s004 22-23 s005 24-25;
    output;
  end;
end;
```

Programa 2

Para a ENDES 86:

```

data dhs86a;
infile "c:\peir01.dat";
input secas 16-17 @;
retain i caseid v005 v008 v011 v014;
retain v101 v102 v106 v107 v133 v138;
retain v208 v214;
retain v501 v502 v503 v509 v510;
retain nasfil1-nasfil6;
array vetor{6} nasfil1-nasfil6;

if secas = 01 then do;
  input caseid 8-15 v005 39-46 v008 51-54 v011 59-62 v014 66;
  i = 0;
  do j=1 to 6;
    vetor{j} = 0;
  end;
end;

if secas = 11 then
input v101 18 v102 19 v106 24 v107 25-26 v133 58-59 v138 66-67;

if secas= 21 then do;
  input bidx 11-19 b3 27-39 @;
  if ( bidx < 6 ) & (b3 >= v008-60) then do;
    i = i+1;
    input x 27-30;
    vetor{i}= x;
  end;
end;

if secas = 22 then
input v208 32 v213 41 v214 42-43;

if secas = 51 then do;
  input v501 18 @;
  if (v501 = 1) or (v501 = 2) then do;
    input v501 5 18 v502 19 v503 20 v509 30-33 v510 34;
    output;
  end;
end; run;

```

Programa 3

Para a ENDES 91/92:

```

data dhs92a;
infile "c:\peiq22.dat";
input secac 9-10 @;
retain i caseid qweight qinic q103c;
retain qwurbrur qwreg3 q107n q107g;
retain q212 q215c;
retain q222;
retain q504 q506 q507c;
retain nasfil1-nasfil6;
array vetor{6} nasfil1-nasfil6;

if secac = 11 then do;
  input caseid 1-8 qweight 52-59 qinic 48-51 qwreg3 65 qwurbrur 66;
  i = 0;
  do j=1 to 6;
    vetor{j} = 0;
  end;
end;

if secac = 12 then
input q103c 60-63 q107n 23 q197g 24;

if secac= 22 then do;
  input q212 11-12 q215c 27-30 @;
  if ( q212 < 6) & (q215c >= v007-60) then do;
    i = i+1;
    input x 27-30;
    vetor{i}= x;
  end;
end;

if secac = 24 then
input q222 15;

if secac = 51 then do;
  input q504 13 @;
  if (q504 = 1) or (q504 = 2) then do;
    input q504 13 q506 14 q507c 34-37;
    output;
  end;
end;

```

Programa 4

```

data dhs92b;
set dhs92a;
verif = 0;
array vetor{6} nasfil1-nasfil6;
if ((qinic - q507c) >=60) then tttotal = 60;
                        else totaal = qinic-q507c;
datainic=qinic-ttotal;
do j = 1 to i;
  if (vetor{j} - datainic >= 9) then v = 0;
  if (0 <=vetor{j} - datainic) and ( < 9) then v = 0;
  if (vetor{j} -datainic < 0 ) then v= 1;
  verif = verif +v;
end;
texp = tttotal;
edad = qinic - q103c;
edad = int ((edad-1)/12);
durac = qinic -q507c;
durac = int((durac-1)/12);
if (verif = 0 ) and (edad-durac<= 12) and (edad <= 49) and ( durac <= 20)
  and (q506 = 1) then output;
run;

```

Programa 5

```
$ assign n(a)= .175, .225, .275, ! 12-14
    .325, .375, .421, .460, .475, ! 15-19
    .477, .475, .470, .460, .475, !20-24
    .455, .449, .442, .435, .428, !25-29
    .420, .410, .400, .389, .375, !30-34
    .360, .343, .325, .305, .280, !35-39
    .247, .207, .167, .126, .087, !40-44
    .055, .035, .021, .011, .003 !45-49$ → Define a taxa de fecundidade natural de Henry
```

```
$ calculate n = n(i-11)$
```

```
$ calculate e = te/12 $
```

```
$ calculate o=%log(n)+%log(e) $ → Cálculo do "offset"
```

```
$ yvar f $→Define b como variável resposta
```

```
$ error Poisson $→Define b com distribuição Poisson
```

```
$ link log $→ Função de ligação do tipo logaritmico
```

```
$ factor cx $→Define f como fator com x níveis
```

```
$ fit c*d $→Ajusta o modelo
```

```
$ display e $→Mostra os estimadores e os seus desvios padrão
```

Para calcular as TFMTE (taxas de fecundidade marital estimadas) foi utilizada a seguinte lista de comandos:

```
$ tabulate the %fv total for i into tf $
```

```
$ tabulate the e total for i.into te$
```

```
$ calculate t=tf/te $
```

```
$ calculate t=%cu(t) $
```

```
$ look t $
```

Programa 6

```
tabulate c, generate(c)→define os níveis do fator : c1, c2, ...,cx
```

```
generate e= te/12
```

```
generate o=log(n)+log(e)→define o "offset "
```

```
poisson f , offset o →define a variável resposta, com distribuição Poisson, a função de ligação do tipo logaritmico é assumida por "default ". Ajusta o modelo.
```

Para ver o efeito do fator classificatório:

```
poisson f c2 -cx, offset (o)→estimadores em referência ao nível c1 do fator
```


Apêndice C

A Evolução da Fecundidade no Peru¹

As mulheres peruanas têm reduzido consideravelmente a sua fecundidade nos últimos quarenta anos. No começo dos anos 50 tinham uma média de sete filhos durante o seu período reproductivo (taxa de fecundidade total), enquanto no inicio dos anos 90 esta média foi reduzida a 3.6 filhos. Este forte descenso da fecundidade é o fator mais influente na queda do crescimento da população.

Esta redução da fecundidade produziu-se na sua maior parte nas mulheres urbanas, cuja taxa de fecundidade total era 2.8 em 1992, enquanto esta taxa nas mulheres rurais era ainda de 6.2. Existe também uma forte diferença da fecundidade segundo o nível de educação: em 1992 a mulheres sem educação nenhuma tinham uma média de sete filhos, enquanto as mulheres com educação universitária tinham somente dois filhos.

As mulheres peruanas reduziram sua fecundidade em todos os grupos etários. Este fato têm sido mais notável nas mulheres maiores de 35 anos. O grupo etário mais fecundo continua sendo aquele de mulheres entre 25 e 30 anos, que duas décadas atrás tinham 24.4% do filhos e no começo dos anos 90 têm o 27% dos filhos.

É importante resaltar que as mulheres menores de 20 anos têm reduzido menos a sua fecundidade que otros grupos etários nos últimos 20 anos, de tal forma

¹Ver: Mujeres Latinoamericanas en Cifras (1994).

que se no começo dos anos 70 tinham 7.2% dos filhos, vinte anos depois têm 8.2% dos mesmos.

<u>ANO</u>	<u>TGT</u>
1950	6.9
1960	6.9
1970	6.0
1975	5.4
1980	4.7
1985	4.0
1990	3.6

Tabela C.1: Taxa Fecundidade Total (TGF).Peru 1950-1990.

	ENAF 77/78	ENDES 86	ENDES 91/92
Area			
Urbano	4.5	3.1	2.8
Rural	7.4	6.3	6.2
Educacao			
Sem educacao	7.3	6.6	7.1
Primaria incompleta	6.7	5.4	
Primaria completa	5.1	4.3	5.1
Secundaria e mais	3.0	2.9	2.0
Regiao			
Lima Metropolitana	3.4	2.5	n.d.
Resto da Costa	4.9	3.8	n.d.
Serra	6.7	5.4	n.d.
Selva	7.3	6.0	n.d.

Tabela C.2: Fecundidade segundo fatores diferenciais. Peru 1977- 1992

Idade	1970-1975		1980-1985		1990-1995	
	Taxa	%	Taxa	%	Taxa	%
15-19	8.6	7.2	8.5	8.5	6.5	8.2
20-24	24.7	20.6	22.7	22.7	20.1	25.3
25-29	29.2	24.4	25.1	25.1	21.5	27.1
30-34	26.6	22.2	20.5	20.5	16.1	20.3
35-39	20.1	16.8	15.8	15.8	11.0	13.8
40-44	8.9	7.4	6.1	6.1	3.6	4.5
45-49	1.9	1.6	1.4	1.4	0.6	0.8
Total		100.0		100.0		100.0

Tabela C.3: Fecundidade segundo grupos etarios

Apêndice D

Mapas

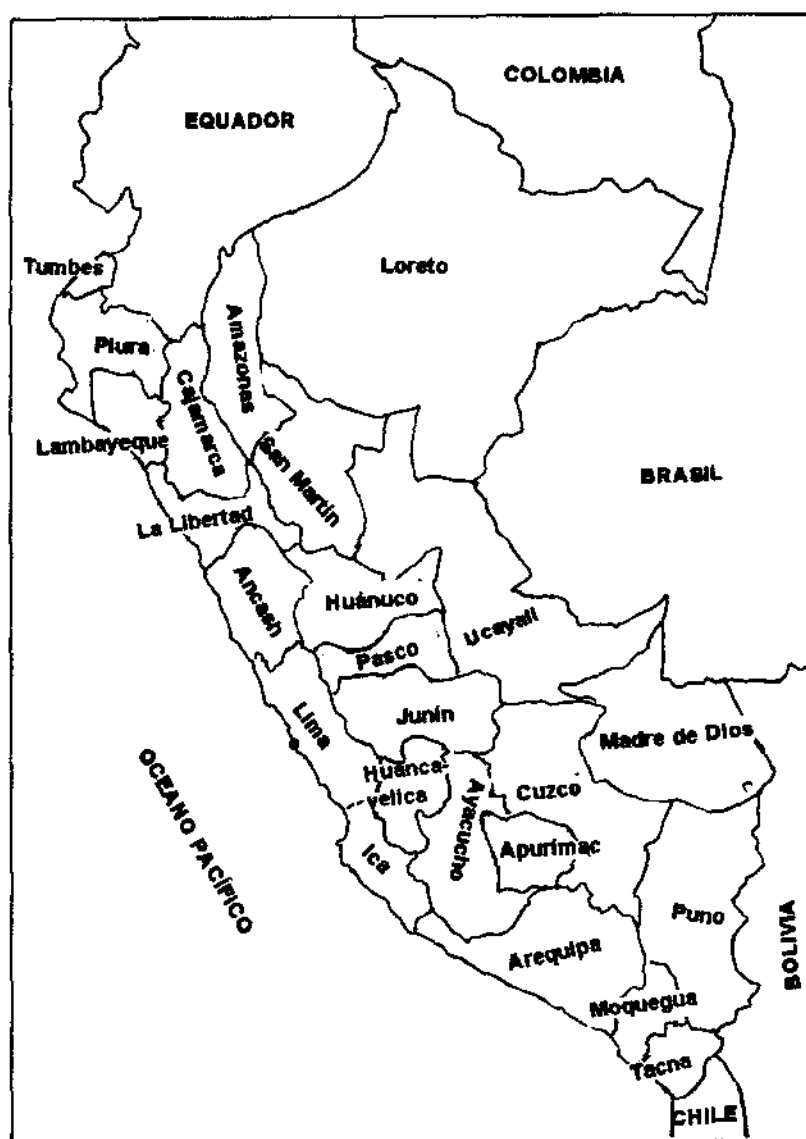


Figura D.1: Peru: Divisão Política

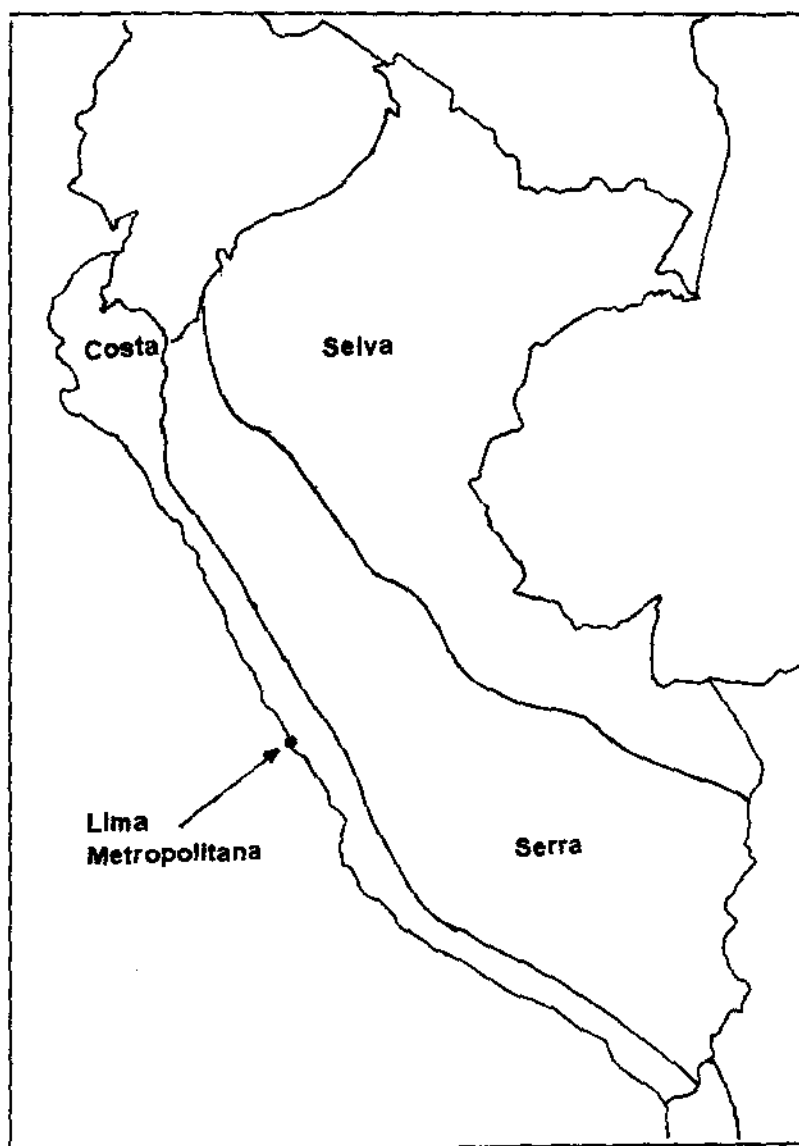


Figura D.2: Peru: Regiões Geográficas

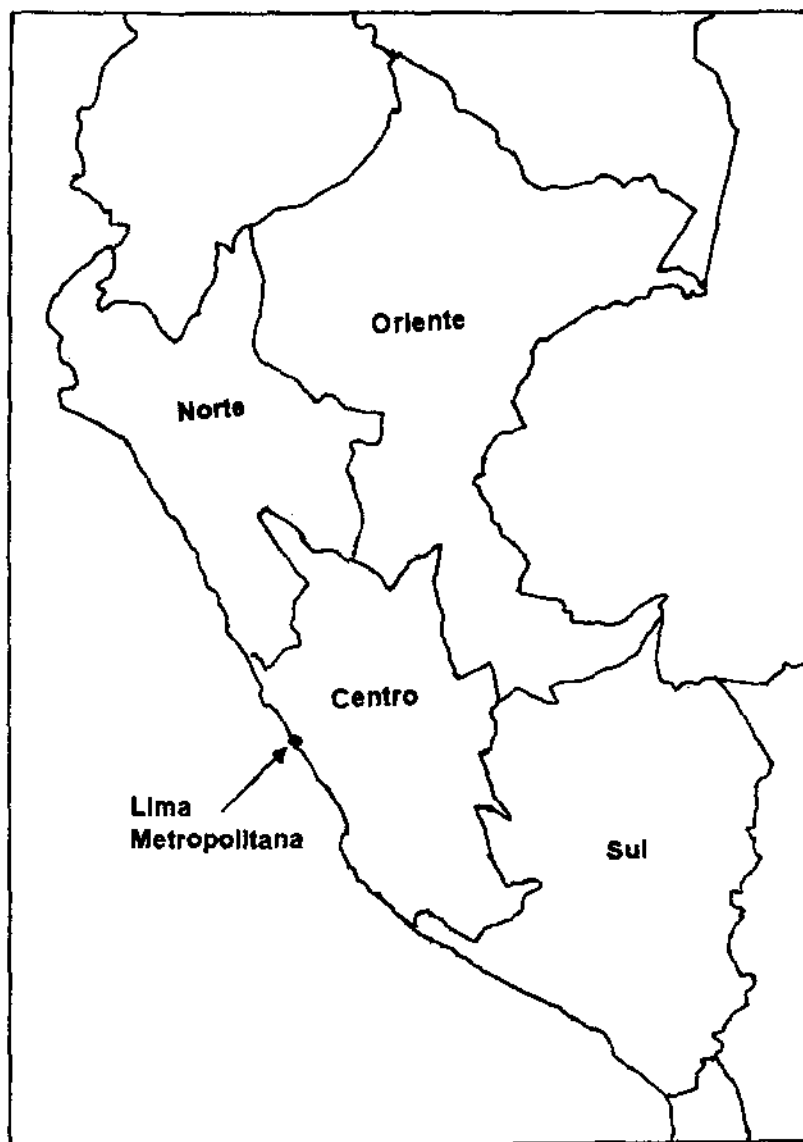


Figura D.3: Peru: Regiões de Planejamento (1977)

Referências Bibliográficas

- [1] Anuário de América Latina y el Caribe (1990) OEA
- [2] Atmann, A.M.G. e Wong, L.R.(1980) Padrões e Tendências da Nupcialidade no Brasil. *Anais Segundo Encontro Nacional de Estudos Populacionais*. Águas de São Pedro, S.P.
- [3] Altmann, A.M.G. e Wong, L.R. (1983) Estimativas de Fecundidade para o Brasil e suas Regiões a partir de informações sobre Nupcialidade e Fecundidade Marital. *Informe Demográfico*, Nº9, Fund.SEADE, São Paulo.
- [4] Berquó, E. (1980) A Análise Demográfica e suas Técnicas: Fatores Estáticos e Dinâmicos - mortalidade e fecundidade.In.*Dinâmica da População: Teoria, Métodos e Técnicas de Análise*. Org. Santos et al. Ed. T.A. Queiroz, São Paulo.
- [5] Becker, G. (1960) *An Economic Analysis of Fertility: Demographic Changes in Developed Countries*. National Bureau of Economic Research, Princeton University Press.
- [6] Bickel, P.J. e Docksum K.A. (1977) *Mathematical Statistics: Basic Ideas and Select Topics*. Holden -Day, Inc.
- [7] Boongarts, J. (1978) A Framework for Analysing the Proximate Determinants of Fertility. *Population and Development Review*, Vol 4, pp.105-132.
- [8] Burden, R.C. e Faires, J.D (1989) *Numerical Analysis*. Boston PWC - Kent.
- [9] Christensen, R. (1990) *Log-Linear Models*. Springer-Verlag.

- [10] Coale, A. (1971) Age Patterns of Marriage. *Population Index*, Vol 25, N°2, pp. 193-214.
- [11] Coale, A. e Mc.Neil, D.R. (1972) The Distribution by Age of the Frequencies of First Marriage in a Female Cohort. *Journal of the American Statistical Association*, Vol 67, N°340, Applications Section.
- [12] Coale, A. e Trussell, J. (1974) Model Fertility Schedules: Variation in the Age Structure of Childbearing in Human Populations. *Population Index*, Vol 40, pp.185-258.
- [13] Coale, A. (1979) El desarrollo de Nuevos Modelos de Nupcialidad y Fecundidad. *CELADE*, Serie D, N°97, Santiago de Chile.
- [14] Cordeiro, G.(1983) Improved Likelihood Ratio Statistics for Generalized Linear Model. *Journal of the Royal Statistical Society, B*, Vol 45, N° 3.
- [15] Cordeiro, G.(1986) *Modelos Lineares Generalizados*. VII Simpósio Nacional de Estatística. Campinas, S.P.
- [16] Draper, N.R e Smith, H. (1981) *Applied Regression Analysis*. John Wiley and Sons.
- [17] Dobson, A (1990) *An Introduction to Generalized Linear Models*. Chapman and Hall.
- [18] Green (1984) Iteratively Rewighted Least Squares for Maximum Likelihood Estimation and Some Robust and Resistant Alternatives. *Journal of the Royal Statistical Society, B*, Vol 46, pp. 149-192.
- [19] Guadalupe, S.E.(1988) *Desarrollo Económico Social y Comportamiento Reproductivo en el Perú*. Concytec, Lima, Peru.
- [20] Henry, L.(1961) Some Data on Natural Fertility. *Eugenics Quarterly*. B(2), pp.81-91.

- [21] Instituto de la Mujer, Ministerio de Asuntos Sociales de España e Facultad de Ciencias Sociales - FLACSO (1994) *Mujeres Latinoamericanas en Cifras*
- [22] Lazo, A. (1994) Marital Fertility in Brazil: Differential by Type of Union and its Importance in the Fertility Transition 1976-1991. *DHS Technical Report* (A ser publicado).
- [23] Lesevic, B. (1987) La Fecundidad desde un Punto de Vista Económico
- [24] Liebenstein, H (1967) *Atrazo e Desenvolvimento Económico*. Fundação Getúlio Vargas, Rio de Janeiro.
- [25] Mauldin, W.P. (1984) Um Marco Conceptual para el Estudio de los Determinantes de la Fecundidad. *Ensayos Sobre Población y Desarrollo*, N°21.
- [26] Moreno, L. (1990) Cambios en la fecundidad marital en México: resultados de dos modelos. *Notas de Población* N°50. CELADE, Santiago de Chile.
- [27] Nelder, J.A. e Wedderburn, R.W. M.(1972) Generalized Linear Models. *Journal of the Royal Statistical Society, A*, 135, pp.370-384.
- [28] Nelder, J.A. e Mc.Cullagh, P. (1989) *Generalized Linear Models*. London Chapman and Hall.
- [29] Oliveira, J. de C. (1991) *Fecundidade e Nupcialidade no Brasil e nos Estados de São Paulo e Rio Grande do Norte: Tendências Passadas e Perspectivas*. IBGE Departamento de População, Rio de Janeiro.
- [30] Ortiz, J.del C. (1987) Perú: Fecundidad Regional y Variables Intermediarias por Nivel Educativo y Grado de Urbanización. In. *Dos Estudios de Fecundidad en el Perú*. Org. Lesevic, B. e Ortiz, J., Ed. Pontificia Universidad Católica del Perú, Facultad de Ciencias Sociales, Lima, Peru.
- [31] Page, H.(1977) Patterns Underlying Fertility Schedules: A decomposition by Both Age and Marriage Duration. *Population Index*, Vol 3, N° 3, pp 85-106.

- [32] Rohatgi, V.K. (1984) *Statistical Inference*. John Wiley and Sons.
- [33] Rodriguez, G. e Aravena, R. (1991) Socio-economic Factors and the Transition to Low Fertility in Less Developed Countries: A Comparative Analysis. Universidad Católica de Chile, Santiago de Chile.
- [34] Rodriguez, G. e Trussell, J. (1980) Maximum Likelihood Estimation of the Parameters of Coale's Model Nuptiality Schedule from Survey Data. *WFS Technical Reports*.
- [35] Rodriguez, G. e Cleland, J. (1988) Modelling Marital Fertility by Age and Duration: An Empirical Appraisal of the Page Model. *Population Studies*, Vol 42, pp 241-257.
- [36] Tuckey, J. (1977) *Exploratory Data Analysis*. New York, Addison Wesley.
- [37] Wedderburn, R.W.M. (1974) Quase-likelihood for Generalized Linear Models and the Gauss-Newton Method. *Biometrika*, Vol 61, pp 439-447.
- [38] Wedderburn, R.W.M. (1978) On the Existence and Uniquess of the Maximum Likelihood Estimators for Certain Generalized Linear Models *Biometrika*, Vol.63. N^o, pp.27-32.
- [39] Weinberger, M., Lloyd, C. e Blanc, A. (1989) Women's Education and Fertility: A Decade of Change in Four Latin American Countries. *International Family Planning Perspectives Vol 15 N^o 1 pp 4-14*.
- [40] Westoff, Moreno e Goldman (1989) The Demographic Impact of Changes in Contraceptive Practice in the Third World Population *Population and Development Review*, Vol 15. N^o1, pp.91-106.
- [41] Wong, L.R.(1983) A Relação Fecundidade-Nupcialidade. Estimativas para São Paulo: 1970-1976. *Informe Demográfico*, N^o9, Fund. SEADE, São Paulo.