O MODELO DE ZEGER PARA ANALISE
DE SÉRIES DE CONTAGENS

Mitti Ayako Hara Koyama

DISSERTAÇÃO APRESENTADA
AO
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA
DA UNIVERSIDADE DE SÃO PAULO
PARA OBTENÇÃO DO GRAU DE MESTRE
EM ESTATÍSTICA

Área de Concentração: Estatística

Orientador: **Prof. Dr. Julio da Motta Singer**

-São Paulo, Janeiro de 1997 -

da dissertação devidamente corrigida
e apresentada por Mitti Ayako Hara Koyama
e aprovada pela Comissão Julgadora

São Paulo, 25 de abril de 1997

Banca Examinadora:

- Prof. Dr. Júlio da Motta Singer - IME-USP
- Prof. Dra. Clélia Maria de Castro Toloi - IME-USP
- Prof. Dra. Sílvia Lopes - UFRGS

Ao meu marido,

Sergio Mikio

Agradecimentos

Desejamos agradecer a todos que contribuíram para a realização deste trabalho, em particular

- ao Professor Doutor Júlio da Motta Singer, orientador e amigo, pelo carinho, paciência e dedicação com que me incentivou e transmitiu seus conhecimentos por todos esses longos anos sem que desistisse de mim.

- à Professora Doutora Clélia Maria de Castro Toloí, orientadora de meu marido, por seu apoio e conselhos constantes.

- à minha chefe Nádia e à Atsuko, pelo apoio e por terem me concedido as férias para que eu pudesse concluir este trabalho.

- ao Professor Doutor Clóvis, por ensinar, rever e aprofundar meus conhecimentos Estatísticos.

- à Dona Rosa, minha assistente e braço direito no Kumon, que cuidou da minha unidade nos últimos anos.

- à Daniela, Marcos e Helena que me apoiaram sempre que precisei.

- aos meus pais e minha sogra pelo apoio, compreensão e socorro em todos os almoços e jantares à qual fui convidada.

- aos meus alunos do Kumon que tiveram a paciência e persistiram no estudo na minha ausência.

- aos meus colegas do SEADE, pelo apoio e incentivo.

Resumo

Neste trabalho, analisaremos modelos de regressão para a investigação do efeito de concentrações atmosféricas de agentes poluentes, umidade do ar e temperatura sobre o número de óbitos para pessoas acima de 65 anos dos residentes no Município de São Paulo no período de 1991 a 1993. Consideraremos modelos que levam em conta não só o fato de a variável resposta ser proveniente de uma contagem como também o fato de as observações sucessivas poderem estar correlacionadas entre si. Os modelos propostos são uma adaptação dos chamados modelos lineares generalizados. Ilustramos as técnicas estatísticas consideradas através do exemplo citado acima e comparamos os resultados com aqueles obtidos através de modelos de regressão de Poisson e Gaussiana.

Abstract

We present regression models to study the effects of air pollution concentration, humidity and temperature on the number of deaths among people over 65 years old, in the city of São Paulo, between 1991 and 1993. We consider models that take into account not only the fact that the response variables are counts, but also the fact that the successive observations may be correlated. The proposed models are an extension of generalized linear models. We illustrate such statistical techniques via the example described above and the results are compared to those obtained by Poisson and Gaussian regression models.

Índice

Capítulo 1- Introdução	1
Capítulo 2 - Os dados	3
Capítulo 3 - O modelo de Zeger para análise de séries de contagens	5
Capítulo 4 - Análise de dados	15
Capítulo 5 - Implementação computacional e conclusões.....	28
 Apêndice	
A.1 - Modelos Lineares Generalizados	31
A.2 - Modelos de quasi-verossimilhança de estimação	39
A.3 - Modelos de quasi-verossimilhança para dados correlacionados.....	43
A.4 - Análise do número de óbitos para pessoas acima de 65 anos com variáveis defasadas	46
A.5 - Programas utilizados para implementar o modelo de Zeger	48
 Referências Bibliográficas	 69

Capítulo 1

Introdução

Concentrações atmosféricas de agentes poluidores têm sido associadas à mortalidade urbana em diversos estudos. O Laboratório de Poluição Atmosférica Experimental da Faculdade de Medicina da USP, juntamente com o Centro de Estatística Aplicada (CEA) da USP e a Escola de Saúde Pública de Harvard desenvolveram estudos utilizando modelos convencionais de regressão linear e detectaram uma associação significativa entre a concentração atmosférica de alguns poluentes e número de óbitos, tanto para crianças com menos de 5 anos (Saldiva et al., 1993) quanto para adultos com mais de 65 anos (Saldiva et al., 1994) no período de 1990-1991 na Região Metropolitana de São Paulo (RMSP).

Para confirmar o resultado anterior com base numa série maior de dados, realizou-se um novo estudo (Singer e Conceição, 1994) com dados referentes ao período de 1991 a 1993. Utilizando os mesmos métodos estatísticos, observou-se uma forte associação entre a concentração de alguns poluentes e mortalidade para adultos com mais de 65 anos.

No modelo de regressão linear convencional assumimos que a distribuição do número diário de óbitos é normal. Entretanto, esse número corresponde a um processo de contagem e uma distribuição possivelmente mais adequada seria a distribuição de Poisson. Isto nos leva a considerar a utilização de Modelos Lineares Generalizados (McCullagh e Nelder, 1989) como uma alternativa para a análise. Como os registros de número de óbitos de que dispomos constituem uma série de tempo com medidas diárias, também seria interessante levar em consideração uma possível dependência entre as observações a fim de produzirmos resultados mais sensíveis. Tanto os modelos clássicos de regressão linear quanto os de regressão Poisson não levam em conta essa possível dependência.

Zeger (1988) estendeu a teoria de Modelos Lineares Generalizados para as situações com observações dependentes. Campbell (1994), num estudo semelhante ao da poluição atmosférica considerado aqui, utilizou o método de Zeger (1988) para analisar os dados de uma investigação sobre Síndrome da Morte Súbita Infantil comparando-o com o de regressão Poisson. Nesse estudo, ele procurou investigar a relação entre essas mortes e a temperatura ambiente. Foram ajustados quatro modelos de regressão, obtendo-se estimadores com erros padrões ligeiramente menores sob o método de Zeger. Num estudo anterior (Murphy e Campbell, 1987) foi utilizada uma análise convencional de Box-Jenkins (ver Wei, 1990, por exemplo) baseada na raiz quadrada da soma de número de óbitos de dois dias consecutivos. Esta análise sugeriu uma associação entre uma mudança na temperatura e nas mortes com defasagem entre 4 e 8 dias. Essa análise indica entretanto, apenas uma associação entre as variáveis defasadas de alguns dias, não permitindo que avaliemos a magnitude do efeito de uma sobre a outra.

A proposta deste trabalho é estudar a associação entre mortalidade e os agentes poluidores do ar atmosférico segundo o método proposto por Zeger (1988) comparando-o com as alternativas mencionadas acima.

Capítulo 2

Os dados

Nesta seção descreveremos os dados do estudo de poluição atmosférica (Singer e Conceição, 1994) utilizados para ilustrar o método proposto por Zeger (1988). Foram coletados três conjuntos de dados durante 3 anos no período de 1 de janeiro de 1991 a 31 de dezembro de 1993:

- a) Concentrações diárias de agentes poluidores, fornecidas por 12 estações da CETESB espalhadas em diversos locais da cidade de São Paulo;
- b) Condições climáticas diárias, fornecidas pelas estações da CETESB situadas no Parque D. Pedro e no Parque do Estado na Água Funda;
- c) Dados de mortalidade diária de residentes no Município de São Paulo, fornecidos pelo Programa de Aprimoramento das Informações de Mortalidade (PROAIM).

Mais especificamente, as variáveis consideradas neste estudo podem ser classificadas como:

(i) Variáveis de qualidade do ar:

- Part - Concentração diária de material particulado ($\mu\text{g} / \text{m}^3$);
- SO₂ - Concentração diária de Dióxido de Enxofre ($\mu\text{g} / \text{m}^3$);
- NO₂ - Concentração diária de Nitrogênio ($\mu\text{g} / \text{m}^3$);
- CO - Concentração diária de Monóxido de Carbono (ppm);
- O₃ - Concentração diária de Ozônio (ppb);

A CETESB produz médias horárias das concentrações dos poluentes em cada estação. Os dados fornecidos pela CETESB para cada estação medidora representam médias de 24 horas das concentrações de PART, NO₂, SO₂. A medida fornecida para o CO é a maior média móvel de 8 horas no dia. Já para O₃, é a maior média horária. O período de coleta das observações compreende intervalos de 24 horas contados apartir das 16 horas. Entretanto, nem todas as estações medem as concentrações de todos os poluentes. As concentrações diárias de poluentes utilizadas neste estudo correspondem à média dos valores diários fornecidos pelas estações onde esta medida estava disponível .

(ii) Variáveis climáticas:

Temp - Temperatura mínima (°C);

Umid - Umidade relativa do ar ao meio - dia (%).

A maioria das informações de temperatura e umidade foi obtida na estação do Parque D. Pedro e os demais dados, na estação do Parque da Água Branca.

(iii) Variáveis Demográficas:

Totsem - Número de óbitos diário para crianças de 0 a 4 semanas;

Totmes - Número de óbitos diário para crianças de 1 a 11 meses;

Respmes - Número de óbitos diário por causas respiratórias para crianças de 1 a 11 meses;

Tot5 - Número de óbitos diário para crianças de 1 a 5 anos;

Resp5 - Número de óbitos diário por causas respiratórias para crianças de 1 a 5 anos;

Tot65 - Número de óbitos diário para pessoas com mais de 65 anos;

Capítulo 3

O modelo de Zeger para análise de séries de contagens

Seja $\mathbf{Y} = (y_1, \dots, y_N)^t$ com y_T representando a resposta (número de óbitos, por exemplo) no instante T , para $T = 1, \dots, N$. O correspondente vetor de variáveis explicativas será denotado por

$$\mathbf{x}_T = (x_{1T}, \dots, x_{pT})^t$$

com x_{iT} representando a i -ésima variável explicativa (concentração de SO_2 , por exemplo) observada no instante T . A matriz de variáveis explicativas é expressa por

$$\mathbf{X}^t = (\mathbf{x}_1, \dots, \mathbf{x}_N).$$

Denotando $\boldsymbol{\mu} = (\mu_1, \dots, \mu_N)^t$ com $\mu_T = E(y_T)$ para $T = 1, \dots, N$, os modelos abordados neste trabalho podem ser escritos na forma

$$g(\boldsymbol{\mu}) = \mathbf{X} \boldsymbol{\beta} \quad (3.1)$$

onde $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^t$ é um vetor de parâmetros desconhecidos a ser estimado e $g(\cdot)$ é uma função monotônica diferenciável.

Como já foi apontado na Seção 1, podemos considerar o número de óbitos diários como um processo de contagem e desta forma associar a ele uma distribuição de Poisson. Se pudéssemos supor independência entre as diversas observações do número de óbitos, poderíamos analisar a associação entre elas e os agentes poluidores do ar através dos Modelos Lineares Generalizados descritos no Apêndice A.1. Essa suposição de independência, no entanto, pode não ser adequada, pois as contagens correspondentes

constituem uma série de tempo, na qual, geralmente se esperam correlações entre as observações sucessivas.

Duas classes de modelos podem ser consideradas para a análise de séries de tempo: **modelos condicionais** a observações passadas e **modelos baseados em processos latentes**. Na primeira classe, podemos considerar os **modelos de transferência**. Nesse último caso utiliza-se a metodologia de Box-Jenkins (ver, por exemplo, Wei, 1990) e a grande dificuldade é a identificação do modelo quando o número de variáveis explicativas é maior ou igual a dois. Além disso, existem situações nas quais nem sempre é possível obter séries de observações tão longas que permitam a utilização dessas técnicas, embora este não seja o caso. A proposta deste trabalho é avaliar um modelo baseado em processos latentes, que permita incorporar estruturas de dependência entre observações.

Zeger (1988) sugere um modelo log-linear baseado na suposição de que $\{y_T, T=1, \dots, N\}$ é uma sequência de contagens independentes quando condicionadas a uma sequência de processo latente $\{\varepsilon_T, T=1, \dots, N\}$ tal que $E(\varepsilon_T) = 1$ e $Cov(\varepsilon_T, \varepsilon_{T+\tau}) = \sigma^2 \rho_\varepsilon(\tau)$, com $\rho_\varepsilon(\tau)$ denotando a função de autocorrelação ($\tau \geq 0$). Esse modelo é dado por

$$E(y_T | \varepsilon_T) = \text{Var}(y_T | \varepsilon_T) = \exp(x_T^t \beta) \varepsilon_T. \quad (3.2)$$

Como $E(E(y_T | \varepsilon_T)) = E(y_T)$ e $\text{Var}(y_T) = \text{Var}(E(y_T | \varepsilon_T)) + E(\text{Var}(y_T | \varepsilon_T))$, a especificação acima introduz a seguinte estrutura para os momentos marginais:

$$E(y_T) = \exp(x_T^t \beta) = \mu_T \quad (3.3)$$

$$\text{Var}(y_T) = \mu_T + \sigma^2 \mu_T^2 \quad (3.4)$$

Além disso,

$$\rho_y(T, \tau) = \text{Corr}(y_T, y_{T+\tau}) = \frac{\rho_\varepsilon(\tau)}{\{1 + (\sigma^2 \mu_T)^{-1}\}^{1/2} \{1 + (\sigma^2 \mu_{T+\tau})^{-1}\}^{1/2}}. \quad (3.5)$$

De (3.4) e (3.5), podemos concluir que o processo $\{\varepsilon_T\}$ introduz tanto uma sobredispersão como uma autocorrelação em $\{y_T\}$. Como $\text{Var}(y_T) = \mu_T + \sigma^2 \mu_T^2$ concluímos também que o grau de sobredispersão depende de μ_T , e não necessariamente precisamos supor uma sobredispersão constante como em Cox (1983).

Fazendo $\varepsilon = (\varepsilon_1, \dots, \varepsilon_N)^t$, teremos

$$E(Y | \varepsilon) = A \varepsilon \quad \text{e} \quad \text{Var}(Y | \varepsilon) = A \varepsilon \quad (3.6)$$

onde $A = \text{diag}(\mu_1, \dots, \mu_N)$.

A estrutura para a matriz de covariância induzida pelo modelo é dada por

$$V = \text{Var}(Y) = A + \sigma^2 A R_\varepsilon A \quad (3.7)$$

onde R_ε é uma matriz quadrada de ordem N com $\text{Corr}(\varepsilon_i, \varepsilon_j)$ na posição (i,j) . Denotaremos por θ_p o vetor que contém todos os parâmetros especificadores da matriz R_ε . Observamos aqui que V depende não só de β mas também do vetor $\theta = (\sigma^2, \theta_p^t)^t$.

Discutiremos agora a estimação dos parâmetros da regressão, β , dado um estimador consistente de θ . Num modelo sem dependência entre as observações, e sob um enfoque de quasi-verossimilhança (ver Apêndice A.2) isto corresponde a resolver as equações

$$U(\hat{\beta}) = \sum_{T=1}^N \frac{\partial \mu_T(\hat{\beta})}{\partial \beta} [v_T(\hat{\beta})]^{-1} (y_T - \mu_T(\hat{\beta})) = 0 \quad (3.8)$$

onde $\mathbf{V} = \text{diag} (\text{Var}(y_1), \dots, \text{Var}(y_N))$ e $\text{Var}(y_T)$ dada por (3.4). A expressão (3.8) corresponde às equações de quasi-verossimilhança (A.2.5). O estimador de quasi-verossimilhança $\hat{\beta}$ é consistente e assintoticamente normal além de ser robusto (McCullagh, 1983). A equação (3.8) pode ser reescrita na forma matricial

$$\frac{\partial \mu'(\hat{\beta})}{\partial \beta} [\mathbf{V}(\hat{\beta})]^{-1} (\mathbf{Y} - \mu(\hat{\beta})) = \mathbf{0}. \quad (3.9)$$

O nosso objetivo é generalizar as equações para o caso de séries de tempo. Assim podemos obter um estimador de β sob o modelo definido por (3.1) e (3.2) resolvendo o sistema de equações

$$\frac{\partial \mu(\hat{\beta})'}{\partial \beta} [\mathbf{V}(\hat{\beta}, \hat{\theta})]^{-1} (\mathbf{Y} - \mu(\hat{\beta})) = \mathbf{0} \quad (3.10)$$

onde $\hat{\theta}$ é uma estimativa $\sqrt{N}$ -consistente de θ , isto é, $\sqrt{N}(\hat{\theta} - \theta) = o_p(1)$. Como $\hat{\theta}$ é função de $\hat{\beta}$, podemos reescrever (3.10) como

$$\mathbf{U}(\hat{\beta}) = \frac{\partial \mu(\hat{\beta})'}{\partial \beta} [\mathbf{V}(\hat{\beta}, \hat{\theta}(\hat{\beta}))]^{-1} (\mathbf{Y} - \mu(\hat{\beta})) = \mathbf{0}.$$

Então, sob certas condições de regularidade (Zeger e Liang, 1986), demonstra-se que $\sqrt{N}(\hat{\beta} - \beta)$ tem distribuição assintoticamente normal multivariada com média zero e matriz de covariância

$$\mathbf{V}(\hat{\beta}) = \lim_{N \rightarrow \infty} \frac{1}{N} \left(\frac{\partial \mu'}{\partial \beta} \mathbf{V}^{-1} \frac{\partial \mu}{\partial \beta} \right)^{-1}.$$

De modo semelhante à estimação pelo método de quasi-verossimilhança, o cálculo de β para um dado valor de $\hat{\theta}(\beta)$ pode ser feito através de um procedimento de mínimos quadrados ponderados iterativos. O estimador na j -ésima iteração é obtido através de

$$\hat{\beta}^{(j+1)} = \left\{ \frac{\partial \mu'(\hat{\beta}^{(j)})}{\partial \beta} \left[V(\hat{\beta}^{(j)}, \hat{\theta}^{(j)}) \right]^{-1} \frac{\partial \mu(\hat{\beta}^{(j)})}{\partial \beta} \right\}^{-1} \left\{ \frac{\partial \mu'(\hat{\beta}^{(j)})}{\partial \beta} \left[V(\hat{\beta}^{(j)}, \hat{\theta}^{(j)}) \right]^{-1} \mathbf{Z} \right\} \quad (3.11)$$

com

$$\mathbf{Z} = \left(\frac{\partial \mu'(\hat{\beta}^{(j)})}{\partial \beta} \right) \hat{\beta}^{(j)} + \left(\mathbf{Y} - \mu(\hat{\beta}^{(j)}) \right).$$

A partir de algum procedimento de estimação de $\theta(\beta)$, $\hat{\beta}$ é obtido temporariamente através de (3.11). Com base nesse valor temporário, atualiza-se o valor de $\hat{\theta}(\hat{\beta})$. Este procedimento é repetido até a convergência.

O cálculo de (3.11) requer a inversão da matriz de covariância $\mathbf{V}$ (de ordem N). Para simplificar os cálculos computacionais, Zeger (1988) sugeriu a seguinte aproximação para a matriz de covariâncias de (3.7)

$$\mathbf{V}_R = \mathbf{D}^{1/2} \mathbf{R}_e \mathbf{D}^{1/2} \quad (3.12)$$

onde $\mathbf{D}(\mu, \sigma^2) = \text{diag} (\mu_1 + \sigma^2 \mu_1^2, \dots, \mu_N + \sigma^2 \mu_N^2)$.

Além disso, podemos aproximar a matriz $\mathbf{R}_e$ por uma matriz de autocorrelações de um processo autoregressivo $\mathbf{R}(\alpha)$, onde o vetor α , de dimensão $s \times 1$, contém os correspondentes parâmetros definidores. Em séries temporais não muito longas, um modelo autoregressivo AR(1) pode se ajustar bem aos dados (Lindsey, 1993) e por esse motivo a ele restringiremos a nossa discussão; segundo esse modelo, cada resposta depende diretamente apenas da resposta imediatamente precedente. Nesse caso, o vetor α tem um único elemento ρ , que é a autocorrelação de ordem 1. Desse modo

$$\mathbf{R}(\rho) = \begin{pmatrix} 1 & \rho & \dots & \rho^{N-2} & \rho^{N-1} \\ \rho & 1 & \dots & \rho^{N-3} & \rho^{N-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho^{N-2} & \rho^{N-3} & \dots & 1 & \rho \\ \rho^{N-1} & \rho^{N-2} & \dots & \rho & 1 \end{pmatrix}.$$

com inversa dada por

$$\mathbf{R}^{-1}(\rho) = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho & \dots & 0 & 0 \\ -\rho & 1+\rho^2 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \vdots & 0 & \dots & 1+\rho^2 & -\rho \\ 0 & 0 & \dots & -\rho & 1 \end{pmatrix}$$

que apresenta a forma de uma matriz diagonal em bandas (ver Lindsey (1993), por exemplo). Devemos lembrar aqui, que esta inversa só existirá se tivermos $|\rho| < 1$, que é a condição necessária para estacionariedade (Priestley, 1981). Como $\mathbf{R}(\rho)$ é inversível, então podemos considerar a decomposição de Cholesky

$$\mathbf{R}^{-1}(\rho) = [\mathbf{L}(\rho)]^t \mathbf{L}(\rho)$$

onde $\mathbf{L}(\rho)$ é uma matriz de filtro linear autoregressivo, ou seja

$$\mathbf{L}(\rho) = \frac{1}{\sqrt{1-\rho^2}} \begin{pmatrix} \sqrt{1-\rho^2} & 0 & \dots & 0 & 0 \\ -\rho & 1 & \dots & 0 & 0 \\ 0 & -\rho & 1 & \dots & \vdots \\ \vdots & 0 & \dots & 1 & 0 \\ 0 & 0 & \dots & -\rho & 1 \end{pmatrix}.$$

De um modo geral, o número de diagonais secundárias em $L(\rho)$ corresponde à ordem do AR. Multiplicando-se a matriz de filtro linear autoregressivo pelo vetor de observações Y , a menos do fator $1/\sqrt{1-\rho^2}$, obteremos um vetor cujos elementos são $\sqrt{1-\rho^2} y_1$ e $y_T - \rho y_{T-1}$ ($T > 1$). A inversa da matriz V_R , ignorando os efeitos de aproximação, é então dada por

$$V_R^{-1} \cong D^{-1/2} L^t L D^{-1/2}.$$

Seja $\theta_R = (\sigma^2, \alpha)$ e suponhamos que $\hat{\theta}_R$ seja um estimador $\sqrt{N}$ -consistente de θ_R . Definiremos agora $\hat{\beta}_R$ como a solução das equações de estimação

$$\frac{\partial \mu(\hat{\beta}_R)^t}{\partial \beta} \left[V_R(\hat{\theta}_R) \right]^{-1} (Y - \mu(\hat{\beta}_R)) = 0.$$

Dado $\hat{\theta}_R$, o algoritmo para encontrar $\hat{\beta}_R$ é análogo ao apresentado em (A.1.11).

Se ρ for conhecido ou passível de ser estimado, um modo de ajustar o modelo consiste em multiplicar as variáveis resposta e explicativas por L e prosseguir normalmente com as técnicas de estimação dos parâmetros da regressão (A.1.11). Em outras palavras, a operação apropriada de filtragem, que representa uma combinação ponderada de valores presentes e passados, é aplicada a todas as variáveis antes do ajuste da regressão. Finalmente, o procedimento de mínimos quadrados ponderados iterativos poderá ser dado de um modo bastante simplificado. Assim para a $(j+1)$ -ésima iteração

$$\hat{\beta}_R^{(j+1)} = \left\{ \left(\hat{L} \hat{D}^{-1/2} \frac{\partial \mu(\hat{\beta}^{(j)})}{\partial \beta} \right) \left(\hat{L} \hat{D}^{-1/2} \frac{\partial \mu(\hat{\beta}^{(j)})}{\partial \beta} \right)^t \right\}^{-1} \left\{ \hat{L} \hat{D}^{-1/2} \frac{\partial \mu(\hat{\beta}^{(j)})}{\partial \beta} \right\}^t (\hat{L} \hat{D}^{-1/2} Z). \quad (3.13)$$

com

$$\mathbf{Z} = \left(\frac{\partial \boldsymbol{\mu}(\hat{\boldsymbol{\beta}}^{(j)})}{\partial \boldsymbol{\beta}} \right) \hat{\boldsymbol{\beta}}^{(j)} + \mathbf{Y} - \hat{\boldsymbol{\mu}}(\hat{\boldsymbol{\beta}}^{(j)}), \quad \hat{\mathbf{L}} = \mathbf{L}(\boldsymbol{\alpha}(\hat{\boldsymbol{\beta}}^{(j)})) \text{ e } \hat{\mathbf{D}} = \mathbf{D}(\boldsymbol{\mu}(\hat{\boldsymbol{\beta}}^{(j)})).$$

O algoritmo acima é essencialmente um algoritmo de mínimos quadrados ponderados iterativos filtrados e envolve os seguintes passos:

- (I) Os valores correntes de $\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}}$ e $\mathbf{Z}$ são ponderados por $\mathbf{D}^{-1/2}$.
- (II) Os valores de $\mathbf{D}^{-1/2} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}}$ e $\mathbf{D}^{-1/2} \mathbf{Z}$ são filtrados através de um filtro autoregressivo de ordem 1.
- (III) As equações de mínimos quadrados (3.13) são resolvidas.
- (IV) Os passos de (I) a (III) são repetidos iterativamente até a convergência.

Sob certas condições de regularidade, $\sqrt{N}(\hat{\boldsymbol{\beta}}_R - \boldsymbol{\beta})$ tem distribuição assintoticamente normal multivariada com média zero e matriz de covariância

$$\mathbf{V}(\hat{\boldsymbol{\beta}}_R) = \mathbf{I}_0^{-1} \mathbf{I}_1 \mathbf{I}_0^{-1} \quad (3.14)$$

onde

$$\mathbf{I}_0 = \lim_{N \rightarrow \infty} \frac{1}{N} \left(\frac{\partial \boldsymbol{\mu}^t}{\partial \boldsymbol{\beta}} \mathbf{V}_R^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}} \right) \quad (3.15)$$

e

$$\mathbf{I}_1 = \lim_{N \rightarrow \infty} \frac{1}{N} \left(\frac{\partial \boldsymbol{\mu}^t}{\partial \boldsymbol{\beta}} \mathbf{V}_R^{-1} \mathbf{V} \mathbf{V}_R^{-1} \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\beta}} \right) \quad (3.16).$$

quando estas matrizes existem (Zeger ,1988).

A expressão para I_1 em (3.14) pode ser reescrita como

$$I_1 = \frac{1}{N} (\mathbf{LD}^{-1/2} \mathbf{AX})' (\mathbf{LD}^{-1/2} \mathbf{V} \mathbf{D}^{-1/2} \mathbf{L}) (\mathbf{LD}^{-1/2} \mathbf{AX}) \quad (3.17)$$

Entretanto (3.17) envolve a matriz $\mathbf{V}$ cuja ordem é N . Para situações como no caso analisado aqui onde N é da ordem de 1000, o seu armazenamento apresenta problemas computacionais. Uma alternativa sugerida por Campbell (1994) para contornar esse problema é construir o vetor $\mathbf{S}(\hat{\beta}) = \mathbf{Y} - \mu(\hat{\beta})$, e estimar $\mathbf{V}$ através de

$$\mathbf{S}(\hat{\beta}) [\mathbf{S}(\hat{\beta})]' \quad (3.18)$$

Desta forma, I_1 pode ser reescrita como

$$I_1 = \left(\hat{\mathbf{L}} \hat{\mathbf{D}}^{-1/2} \frac{\partial \mu(\hat{\beta})}{\partial \beta} \right)' \left(\hat{\mathbf{L}} \hat{\mathbf{D}}^{-1/2} \mathbf{S}(\hat{\beta}) \right) \left(\hat{\mathbf{L}} \hat{\mathbf{D}}^{-1/2} \mathbf{S}(\hat{\beta}) \right)' \left(\hat{\mathbf{L}} \hat{\mathbf{D}}^{-1/2} \frac{\partial \mu(\hat{\beta})}{\partial \beta} \right).$$

Assim, basta armazenar matrizes de ordem $p \times N$ e multiplicá-las por vetores de dimensão $N \times 1$.

Zeger (1988) analisou a perda de precisão com a utilização de $\hat{\beta}_R$ no lugar de $\hat{\beta}$ em três casos específicos: regressão com tendência, função escada e uma regressão utilizando uma variável explicativa com distribuição $N(0,1)$ para três valores de σ^2 e de ρ_e . Ele observou numericamente que $\hat{\beta}_R$ é tão eficiente quanto $\hat{\beta}$. Além disso, Zeger observou que a eficiência independe dos valores de σ^2 e ρ_e , o que sugere que a perda de precisão de $\hat{\beta}_R$ relativamente a $\hat{\beta}$ é pequena frente às suas vantagens computacionais. O cálculo do estimador $\hat{\beta}$ (3.11) é computacionalmente mais dispendioso.

Os parâmetros de perturbação podem ser estimados pelo método dos momentos. Lembrando que $\text{Var}(y_T) = \mu_T + \sigma^2 \mu_T^2$, na j -ésima iteração do processo de estimação, σ^2 pode ser estimado por

$$[\hat{\sigma}^{(j)}]^2 = \frac{\sum_{T=1}^N \left\{ \left(y_T - \mu_T(\hat{\beta}^{(j)}) \right)^2 - \mu_T(\hat{\beta}^{(j)}) \right\}}{\sum_{T=1}^N [\mu_T(\hat{\beta}^{(j)})]^2} \quad (3.19)$$

e a função de autocorrelação de ε_T pode ser estimada por

$$\hat{\rho}_\varepsilon^{(j)}(\tau) = [\hat{\sigma}^{(j)}]^2 \frac{\sum_{T=\tau+1}^N \left\{ \left(y_T - \mu_T(\hat{\beta}^{(j)}) \right) \left(y_{T-\tau} - \mu_{T-\tau}(\hat{\beta}^{(j)}) \right) \right\}}{\sum_{T=\tau+1}^N \mu_T(\hat{\beta}^{(j)}) \mu_{T-\tau}(\hat{\beta}^{(j)})} \quad (3.20)$$

onde $\mu(\hat{\beta}^{(j)}) = \exp(\mathbf{X}\hat{\beta}^{(j)})$.

Supondo que ε_T seja um processo autoregressivo de primeira ordem, teremos $\hat{\alpha} = (\hat{\rho}_\varepsilon(1))$ que é a autocorrelação com defasagem 1 (lag 1). Para casos mais gerais, $\hat{\rho}_\varepsilon(\tau)$ pode ser obtido através das equações de Yule-Walker (ver Morettin e Toloi, 1987 por exemplo). Uma grande desvantagem deste método é de que o estimador da variância nem sempre é positivo e ocasionalmente, durante o processo iterativo podemos obter valores negativos. Caso isto ocorra, Campbell (1994) sugere que o estimador de $\hat{\sigma}^2$ acima seja substituído por

$$[\hat{\sigma}^{(j)}]^2 = \sum_{T=1}^N (y_T - \hat{\mu}_T(\hat{\beta}^{(j)}))^2 / (N-1).$$

Capítulo 4

Análise de dados

O nosso objetivo é aplicar as técnicas descritas no Capítulo 3 ao problema descrito no Capítulo 2. Na Figura 1, apresentamos a série de números de óbitos para pessoas com mais de 65 anos; nela podemos observar uma aparente sazonalidade anual e uma tendência positiva.

FIGURA 1. ÓBITOS DE PESSOAS COM MAIS DE 65 ANOS NO PERÍODO DE 1991-1993 NA RMSP

Com o objetivo de controlar esta sazonalidade, utilizaremos o seguinte conjunto de variáveis indicadoras nas análises: dias frios (DIASF) com temperatura mínima abaixo de 14,0°C, dias quentes (DIASQ) com temperatura mínima acima de 18,1°C, dias secos (DIASS) com umidade relativa do ar ao meio-dia menor que 57%, dias úmidos (DIASU) com umidade relativa do ar ao meio-dia maior que 75%, primavera (PRI), outono (OUT) e inverno (INV). Os pontos de corte para os dias frios/quente e úmidos/secos correspondem aos primeiros e terceiros quartis das respectivas distribuições. Essas variáveis serão designadas Condições Climáticas. É conveniente lembrar que a ausência da variável indicadora associada ao verão decorre do fato de estarmos considerando a taxa média de óbitos desta estação como referência, sendo os coeficientes associados às demais estações interpretados como as diferenças entre as taxas médias de óbitos de cada uma delas e aquela correspondente ao verão. Essa parametrização também vale para as variáveis referentes à temperatura diária; nesse caso adotamos a taxa média de óbitos nos dias com temperaturas mínima entre 14,0°C e 18,1°C como referência; no caso da umidade, a referência é a taxa média de óbitos em dias onde a umidade relativa do ar ao meio-dia está entre 57% e 75%.

Para efeitos descritivos apresentamos as médias das taxas diárias de óbitos na TABELA 4.1, mensalmente nos três anos considerados. Os valores mais altos dessas taxas ocorrem nos meses de julho a setembro (inverno) e evidenciam a sazonalidade mencionada anteriormente. Pode-se também observar um aumento no número de óbitos, ao longo dos três anos indicando uma tendência de crescimento.

**TABELA 4.1.MÉDIA DAS TAXAS DIÁRIAS DE ÓBITOS
PARA PESSOAS COM MAIS DE 65 ANOS**

Meses	1991	1992	1993
janeiro	55	60	65
fevereiro	59	64	60
março	57	60	67
abril	60	63	77
maio	65	70	74
junho	67	74	87
julho	71	80	79
agosto	73	82	84
setembro	73	74	75
outubro	69	71	70
novembro	60	67	65
dezembro	61	66	65
Total	64	69	72

Uma investigação anterior envolvendo os mesmos dados (Singer e Conceição, 1994) evidencia uma associação entre a mortalidade de pessoas acima de 65 anos e temperatura, umidade e concentrações de poluentes (com exceção de O_3).

Para explorarmos uma possível correlação com defasagem de tempo entre o número de óbitos e essas variáveis, foram calculadas as correspondentes correlações cruzadas. Os resultados sugerem uma associação entre a concentração de SO_2 e o número de óbitos para pessoas com mais de 65 anos com uma defasagem de 3 dias, ou seja, que o número de óbitos pode estar associado com a concentração de SO_2 de três dias atrás. Já a associação entre número de óbitos de pessoas com mais de 65 anos e temperatura sugere que o valor presente de óbitos de pessoas com mais de 65 anos possa depender das temperaturas de 2 e 3 dias anteriores. Estes resultados devem ser vistos com cautela, uma vez que seria mais adequado calcular as correlações cruzadas após o ajuste de modelos. Além disso, por motivos de comparabilidade com estudos anteriores (Singer e Conceição, 1994) não

usaremos neste capítulo as variáveis defasadas. Os resultados das análises com as defasagens estão apresentados no Apêndice A.5.⁴

Realizamos análises com taxa de mortalidade por causas respiratórias para crianças com menos de 1 ano e crianças de 1 a 5 anos, entretanto, tanto os modelos de regressão linear simples como os de Poisson apresentaram-se não significantes. Desta forma, investigaremos particularmente o comportamento do número de óbitos diários para pessoas com mais de 65 anos (TOT65).

Para investigarmos a magnitude dos efeitos dos agentes poluidores do ar sobre a taxa de mortalidade diária de pessoas com mais de 65 anos (TOT65) ajustamos regressões de Poisson com função de ligação logarítmica, tendo as características climáticas mencionadas anteriormente como covariáveis.

A Tabela 4.2 contém os coeficientes estimados e os correspondentes erros padrões dos diferentes modelos de regressão Poisson considerados. No primeiro modelo incluímos todas as variáveis associadas às condições climáticas e aquelas correspondentes aos agentes poluidores em conjunto. Nos demais modelos, cada agente poluidor é considerado individualmente além das variáveis associadas às condições climáticas. Observamos uma associação significativa positiva entre a taxa de mortalidade e concentrações de NO₂, PART, SO₂ quando incluídos individualmente na presença das variáveis climáticas. No modelo que inclui todos os poluentes, observamos uma associação positiva apenas entre a taxa de mortalidade e a concentração de SO₂. Além disso, podemos verificar que existe efeito significativo das estações do ano. Este resultado é semelhante àquele do estudo realizado por Singer e Conceição (1994) utilizando regressão linear simples.

TABELA 4.2-COEFICIENTES ESTIMADOS E ESTIMATIVAS DOS RESPECTIVOS ERROS PADRÕES

Variáveis Independentes	Modelos de Poisson								Modelos de Zeger							
	1	2	3	4	5	6	7	8	1	2	3	4	5	6	7	8
INTERCEPTO	4,024	4,063	4,054	4,041	4,081	4,072	4,162	4,043	3,996	4,077	4,062	4,058	4,087	4,074	4,142	4,017
e p	0,063	0,012	0,012	0,012	0,012	0,013	0,041	0,042	0,075	0,017	0,018	0,017	0,018	0,019	0,047	0,055
COEFICIENTES (10 ⁻³)																
NO2	0,063 0,054	0,188 0,048							0,033 0,067	0,145 0,060						
PART	-0,043 0,225		0,711 0,154						0,106 0,296		0,688 0,204					
SO2	3,252 0,608			3,031 0,427					2,137 0,784			2,391 0,553				
O3	-0,007 0,098				0,114 0,094				0,044 0,125				0,126 0,123			
CO	-3,576 2,137					3,362 1,813			-1,071 2,780					4,388 2,410		
UMID	-0,513 0,627						-1,119 0,617		-0,278 0,699						-0,699 0,695	
TEMP	3,593 2,486							2,838 2,451	4,559 3,187							4,808 3,188
DIASF	32,833 13,962	22,002 10,000	21,621 9,955	19,908 9,900	18,144 9,999	19,492 10,034	17,200 9,912	27,605 13,810	30,006 16,871	13,069 13,170	14,555 13,137	13,224 12,973	10,839 13,219	12,265 13,216	9,832 13,142	25,474 16,958
DIASQ	24,329 12,525	34,030 9,515	34,012 9,511	36,904 9,506	36,250 9,508	35,602 9,503	34,758 9,522	26,441 12,502	14,801 14,930	26,590 12,143	25,692 12,122	28,384 12,026	28,175 12,165	27,149 12,157	27,544 12,173	14,560 15,041
DIASS	11,564 12,550	27,232 9,377	22,552 9,549	19,740 9,437	33,324 9,233	32,347 9,259	18,951 12,442	33,711 9,216	10,319 14,175	18,964 11,282	15,700 11,348	15,333 11,250	22,366 11,199	21,067 11,210	14,146 14,208	22,396 11,180
DIASU	0,603 16,543	-18,687 9,144	-14,113 9,262	-12,376 9,207	-18,767 9,465	-21,846 9,106	2,921 16,432	-21,459 9,115	-8,765 17,819	-20,795 10,469	-17,221 10,537	-16,307 10,504	-19,110 10,824	-22,055 10,454	-7,379 17,796	-21,543 10,466
PRI	90,497 10,981	94,883 10,955	90,490 11,040	91,546 10,967	97,541 10,934	96,979 10,938	98,082 10,939	96,668 10,956	90,594 17,966	94,888 18,269	89,788 18,339	92,037 17,849	97,001 18,419	96,143 18,438	97,274 18,380	96,126 18,470
OUT	125,684 11,760	134,419 11,148	122,744 11,440	120,704 11,323	136,191 11,226	131,138 11,298	135,027 11,149	135,806 11,196	124,713 18,700	131,346 18,441	120,250 18,707	120,405 18,190	133,121 18,698	127,083 18,764	131,748 18,572	134,558 18,779
INV	223,670 13,666	232,973 12,454	220,948 13,102	218,480 12,667	243,908 12,188	235,802 12,406	243,180 12,176	246,813 12,561	228,924 21,088	234,824 19,756	220,661 20,523	222,560 19,630	243,251 19,688	236,924 19,898	242,492 19,632	250,681 20,465

A Tabela 4.3 apresenta a análise de desvios (“deviances”) correspondentes. Para avaliar o ajuste de um determinado modelo, consideraremos a diferença entre o desvio desse modelo e aquele do primeiro modelo que possui como variáveis explicativas as Condições Climáticas e a concentração de SO₂.

TABELA 4.3-ANÁLISE DE DESVIOS

Termos do Modelo	Desvios	graus de liberdade	Diferença dos desvios	graus de liberdade	nível descritivo (p-value)
Condições Climáticas + SO ₂	1584.795	1086			
Condições Climáticas + SO ₂ +NO ₂	1583.710	1085	1.085	1	0.298
Condições Climáticas + SO ₂ +PART	1584.752	1085	0.043	1	0.836

Observando a tabela, verificamos que a inclusão no modelo das concentrações de NO₂ e de material particulado não é significativa; em função disto, analisaremos detalhadamente o modelo que possui somente o SO₂ como variável explicativa além das Condições Climáticas. Visto que um dos pressupostos da regressão de Poisson é a independência entre as observações da variável dependente, aplicamos o teste de Box-Pierce (Morettin e Toloí, 1987) aos resíduos do modelo ajustado para verificar se eles são não correlacionados. Os resultados apresentados na Tabela 4.4, sugerem que os resíduos são autocorrelacionados. Numa tentativa de eliminar essa autocorrelação, incluímos no modelo um termo associado a uma possível tendência linear e os seguintes pares de variáveis: $\cos(\omega t)$, $\sin(\omega t)$ e $\cos(2\omega t)$, $\sin(2\omega t)$, onde ω é dado por $2\pi/365$. Esses pares de variáveis explicativas tendem a diminuir a magnitude dos resíduos e, em geral, reduzem as autocorrelações residuais.

TABELA 4.4 - ESTATÍSTICAS PARA AVALIAÇÃO DO AJUSTE DOS MODELOS

Termos do Modelo	Desvios	graus de liberdade	Diferença dos desvios	nível descritivo (p-value)	Estatística de Box-Ljung	graus de liberdade	nível descritivo (p-value)
Condições Climáticas							
+ SO ₂	1584,795	1086		0,000	870,258	16	0,000
+ cos(ωt)+sen(ωt)	1463,515	1084	121,280	0,000	609,733	14	0,000
+ cos(2 ωt)+sen(2 ωt)	1460,239	1082	3,276	0,194	601,309	12	0,000
+ t	1321,200	1081	139,039	0,000	130,745	11	0,000

Nessa tabela, pode-se notar que tanto o efeito anual ($\cos(\omega t)$, $\sin(\omega t)$) e de tendência (t) no modelo são bastante significativos ($p < 0,001$), o mesmo não ocorrendo com o efeito semestral ($\cos(2\omega t)$, $\sin(2\omega t)$) para o qual $p = 0,194$. A inclusão desses termos também acarreta uma sensível redução nas autocorrelações residuais. Entretanto, essa redução não é suficiente para tornar os resíduos não correlacionados.

As Figuras 2.1 e 2.2 apresentam as funções de autocorrelação parcial e autocorrelação para os resíduos do ajuste do modelo que inclui somente a variável concentração de SO₂ além das condições climáticas. No gráfico apresentamos os intervalos de confiança de 95% de forma a testar a hipótese de inexistência de autocorrelações e autocorrelações parciais.

FIGURA 2.1- AUTOCORRELAÇÃO PARCIAL DO MODELO DE POISSON COM TERMO SO2

FIGURA 2.2- AUTOCORRELAÇÃO DO MODELO DE POISSON COM TERMO SO2

Podemos observar que as autocorrelações residuais decrescem suavemente para zero, o que sugere, que a suposição de independência do modelo de regressão Poisson pode não ser

verdadeira. Como o intuito deste trabalho é ajustar um modelo com uma estrutura de correlação simples para facilitar os cálculos computacionais, consideraremos o enfoque de Zeger, apresentado no Capítulo 3 com uma estrutura de autocorrelação autoregressiva de ordem 1 (AR(1)) para o processo latente auto-regressivo. Inicialmente consideraremos modelos semelhantes àqueles indicados na Tabela 4.2, isto é, sem os componentes sazonais e de tendência linear. Estimativas dos coeficientes dos termos desses modelos e seus erros padrões encontram-se na Tabela 4.2. Nessa tabela, os coeficientes significativos com um nível de 5%, foram assinalados em negrito. Podemos observar que os dois enfoques produzem resultados coerentes. Os coeficientes obtidos sob o enfoque de Zeger para os modelos 1, 2, 3 e 4 têm magnitudes ligeiramente menores e erros padrões um pouco maiores do que aqueles obtidos sob o enfoque de regressão Poisson. A Tabela 4.5 apresenta os níveis descritivos dos testes de significância para os coeficientes dos modelos 1, 2, 3 e 4. Podemos observar que a importância dos efeitos das concentrações de NO₂ ou material particulado na presença das condições climáticas e de sobre a mortalidade de pessoas acima de 65 anos diminui ligeiramente para os modelos ajustados sob o enfoque de Zeger. O mesmo ocorre para a concentração de SO₂ na presença de demais agentes poluentes.

TABELA 4.5- NÍVEIS DESCRITIVOS PARA TESTES DE SIGNIFICÂNCIA DOS COEFICIENTES DOS MODELOS 2, 3 E 4

Termos do modelo na presença de condições climáticas	Modelo de Poisson	Modelo de Zeger
	nível descritivo	
NO ₂	0,000	0,016
PART	0,000	0,001
SO ₂	0,000	0,000
SO ₂ (modelo completo)	0,000	0,006

Ainda sob o enfoque de Zeger, acrescentamos os efeitos anual, semestral e de tendência linear ao modelo 4. Os coeficientes associados à concentração de SO₂ após a inclusão desses efeitos encontram-se na Tabela 4.6. Podemos notar que analogamente ao caso anterior os coeficientes do modelo estimados sob o enfoque de Zeger têm magnitudes ligeiramente menores com erros padrões maiores do que sob o enfoque de regressão Poisson. Entretanto, a importância da concentração de SO₂ na explicação do número de óbitos de pessoas com mais de 65 anos é semelhante sob os dois enfoques. Observamos ainda que, tanto para o enfoque de Poisson como para o de Zeger, na medida que incluímos efeitos sazonais e de tendência linear, a importância dos agentes poluentes no modelo diminui. Isto pode ser uma indicação de que como o comportamento do número de óbitos para pessoas acima de 65 anos é sazonal, parte de sua sazonalidade é explicada através da sazonalidade da concentração de SO₂. Entretanto, quando incluímos os outros efeitos sazonais, a importância dessa concentração no modelo diminui.

TABELA 4.6 - ESTIMATIVAS DOS COEFICIENTES DE SO₂ SOB O ENFOQUE DE ZEGER E POISSON NA PRESENÇA DE EFEITOS SAZONAIS E DE TENDÊNCIA LINEAR

Termos do modelo na presença de condições climáticas	Modelo de Poisson			Modelo de Zeger		
	Estimativa do coeficiente de SO ₂	Erro Padrão	nível descritivo	Estimativa do coeficiente de SO ₂	Erro Padrão	nível descritivo
	(10 ⁻³)			(10 ⁻³)		
SO2	3,031	0,427	0,000	2,391	0,553	0,000
+cos(wt)+sen(wt)	2,353	0,434	0,000	1,998	0,539	0,000
+cos (2wt)+sen(2wt)	2,272	0,437	0,000	1,951	0,541	0,000
+t	1,979	0,438	0,000	1,823	0,510	0,000

O valor da estatística do teste de Box-Ljung para o modelo completo sob o enfoque de Zeger foi de 148,9803 com 16 graus de liberdade ($p < 0,001$), sugerindo ainda a existência de resíduos autocorrelacionados. Isto pode ser explicado pelo fato de termos

utilizado uma estrutura AR(1) quando na verdade seria necessário utilizarmos um modelo com estrutura autoregressiva de ordem maior. Podemos verificar através do periodograma dos resíduos Figura 3 do modelo completo sob o enfoque de Zeger, que aparentemente não há nenhuma periodicidade importante a ser incorporada no modelo.

FIGURA 3- PERIODOGRAMA DOS RESÍDUOS DO MODELO COMPLETO SOB ENFOQUE DE ZEGGER

As estimativas dos erros padrões dos coeficientes estimados foram baseados na inversa da matriz apresentada em (3.15) uma vez que a proposta de Campbell (1993) de estimar V através de (3.18) apresentou resultados instáveis no exemplo analisado aqui. Esperamos que esta aproximação seja razoável, pois para um número grande de observações, é bastante provável que a matriz V_R esteja muito próxima de V . Como o tamanho da série dos dados é da ordem de 1000, também não foi possível estimar a matriz de covariâncias das observações através de (3.7) por problemas computacionais. Utilizamos um computador PC com processador 486 e memória RAM de 16 Mbytes, que não é o suficiente para armazenar uma matriz com essa dimensão. Entretanto, dividindo a série

completa em três séries anuais, esses problemas computacionais podem ser contornados. Isto permite a comparação entre os resultados das seguintes aproximações para a estimação da matriz de covariância em questão:

Alternativa 1: $V(\hat{\beta}) = I_o^{-1}$

Alternativa 2: $V(\hat{\beta}) = I_o^{-1} I_1 I_o^{-1}$ onde $I_1 = (AX)' V_R^{-1} (A + \sigma^2 AR_e A) V_R^{-1} (AX)$

Alternativa 3: $V(\hat{\beta}) = [(AX)' (A + \sigma^2 AR_e A)^{-1} (AX)]^{-1}$.

Para a comparação dessas alternativas, usamos o modelo que inclui as Condições Climáticas, a concentração de SO₂, os efeitos semestrais e a tendência linear. Procuramos ajustar um efeito trimestral, mas não houve evidências suficientes para a sua inclusão. Os resultados estão resumidos na Tabela 4.7.

TABELA 4.7 - COMPARAÇÃO DOS ERROS PADRÕES ESTIMADOS DO COEFICIENTE DE SO_2

ANO	Desvio padrão estimado do coeficiente do termo SO_2 (10^{-3})		
	Alternativas		
	1	2	3
1991	0.850	0.016	0.999
1992	0.909	0.012	0.869
1993	0.805	0.009	0.820

Analisando as diferentes alternativas, observamos que as alternativas 1 e 3 apresentam resultados coerentes. A alternativa sugerida por Campbell (1993) em (3.18) difere bastante das demais sugerindo que a aproximação de V por V_R não parece adequada quando usamos uma estrutura de dependência autoregressiva de ordem 1.

Capítulo 5

Implementação computacional e conclusões

O enfoque de Zeger para ajuste de modelos parece ser adequado para séries que não apresentem grandes sazonalidade e tendência linear, uma vez que a estrutura de dependência do processo latente $\{\varepsilon_t\}$ é de um processo autoregressivo estacionário. Como pudemos verificar, na medida que foram introduzidas variáveis associadas aos efeitos sazonais, a importância da concentração de SO_2 diminui, sugerindo que o mesmo apresenta um ciclo semelhante ao da mortalidade para pessoas acima de 65 anos, porém não na mesma fase. O fato de termos utilizado um modelo autoregressivo de ordem 1 para a estrutura de dependência do processo latente, quando na realidade haveria necessidade de um modelo com ordem maior, pode não ter eliminado ou reduzido satisfatoriamente a autocorrelação residual. Uma alternativa que consideraremos em futuras pesquisas é utilizar filtros autoregressivos de ordens maiores.

Durante o processo iterativo, ocasionalmente $\hat{\rho}_v^{(j)}(\tau)$, dada por (3.20) do processo AR(1) se torna maior que 1. Neste caso, foram utilizadas duas alternativas: utilização do cálculo alternativo proposto por Campbell (1993) para a estimativa de $\bar{\sigma}^2$ na equação de cálculo de (3.19) ou baseada no resíduo de Pearson (Morettin e Toloi, 1987). Optamos pela segunda alternativa dado que sob a primeira proposta não houve convergência em alguns casos.

Para séries não muito longas, com cerca de 200 observações ou menos, seria possível construir uma matriz de correlação com estrutura autoregressiva de ordem maior que 1, e realizar a decomposição de Cholesky para a construção de filtros autoregressivos.

Isto permitiria usarmos a regressão sob o enfoque de Zeger com estruturas de dependência de ordens maiores.

Apêndice

A.1 Modelos Lineares Generalizados

Consideremos, o modelo linear clássico

$$y_T = \mathbf{x}_T^t \boldsymbol{\beta} + \varepsilon_T, \quad T = 1, \dots, N$$

onde ε_T são erros aleatórios, tais que $E(\varepsilon_T) = 0$ e $\text{Var}(\varepsilon_T) = \sigma^2$. Então $\mu_T = E(y_T) = \mathbf{x}_T^t \boldsymbol{\beta}$ e o valor de $\boldsymbol{\beta}$ que minimiza a função

$$Q(\boldsymbol{\beta}) = (\mathbf{Y} - \mathbf{X} \boldsymbol{\beta})^t (\mathbf{Y} - \mathbf{X} \boldsymbol{\beta}) \quad (\text{A.1.1})$$

é o estimador de mínimos quadrados de $\boldsymbol{\beta}$. Este estimador, denotado por $\hat{\boldsymbol{\beta}}$, é a solução do sistema de equações obtido quando igualamos a zero a derivada de (A.1.1) em relação a $\boldsymbol{\beta}$, ou seja,

$$\mathbf{X}^t (\mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}) = \mathbf{0}. \quad (\text{A.1.2})$$

As equações (A.1.2) são chamadas de **equações de estimação de mínimos quadrados**. O estimador $\hat{\boldsymbol{\beta}}$ é não viesado, isto é, $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ e sua matriz de covariância é dada por

$$\text{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{X}^t \mathbf{X})^{-1} \quad (\text{A.1.3})$$

Além disto, se os erros ε_T tiverem distribuição normal, $\hat{\boldsymbol{\beta}}$ terá distribuição normal multivariada. O conhecimento desta distribuição permite a construção de intervalos de confiança e testes de hipóteses para os coeficientes de regressão.

As equações (A.1.2) podem ser generalizadas para comportar situações onde as observações têm variâncias desiguais, isto é, $\text{Var}(\varepsilon_T) = w_T \sigma^2$ com w_T denotando uma constante positiva conhecida. Neste caso, o valor de β que minimiza a função

$$Q_w(\beta) = (Y - X\beta)' W^{-1}(Y - X\beta)$$

onde $W = \text{diag}(w_1, \dots, w_N)$, é o estimador de mínimos quadrados ponderados de β e corresponde à solução das **equações de mínimos quadrados ponderados**

$$X'W^{-1}(Y - X\hat{\beta}_w) = 0 \quad (\text{A.1.4})$$

A solução $\hat{\beta}_w$ das equações de mínimos quadrados ponderados também é não viesada com

$$\text{Cov}(\hat{\beta}_w) = \sigma^2 (X'W^{-1}X)^{-1}.$$

Se os erros ε_T tiverem distribuição normal, $\hat{\beta}_w$ terá distribuição normal multivariada.

A denominação mínimos quadrados ponderados refere-se aos pesos correspondentes à matriz diagonal W^{-1} usada na estimação de β . As equações de estimação (A.1.2) representam um caso especial de (A.1.4) quando $W = I$. No caso mais geral em que as observações são correlacionadas, é possível generalizar as equações (A.1.4). Nesse caso W é substituída por uma matriz não diagonal V e o estimador correspondente é denominado de **mínimos quadrados generalizados**.

O modelo linear clássico possui suas limitações. Entre eles destacamos o fato de o intervalo de variação de μ_T ser irrestrito; na prática, muitas situações requerem intervalos estritamente não-negativos (por exemplo, em modelos para o tempo de sobrevivência ou contagens) ou ainda restrito ao intervalo $[0, 1]$ (como nos casos de dados binários e

modelagem de proporções). Além disso, o modelo linear assume que os efeitos das covariáveis são aditivos; em muitos casos, efeitos multiplicativos podem ser de interesse, como na modelagem de razões de chances (odds ratio) em Epidemiologia. Finalmente, as inferências sob o modelo linear clássico são usualmente baseadas na normalidade da distribuição dos erros, suposição que pode não ser válida para muitas situações. Os Modelos Lineares Generalizados (MLG) não exigem essa suposição.

Um dos componentes chave desses modelos é uma função monotônica diferenciável denominada **função ligação**, que relaciona μ_T com as variáveis explicativas na forma

$$g(\mu_T) = \mathbf{x}_T^t \boldsymbol{\beta}.$$

No modelo linear descrito acima, $g(\cdot)$ é a função identidade. No caso da modelagem de proporções é comum utilizar $g(\mu_T) = \log(\mu_T / 1 - \mu_T) = \mathbf{x}_T^t \boldsymbol{\beta}$. Consequentemente, $\mu_T = \exp(\mathbf{x}_T^t \boldsymbol{\beta}) / (1 + \exp(\mathbf{x}_T^t \boldsymbol{\beta}))$ o que restringe μ_T a $[0,1]$.

Os MLG também não requerem que as variâncias dos y_T 's sejam constantes. É necessário assumir apenas que a variância de y_T seja uma função conhecida de μ_T ou mais explicitamente, que $\text{Var}(y_T) = \sigma^2 V(\mu_T)$ onde σ^2 (**parâmetro de escala**) pode ser desconhecido e $V(\cdot)$ é uma função conveniente. No caso de variáveis binárias, onde μ_T representa a probabilidade de sucesso, a variância é dada por $V(\mu_T) = \mu_T (1 - \mu_T)$ e, nesse caso, $\sigma^2 = 1$. Na família exponencial é sempre possível especificar a função de variância.

Para observações independentes, a matriz de covariância de $\mathbf{Y}$ é dada por $\sigma^2 \mathbf{V}(\boldsymbol{\mu})$ onde $\mathbf{V}(\boldsymbol{\mu}) = \mathbf{V} = \text{diag}[V(\mu_1), \dots, V(\mu_N)]$.

O estimador de $\boldsymbol{\beta}$ sob um MLG pode ser obtido através da minimização da função

$$Q_{\text{MGL}}(\boldsymbol{\beta}) = (\mathbf{Y} - \boldsymbol{\mu}(\boldsymbol{\beta}))^t [\mathbf{V}(\boldsymbol{\mu}(\boldsymbol{\beta}))]^{-1} (\mathbf{Y} - \boldsymbol{\mu}(\boldsymbol{\beta}))$$

onde $\mu(\beta) = g^{-1}(X\beta)$. Essa minimização pode ser concretizada derivando $Q_{MGL}(\beta)$ em relação a β e igualando o resultado a zero, obtendo as equações

$$U(\hat{\beta}) = D(\hat{\beta})' [V(\mu(\hat{\beta}))]^{-1} (Y - \mu(\hat{\beta})) = 0 \quad (A.1.5)$$

onde

$$D(\hat{\beta}) = \frac{\partial \mu(\hat{\beta})}{\partial \beta} = \begin{bmatrix} \frac{\partial \mu_1(\hat{\beta})}{\partial \beta_1} & \dots & \frac{\partial \mu_1(\hat{\beta})}{\partial \beta_p} \\ \vdots & \ddots & \vdots \\ \frac{\partial \mu_N(\hat{\beta})}{\partial \beta_1} & \dots & \frac{\partial \mu_N(\hat{\beta})}{\partial \beta_p} \end{bmatrix}.$$

As equações de estimação de mínimos quadrados ponderados (A.1.4) constituem um caso particular das equações de estimação dos MLG com $g(\cdot)$ dada pela função identidade, ou seja, com $\mu = X\beta$ e com a matriz de ponderação W independente de β . As equações (A.1.5) requerem uma solução iterativa (Green, 1984). McCullagh e Nelder (1989) sugerem o uso de um algoritmo de mínimos quadrados ponderados iterativos para resolvê-las. Uma justificativa para esse algoritmo está apresentada em Sen and Singer (1993) e está reproduzida a seguir.

A estatística $U(\beta)$ em (A.1.5) corresponde à derivada do logaritmo da verossimilhança em relação a β . A matriz de Informação de Fisher correspondente é dada por

$$I(\beta) = -E\left(\frac{\partial U(\mu(\beta))}{\partial \beta}\right) = D'(\beta) [V(\mu(\beta))]^{-1} D(\beta) \quad (A.1.6)$$

(McCullagh e Nelder, 1989). Os procedimentos computacionais para a resolução de (A.1.5) são baseados numa expansão em série de Taylor de $U(\beta)$, ou seja,

$$U(\beta^{(k+1)}) \cong U(\beta^{(k)}) + \frac{\partial U(\beta^{(k)})}{\partial \beta} (\beta^{(k+1)} - \beta^{(k)})$$

o que implica

$$\beta^{(k+1)} = \beta^{(k)} + \mathbf{I}^{-1}(\beta^{(k)}) U(\beta^{(k)}). \quad (\text{A.1.7})$$

Multiplicando ambos os termos por $\mathbf{I}(\beta^{(k)})$, temos

$$\mathbf{I}(\beta^{(k)}) \beta^{(k+1)} = \mathbf{I}(\beta^{(k)}) \beta^{(k)} + U(\beta^{(k)}). \quad (\text{A.1.8})$$

Utilizando (A.1.6), o segundo membro de (A.1.8) pode ser reescrito como

$$\mathbf{D}'(\beta^{(k)}) [\mathbf{V}(\mu(\beta^{(k)}))]^{-1} \mathbf{D}(\beta^{(k)}) \beta^{(k)} + \mathbf{D}'(\beta^{(k)}) [\mathbf{V}(\mu(\beta^{(k)}))]^{-1} (\mathbf{Y} - \mu(\beta^{(k)})). \quad (\text{A.1.9})$$

Então, considerando

$$\begin{aligned} \mathbf{W}(\beta^{(k)}) &= \mathbf{diag}[w_1(\beta^{(k)}), \dots, w_N(\beta^{(k)})] \\ \dot{\mathbf{g}}(\beta^{(k)}) &= \mathbf{diag}[\dot{g}_1(\beta^{(k)}), \dots, \dot{g}_N(\beta^{(k)})] \end{aligned}$$

onde $w_i(\beta^{(k)}) = \left\{ \left[g'(\mu_i(\hat{\beta}^{(k)})) \right]^2 \left[v_i(\hat{\beta}^{(k)}) \right] \right\}^{-1}$ e $\dot{g}_i(\beta^{(k)}) = [g'(\mu_i(\beta^{(k)}))]$ e observando que

$$D(\beta) = \frac{\partial \mu(\beta^{(k)})}{\partial \beta} = X \dot{g}(\beta^{(k)})$$

podemos reescrever (A.1.9) como

$$X^t W(\beta^{(k)}) X \beta^{(k)} + X^t W(\beta^{(k)}) \dot{g}(\beta^{(k)}) (Y - \mu(\beta^{(k)}))$$

que ainda pode ser reexpressa como

$$X^t W(\beta^{(k)}) [X \beta^{(k)} + \dot{g}(\beta^{(k)}) (Y - \mu(\beta^{(k)}))] \quad (A.1.10)$$

Por outro lado, consideremos a seguinte aproximação baseada numa expansão de Taylor de primeira ordem para $g(Y)$

$$g(Y) \cong g(\mu(\beta^{(k)})) + \dot{g}(\beta^{(k)}) (Y - \mu(\beta^{(k)}))$$

Consequentemente, (A.1.10) pode ser escrita como

$$X^t W(\beta^{(k)}) g(Y)$$

Finalmente, (A.1.7) será dada por

$$\beta^{(k)} = (X^t W(\beta^{(k-1)}) X)^{-1} X^t W(\beta^{(k-1)}) g(Y) \quad (A.1.11)$$

Nelder e Wedderburn (1972), McCullagh e Nelder (1989), e Dobson (1990) justificam este método iterativo ao mostrar a sua equivalência com o método “scoring” de Fisher. Sen and Singer (1993) mostram que sob certas condições, a solução $\hat{\beta}$ para as

equações de estimação (A.1.5) é assintoticamente normal multivariada com vetor de médias β e matriz de covariância

$$V(\hat{\beta}) = \sigma^2 \{D'(\hat{\beta}) [V(\hat{\beta})]^{-1} D(\hat{\beta})\}^{-1} \quad (A.1.12)$$

Consideremos uma regressão Poisson para ilustrar a aplicação de um MLG ao estudo de N contagens independentes. Seja y_T a variável resposta para a qual assumimos uma distribuição de Poisson. Considerando uma função de ligação logarítmica, ou seja

$$\mu_T = \exp(\mathbf{x}_T' \beta), \quad (A.1.13)$$

as equações (A.1.5) se reduzem a

$$\begin{bmatrix} \sum_{i=1}^N x_{i1}(y_i - \mu_i(\beta)) \\ \vdots \\ \sum_{i=1}^N x_{ip}(y_i - \mu_i(\beta)) \end{bmatrix} = \mathbf{0}$$

pois

$$D(\beta)' = \begin{bmatrix} x_{11}\mu_1(\beta) & \dots & x_{N1}\mu_N(\beta) \\ \vdots & \ddots & \vdots \\ x_{1p}\mu_1(\beta) & \dots & x_{Np}\mu_N(\beta) \end{bmatrix} \quad \text{e} \quad [V(\mu(\beta))]^{-1} = \begin{bmatrix} \frac{1}{\mu_1(\beta)} & & 0 \\ & \ddots & \\ 0 & & \frac{1}{\mu_N(\beta)} \end{bmatrix}$$

Essas equações são não lineares nos β_i 's, mas podem ser linearizadas se usarmos a função de ligação para obter as variáveis transformadas, ou seja, $z_T = g(y_T) = \log(y_T)$. Usando uma aproximação de Taylor de primeira ordem teremos

$$E(z_T) \cong \mathbf{x}_T' \boldsymbol{\beta} \text{ e } \text{Var}(z_T) \cong 1/\mu_T$$

Logo, a nova variável $\mathbf{z}' = (z_1, \dots, z_N)$ terá a matriz de covariâncias dada por $\mathbf{V} = \mathbf{diag}[\text{Var}(z_1), \dots, \text{Var}(z_N)]$. Neste caso as equações (A.1.5) se reduzem a

$$\mathbf{X}' \mathbf{V}^{-1} (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}) = 0 \quad (\text{A.1.14})$$

onde $\mathbf{Z} = \mathbf{g}(\mathbf{Y})$.

A.2 Método de quasi-verossimilhança para estimação

Quando não há informações suficientes sobre a distribuição de uma variável para a construção da sua função de verossimilhança, podemos recorrer à Teoria da Quasi-verossimilhança desenvolvida por Wedderburn (1974), e analisada posteriormente com mais detalhes por McCullagh (1983). A principal ideia por trás dessa teoria é evitar a especificação completa da distribuição da variável e basear os métodos de inferência apenas nos dois primeiros momentos. Para isto, especifica-se apenas uma relação entre a variância e a média da variável resposta, a chamada **função de variância**, denotada por $V(.)$. Desta forma, para o vetor de observações $\mathbf{Y}$ teremos:

$$\text{Var}(\mathbf{Y}) = \phi \mathbf{V}(\mu) \quad (\text{A.2.1})$$

onde ϕ é tratado como um **parâmetro de perturbação**.

Consideremos inicialmente apenas um componente y_T de $\mathbf{Y}$; então

$$U_T = u_T(\mu_T; y_T) = \frac{y_T - \mu_T}{\phi V(\mu_T)}$$

possui as seguintes propriedades semelhantes à derivada do logaritmo da função de verossimilhança, ou seja

$$\begin{aligned} E(U_T) &= 0 \\ \text{Var}(U_T) &= 1 / \{ \phi V(\mu_T) \} \\ -E\left(\frac{\partial U_T}{\partial \mu_T} \right) &= 1 / \{ \phi V(\mu_T) \} \end{aligned} \quad (\text{A.2.2})$$

Como a teoria assintótica de primeira ordem associada à função de verossimilhança baseia-se nas propriedades (A.2.2), por extensão, se a integral

$$Q(y_T, \mu_T) = \frac{1}{\phi} \int_{y_T}^{\mu_T} (y_T - s) V'(s)^{-1} ds \quad (\text{A.2.3})$$

existir, esta deverá se comportar de forma semelhante à função logarítmica da verossimilhança. A integral acima é chamada de **função logaritmo de quasi-verossimilhança**. Em situações nas quais $V(\mu_T) = k, \mu_T, \mu_T(1 - \mu_T), \mu_T + \mu_T^2 / k$ e μ_T^3 , (sendo k uma constante), as integrais resultantes de (A.2.3) correspondem aos logaritmos das funções de verossimilhança das distribuições Normal, Poisson, Binomial, Binomial Negativa e Normal Inversa, a menos de constantes. Logo, para essas funções de variância, os modelos de quasi-verossimilhança são equivalentes aos MLG.

Supondo que os componentes de $\mathbf{Y}$ são independentes entre si, a função de quasi-verossimilhança é dada pela soma das contribuições individuais

$$Q(\mu, \mathbf{Y}) = \sum_{T=1}^N Q(y_T, \mu_T).$$

Diferenciando $Q(\mu, \mathbf{Y})$ em relação a μ temos

$$\frac{\partial Q(\mu, \mathbf{Y})}{\partial \mu} = \mathbf{V}^{-1}(\mu) (\mathbf{Y} - \mu). \quad (\text{A.2.4})$$

Se considerarmos o modelo $\mu = \mu(\beta)$, (A.2.4) pode ser reescrita como

$$U(\mu(\beta), \mathbf{Y}) = \frac{\partial Q[\mu(\beta)]}{\partial \beta} = \frac{\partial \mu'(\beta)}{\partial \beta} \mathbf{V}^{-1}[\mu(\beta)] [\mathbf{Y} - \mu(\beta)] / \phi \quad (\text{A.2.5})$$

que é considerada como a função de “quasi-score” da quasi-verossimilhança e tem as seguintes propriedades (McCullagh, 1983):

$$(a) E(U(\mu(\beta), Y)) = \mathbf{0} ;$$

$$(b) \text{Cov} (U(\mu(\beta), Y)) = \mathbf{I}_\beta; \quad (A.2.6)$$

$$(c) -E\left(\frac{\partial U(\mu(\beta), Y)}{\partial \beta}\right) = \mathbf{I}_\beta;$$

onde $-\mathbf{I}_\beta = \frac{\partial \mu^t(\beta)}{\partial \beta} \mathbf{V}^{-1} \frac{\partial \mu(\beta)}{\partial \beta}$ desempenha o mesmo papel que a matriz de informação de

Fisher para funções de verossimilhança usuais, ou seja, $[\mathbf{I}_\beta]^{-1}$ corresponde à $\text{Var}(\beta)$.

Os estimadores de quasi-verossimilhança são obtidos através da resolução das equações de estimação dadas por

$$U(\mu(\hat{\beta}), Y) = \mathbf{0}. \quad (A.2.7)$$

As equações acima são chamadas de **equações de “quasi-score”** para β . Quando Y segue uma distribuição da família exponencial, (A.2.7) é equivalente às equações de “score” para β nos MLG. O estimador $\hat{\beta}$ é calculado através do método de mínimos quadrados ponderados iterativos de forma análoga ao caso dos Modelos Lineares Generalizados. Demonstra-se que o estimador resultante é assintoticamente normal sob certas condições de regularidade (Mc Cullagh e Nelder, 1989).

Para observações independentes da variável resposta, teremos $\mathbf{V} = \text{diag}\{V(\mu_T)\}$. Entretanto, se desejarmos incluir alguma estrutura de dependência no modelo, a matriz $\mathbf{V}$

terá elementos não nulos fora da diagonal principal e algumas adaptações (discutidas na próxima seção) são necessárias.

A.3 Modelos de quasi-verossimilhança para dados correlacionados

Se os dados forem dependentes, a matriz V não será diagonal. Neste caso, as equações de estimação (A.2.5) podem incluir modelos com correlações entre as observações conforme sugerem Liang e Zeger (1986). Consideremos unidades amostrais rotuladas por $i = 1, \dots, L$ e observadas nos tempos $t = 1, \dots, T_i$, resultando um total de $N = \sum_{i=1}^L T_i$ observações. O vetor de respostas para a i -ésima unidade amostral é dado por $y_i = (y_{i1}, \dots, y_{iT_i})^t$ e as informações das covariáveis associadas a ele são resumidas no vetor $x_i = (x_{i1}, \dots, x_{iT_i})^t$. Embora as respostas de uma mesma unidade amostral possam ser correlacionadas, assumimos que as respostas das diferentes unidades amostrais são independentes. A matriz de covariância de y_i tem a forma $\sigma^2 V_i$ onde

$$V_i = \{ \text{diag}[V(\mu_{i1}), \dots, V(\mu_{iT_i})] \}^{1/2} R_i \{ \text{diag}[V(\mu_{i1}), \dots, V(\mu_{iT_i})] \}^{1/2} \quad (\text{A.3.1})$$

e R_i é a matriz de correlação entre as respostas nos diferentes instantes. A matriz de correlação de todas as N observações $Y = (Y_1^t, \dots, Y_L^t)^t$ é uma matriz diagonal em blocos $\sigma^2 V = \sigma^2 \text{diag}(V_1, \dots, V_L)$. Dada essa estrutura diagonal em blocos, as equações de estimação (A.2.5) podem ser reescritas na forma

$$\sum_{i=1}^L D_i^t(\hat{\beta})^t [V_i(\mu_i(\hat{\beta}))]^{-1} (Y_i - \mu_i(\hat{\beta})) = 0 \quad (\text{A.3.2})$$

onde $D_i^t(\hat{\beta}) = \partial \mu_i(\hat{\beta}) / \partial \beta$. Estas equações constituem um caso particular das chamadas **equações de estimação generalizadas (EEG)**. Elas têm forma idêntica às equações de estimação (A.2.5) exceto pelo fato de que a matriz V agora contém elementos não nulos

fora da diagonal principal. Os estimadores de β , obtidos como soluções das equações (A.3.2) através de métodos iterativos, têm distribuição assintoticamente normal multivariada com matriz de covariância

$$\begin{aligned} V(\hat{\beta}) &= \sigma^2 (\mathbf{D}'(\hat{\beta}) [\mathbf{V}(\mu(\hat{\beta}))]^{-1} \mathbf{D}(\hat{\beta}))^{-1} \\ &= \sigma^2 \left\{ \sum_{i=1}^L (\mathbf{D}_i'(\hat{\beta}) \mathbf{V}_i^{-1}(\mu(\hat{\beta})) (\mathbf{D}_i(\hat{\beta}))) \right\}^{-1} \end{aligned} \quad (\text{A.3.3})$$

sob certas condições de regularidade (Liang and Zeger, 1986).

Quando $\mathbf{R}_i$ é conhecida, é possível resolver as equações (A.3.2) através dos métodos de estimação iterativos já citados. No entanto, em geral a verdadeira estrutura de correlação (por exemplo, autoregressiva, correlação permutável, etc.) é desconhecida. Liang e Zeger (1986) usam um **estimador de "trabalho"** $\hat{\mathbf{R}}_i$ para aproximar essa estrutura de correlação em (A.3.1). As equações de estimação resultantes são dadas por

$$\sum_{i=1}^L \mathbf{D}_i'(\hat{\beta}) [\hat{\mathbf{V}}_i(\mu(\hat{\beta}))]^{-1} (\mathbf{Y}_i - \mu_i(\hat{\beta})) = \mathbf{0} \quad (\text{A.3.4})$$

onde $\hat{\mathbf{V}}_i^{-1} = \mathbf{V}_i(\hat{\mathbf{R}}_i)$ é um estimador de (A.3.1) baseado na matriz de correlação de trabalho $\hat{\mathbf{R}}_i$. Liang e Zeger (1986) propõem várias formas de estimação de $\hat{\mathbf{R}}_i$. O estimador de β obtido como solução das equações (A.3.2) tem distribuição assintótica normal multivariada sob certas condições (Liang e Zeger, 1986), com matriz de covariância dada por

$$v(\hat{\beta}) = \sigma^2 \left\{ \sum_{i=1}^L (\mathbf{D}_i' \mathbf{V}_i^{-1} \mathbf{D}_i) \right\}^{-1} \left\{ \sum_{i=1}^L \mathbf{D}_i' \mathbf{V}_i^{-1} \mathbf{V}(\mathbf{Y}_i) \mathbf{V}_i^{-1} \mathbf{D}_i \right\}^{-1} \left\{ \sum_{i=1}^L (\mathbf{D}_i' \mathbf{V}_i^{-1} \mathbf{D}_i) \right\}^{-1} \quad (\text{A.3.5})$$

onde $\mathbf{D}_i^t = \mathbf{D}_i^t(\beta)$. Quando não impomos nenhuma estrutura para $\mathbf{R}_i$ temos $\mathbf{V}(\mathbf{Y}_i) = \mathbf{V}_i(\mu(\beta))$ e a expressão (A.3.5) se reduz a (A.3.3). Uma característica importante da metodologia das EEG é de que tanto $\hat{\beta}$ quanto $\hat{\mathbf{V}}(\hat{\beta})$ são consistentes mesmo que a estrutura de correlação seja mal especificada. Desta forma, os intervalos de confiança para $\hat{\beta}$ serão assintoticamente corretos mesmo que a matriz de correlação de trabalho seja incorreta. Além disto, as EEG permitem analisar dados desbalanceados no tempo se as perdas de algumas observações forem aleatórias.

O processo iterativo para resolver as EEG (A.3.4) é semelhante àquele das equações de estimação para MLG.

A.4 Análise do número de óbito para pessoas acima de 65 anos com variáveis defasadas.

Conforme o Capítulo 4, as correlações cruzadas calculadas entre o número de óbitos para pessoas com mais de 65 anos e concentração de SO_2 indicaram uma forte correlação com defasagem de 0 e 3 dias. Já para o número de óbitos e temperatura verificamos uma significativa correlação com defasagem de 2 e 3 dias. Desta forma foram criadas duas novas variáveis para a concentração de SO_2 e temperatura defasadas de 3 dias. Antes de prosseguirmos com as análises usando as variáveis defasadas, recalculamos as correlações de Pearson entre os número de óbitos e esses dois agentes poluentes. Verificamos que a correlação de Pearson entre os óbitos para pessoas acima de 65 anos e da concentração de SO_2 defasada é de 0,1920, enquanto que a correlação sem a defasagem é de 0,3299, indicando que a defasagem nula (0 dias) é mais informativa que a de 3 dias. Já para a temperatura, verificamos que a correlação de Pearson com a defasagem é de -0,3697 enquanto que sem a defasagem é de -0,2365. Desta forma, realizamos análises semelhantes ao Capítulo 4 exceto pelo fato de utilizarmos a temperatura defasada de 3 dias ao invés da temperatura sem defasagem. Denotaremos por temperatura* à temperatura defasada.

Inicialmente ajustamos um modelo com o enfoque de Poisson com as variáveis de condições climáticas e todos os agentes poluentes. Verificamos que apenas a concentração de SO_2 e temperatura* foram altamente significantes. Como os resíduos desse modelo apresentam forte autocorrelação residual introduzimos tal qual no Capítulo 4 funções senoidais e cossenoidais e tendência. O mesmo foi realizado sob o enfoque de Zeger. Os resultados encontram-se na Tabela A4.1.

TABELA A 4.1 - ESTATÍSTICAS PARA AVALIAÇÃO DO AJUSTE DOS MODELOS

Termos do Modelo	Estatística de Box-Ljung (enfoque de Zeger)	Estatística de Box-Ljung (enfoque de Poisson)	graus de liberdade	nível descritivo modelo Zeger (p-value)	nível descritivo modelo Poisson (p-value)
Modelo 1=CC+SO2+TEMP*+ cos(wt)+sen(wt)	840,189	762,072	18	0,000	0,000
Modelo 2=Modelo 1+ t	114,955	115,273	17	0,000	0,000
Modelo 3=Modelo 2+cos(2wt)+sen(2wt)	114,853	115,391	15	0,000	0,000

Podemos verificar que a tendência é um componente importante do modelo sob os enfoques de Zeger e de Poisson, reduzindo consideravelmente o valor da estatística de Box-Ljung. Entretanto, autocorrelação residual sob os dois enfoques continua presente mesmo após a inclusão dos efeitos anuais e semestrais.

A.5 Programas utilizados para implementar o modelo de Zeger

Os programas abaixo utilizam-se da linguagem S-PLUS, versão 3.3. Nos programas 1 e 3 as matrizes de covariâncias são calculados utilizando-se respectivamente as alternativas 1 e 2 do Capítulo 4.

1.Programa Zeger.txt - Este programa executa a regressão sob o enfoque de Zeger, utilizando uma estrutura de dependência um autorregressivo de ordem 1. Os vetores e matrizes utilizados neste programa são denotados respectivamente com letras minúscula e maiúscula.

```
#=====#
#          Inicialização- Regressão de Poisson          #
#=====#
# Módulo 1_ Entrada de dados e Inicializações do Processo Recursivo #
# Entrada de Dados #
attach(obic)

# Valores iniciais da média e coeficientes obtidos através da regressão Poisson#
inic_glm(TOT65~DIASF+DIASQ+DIASS+DIASU+OUT+INV+PRI+TEMP+UMID+CO
+O3+SO2+PART+NO2,poisson,obic)

# armazenamento do resíduo (y-m), média e coeficientes#
#rp- resíduo#
# m- média#
# cf- coeficientes#
rp_as.vector(residuals(inic,"response"))
m_as.vector(fitted.values(inic))
cf_as.vector(coefficients(inic))
print(cf)

# Construção da Matriz de Especificação do modelo - Dimensão n x p #
```

```

CCLI_ cbind(1, obic$DIASF, obic$DIASQ, obic$DIASS, obic$DIASU, obic$OUT,
obic$INV,obic$PRI)
X_as.matrix(cbind(CCLI,
obic$TEMP,obic$UMID,obic$CO,obic$O3,obic$SO2,obic$PART,obic$NO2))

# Construção do Vetor de Respostas #
y_as.vector(obic$TOT65)

# Módulo 2 -Processo recursivo#
# Controle do processo recursivo #
# Número máximo de iterações #
iterac_30

#tamanho da série#
n_1095

#ordem do vetor de coeficientes #
p_(8+7)

#controle das iterações#
ni_1

repeat
{
#=====
#          Cálculo de sigma2          #
#=====
  rp2m_rp*rp-m
  m2_m*m
  srd_t(um)%*%(rp2m)
  sm_t(um)%*%m2
  sigm2_srd/sm
  print(sigm2)
  # Controle para valores negativos de sigma2#
  if (sigm2 < 0 ) {
    rp2_rp*rp
    sigm2_(t(um)%*%rp2)/(n-1)
    print("sigma2 negativo=")
    print(sigm2)
  }

#=====
#  Cálculo da matriz D = mu+sigma2*mu2  #
#=====

```

```

d_m+sigm2*m2
di_d^(-1)
invrd_sqrt(di)
DELM_m*X
DERIV_as.matrix(invrd* DELM)

#=====#
#  Cálculo de rho                                     #
#=====#
mat_cbind(rp,m)
lmat_rbind(0,mat)
elmat_lmat[1:n,1:2]
lrp_as.vector(elmat[,1])
lmu_as.vector(elmat[,2])
srpl_t(um)%*%( rp*lrp )
sml_t(um)%*%(m*lmu)
rho_srpl/(sml * sigm2)
print(rho)
# Controle para valores de |rho| > 1#
if (abs(rho)>1 ) {

    warning("RHO maior que um")
    rp2_sum((invrd*rp)*(invrd*rp))
    srpl_sum((invrd*rp)*(invrd*lrp))
    rho_srpl/rp2
    print( "rho alternativo = ")
    print(rho)

}

rho_prod(eigen(rho)$values)

# Filtragem de *D-1/2*(delm/delbeta) e D-1/2*z- Passo (II) do algoritmo#
#=====#
#  Cálculo da matriz W=L*D-1/2*(delm/delbeta)         #
#=====#
LDER_rbind(0, DERIV)
ELDER_LDER[1:n,1:p]
frho_rho/(sqrt(1-rho^2))
L1_DERIV[1,]
LR_(1/sqrt(1-rho^2))*DERIV[2:n,1:p]
LDERIV_rbind(L1,LR)
W_LDERIV- frho*ELDER

```

```

#=====#
#  Cálculo do vetor fz=L*D-1/2*z                                     #
#=====#
DELMC_ DELM%*%cf
z_DELM%*%cf +rp
dez_invr*d*z
LDEZ_ rbind(0,dez)
DDEZ_LDEZ[1:(n),1:1]
ddez_as.matrix(DDEZ) *frho
li1_as.matrix(dez[1,])
lres_as.matrix((1/sqrt(1-rho^2))*dez[2:n,1:1])
ldez_rbind(li1,lres)
fz_ldez-ddez

#=====#
#  Regressao de Minimos Quadrados Ordinários                         #
#=====#
zfit_lsfit(W,fz,intercept=F)
# Novos valores da média e coeficientes#
ncf_as.vector(zfit$coef)
rp_(y-exp(X%*%ncf))
m_as.vector(exp(X%*%ncf))
rp_as.matrix(rp)

#=====#
#  Verificacao da Convergencia                                       #
#=====#
if(all(abs((cf-ncf)/ncf) <=1e-4) & ni>1 | ni >= iterac)
  break
cf_ncf
ni_ni+1
print(cf)
print(ni)
}

if (ni>=iterac)
  warning(" O LIMITE DE ITERACOES FOI ATINGIDO")
if (ni<iterac)
{
  warning("Convergiu")
#=====#
# Atualizacao dos valores de m, W, rho para o cálculo do valor final dos coeficientes#
#=====#

```

```

m_as.vector(exp(X%%ncf))
rp_as.vector(y-m)

#=====#
#  Cálculo de sigma2                                     #
#=====#

rp2m_rp*rp-m
m2_m*m1
srd_t(um)%%(rp2m)
sm_t(um)%%m2
sigm2_srd/sm
print(sigm2)
if (sigm2 < 0 ) {
    rp2_rp*rp
    sigm2_(t(um)%%rp2)/(n-1)
    print(sigm2)
}

#=====#
#  Cálculo da matriz D = mu+sigma2*mu2                   #
#=====#

d_m+sigm2*m2
di_d^(-1)
invrd_sqrt(di)
DELM_m*X
DERIV_as.matrix(invrd* DELM)

#=====#
#  Cálculo de rho                                         #
#=====#

mat_cbind(rp,m)
lmat_rbind(0,mat)
elmat_lmat[1:n,1:2]
lrp_as.vector(elmat[,1])
lmu_as.vector(elmat[,2])
srpl_t(um)%%( rp*lrp )
sml_t(um)%%(m*lmu)
print(srpl)
print(sml)
rho_srpl/(sml * sigm2)
print(rho)
if (rho>1 ) {
    warning("RHO maior que um")
    rp2_sum((invrd*rp)*(invrd*rp))
}

```



```

        srpl_sum((invrd*rp)*(invrd*lrp))
        rho_srpl/rp2
        print( "rho alternativo = ")
        print(rho)

        if (rho > 1) {
            warning("RHO maior que um")
            break
        }
    }

    rho_prod(eigen(rho)$values)

#####
# Cálculo da matriz  $W=L*D-1/2*(delm/delbeta)$  #
#####
    LDER_rbind(0, DERIV)
    ELDER_LDER[1:n,1:p]
    frho_rho/(sqrt(1-rho^2))
    L1_DERIV[1,]
    LR_(1/sqrt(1-rho^2))*DERIV[2:n,1:p]
    LDERIV_rbind(L1,LR)
    W_LDERIV- frho*ELDER

#####
# Calculo da Matriz de Covariancia dos coeficientes #
#####
# Calculo da Matriz Io #
    Io_t(W)%*%W
# Calculo da Matriz I1 #

# Calculo da Matriz  $L*D-1/2*rp$  #
    DRP_as.matrix(invrd*rp)
    LDRP_rbind(0,DRP)
    DDRP_LDRP[1:n,1:1]
    ddrp_as.matrix(DDRP)*frho
    li_as.matrix(DRP[1,])
    lres_as.matrix((1/sqrt(1-rho^2))*DRP[2:n,1:1])
    lidrp_rbind(li, lres)
    LDS_lidrp-ddrp

    I1_(t(W)%*%LDS)%*%t(LDS)%*%W)

CVBETA_(solve(Io)%*%I1%*%solve(Io))

```

```

# Extracao dos desvios de beta usando I1#
dcoef.beta_sqrt(as.vector(diag(CVBETA)))

CVBETA_inv(Io)

# Extracao dos desvios de beta usando V=solve(Io)#
dcoef1.beta_sqrt(as.vector(diag(CVBETA)))

print(cbind(coef(inic),ncf,dcoef.beta,dcoef1.beta))

}

```

2. Programa MZeger.txt - Este programa executado com dados anuais, constrói a matriz

de correlações de R_e e calcula a sua inversa.

```
#=====#
#  Inicialização- Regressão de Poisson                                     #
#=====#
# Módulo 1_ Entrada de dados e Inicializações do Processo Recursivo #
# Entrada de Dados #
attach(obic3)
# Valores iniciais da média e coeficientes obtidos através da regressão Poisson#
inic_glm(TOT65~DIASF+DIASQ+DIASS+DIASU+
OUT+INV+PRI+cos(4*pi*IND/365)+ +sin(4*pi*IND/365)+IND + SO2,poisson,obic3)
# armazenamento do resíduo (y-m), média e coeficientes#
#rp- resíduo#
# m- média#
# cf- coeficientes#
rp_as.vector(residuals(inic,"response"))
m_as.vector(fitted.values(inic))
cf_as.vector(coefficients(inic))
print(cf)

# Construção da Matriz de Especificação do modelo - Dimensão n x p #
CCLI_cbind(1, obic3$DIASF, obic3$DIASQ, obic3$DIASS, obic3$DIASU, obic3$OUT,
obic3$INV,obic3$PRI, cos(4*pi*obic3$IND/365), sin(4*pi*obic3$IND/365),obic3$IND)
X_as.matrix(cbind(CCLI, obic3$SO2))

# Construção do Vetor de Respostas #
y_as.vector(obic3$TOT65)
# Módulo 2 -Processo recursivo#
# Controle do processo recursivo #
# Número máximo de iterações #
iterac_30

#tamanho da série#
n_365

#ordem do vetor de coeficientes #
p_ (8+4)

#controle das iterações#
ni_1
```

```

#Módulo 2 - processo iterativo#
rhoun_0 # indica se em algum momento |RHO|>1#
repeat
{

#=====#
#  Cálculo de sigma2                                     #
#=====#
  rp2:n_rp*rp-m
  m2_m*m
  srd_t(um)%*%(rp2m)
  sm_t(um)%*%m2
  sigm2_srd/sm
  print(sigm2)
  if (sigm2 < 0 ) {
    rp2_rp*rp/m
    sigm2_(t(um)%*%rp2)/(n-1)
    print("sigma2 negativo=")
    print(sigm2)
  }

  sigm2_prod(eigen(sigm2)$values)

#=====#
#  Cálculo de rho                                     #
#=====#

  if (rhoun==0)
  {
    mat_cbind(rp,m)
    lmat_rbind(0,mat)
    elmat_lmat[1:n,1:2]
    lrp_as.vector(elmat[,1])
    lmu_as.vector(elmat[,2])
    srpl_t(um)%*%( rp*lrp )
    sml_t(um)%*%(m*lmu)
    rho_srpl/(sml * sigm2)
    print(rho)
  }

  if (rho>1 | rhoun==1 )

```

```

{
  rhoun_1
  warning("RHO maior que um")
  rp2_as.rts(invrd*rp)
  fit_ar(rp2,aic=T,order=1)
  rho_fit$ar
  print( "rho alternativo = ")
  print(rho)
}

```

```
rho_prod(eigen(rho)$values)
```

```

#=====#
#  Matriz de Correlacao de Trabalho  #
#=====#

L_matrix(numeric(n^2),n)
for (i in 1:(n-1))
  for (j in (i+1):n)
    L[i,j]_ abs(rho)^(j-i)
  L_L+t(L)+diag(n)

#=====#
#  Cálculo da matriz V = mu+sigma2*mu2  #
#=====#

M_diag(m)
V_ M+sigm2*(M%*%L%*%M)
VV_inv(V)

```

```

#=====#
#  Cálculo do vetor Z=AXBETA+ (Y-mu)      #
#=====#
MX_M%%X
S_matrix(rp,nrow=n)
z_MX%%cf+S

#=====#
#  Regressao de Minimos Quadrados Ordinários      #
#=====#

XwX_t(MX)%%VV%%MX  # (AX)'V-1(AX)#
XwY_t(MX)%%VV%%z   #(AX)'V-1Z#
ncf_solve(XwX)%%XwY
print("ncf")
print(ncf)
m_as.vector(exp(X%%ncf))
rp_as.vector((y-m))

#=====#
#  Verificacao da Convergencia      #
#=====#

if (all(abs((cf-ncf)/cf)<=1e-4)& ni>1 | ni >= iterac)
  break
cf_ncf
ni_ni+1

}

if (ni>=iterac)
  warning(" O LIMITE DE ITERACOES FOI ATINGIDO")
if (ni<iterac)
{
  warning("Convergiu")
  print("Numero de Iteracoes Ocorridas ")
  print(ni)
}

```

```

#=====#
# Atualizacao dos valores de m, W, rho para o cálculo do valor final dos coeficientes#
#=====#
m_as.vector(exp(X%*%ncf) )
rp_as.vector(y-m)

#=====#
#  Cálculo de sigma                                     #
#=====#
rp2m_rp*rp-m
m2_m*m
srd_sum(rp2m)
sm_sum(m2)
sigm2_srd/sm

print("sigma2")
print(sigm2)
if (sigm2 < 0 ) {
    rp2_rp*rp/m
    sigm2_(sum(rp2))/(n-1)
    print("sigma2 negativo=")
    print(sigm2)
}

sigm2_prod(eigen(sigm2)$values)

#=====#
#  Cálculo de rho                                     #
#=====#
if (rhoun==0)
{
    mat_cbind(rp,m)
    lmat_rbind(0,mat)
    elmat_lmat[1:n,1:2]
    lrp_as.vector(elmat[,1])
    lmu_as.vector(elmat[,2])
    srpl_t(um)%*%( rp*lrp )
    sml_t(um)%*%(m*lmu)
    rho_srpl/(sml * sigm2)
    print(rho)
}

```

```

if (rho>1 | rhoum==1 )
{
  rhoum_1
  warning("RHO maior que um")
  rp2_as.rts(invrd*rp)
  fit_ar(rp2,aic=T,order=1)
  rho_fit$ar
  print( "rho alternativo = ")
  print(rho)
}

rho_prod(eigen(rho)$values)
rhoun_0

#=====#
#  Matriz de Correlacao de Trabalho                                #
#=====#

L_matrix(numeric(n^2),n)
for (i in 1: (n-1))
  for (j in (i+1):n)
    L[i,j]_ abs(rho)^(j-i)
  L_L+t(L)+diag(n)

#=====#
#  Cálculo da matriz V = mu+sigma2*mu2                                #
#=====#
M_diag(m)
V_ M+sigm2*(M%*%L%*%M)
VV_inv(V)

#=====#
#  Cálculo da matriz Io                                              #
#=====#

Io_t(MX)%*%VV%*%MX  # (AX)'V-1(AX)#
print(Io)

#=====#
#  Calculo das Matriz de Variancia e Covariancia de Beta            #
#=====#

```



```

# Calculo da Matriz de BETA #

CVBETA_solve(Io)

# Calculo da Correlação de Y no lag 1 #
rp_as.rts(invrd*rp)
fit_ar(rp,aic=T,order=1)
yrho_fit$ar
print("Correlacao de Y no Lag1")
print(yrho)

# Calculo da Matriz de COV(beta)
dcoef.beta_sqrt(diag(CVBETA))

# Extracao dos desvios de beta#
print(cbind(coef(inic),ncf,dcoef.beta))

}

```

3.Programa I1Zeger.txt - Este programa utiliza o filtro linear. A matriz **I 1** é calculada utilizando-se a matriz **V** de (3.7). No final do programa, calcula a estatística do teste de Box-Pierce.

```
# MODELO COM SO2#
#=====#
#  Inicialização- Regressão de Poisson #
#=====#
# Módulo 1_ Entrada de dados e Inicializações do Processo Recursivo #
# Entrada de Dados #
attach(obic3)
# Valores iniciais da média e coeficientes obtidos através da regressão Poisson#
inic1_glm(TOT65~DIASF+DIASQ+DIASS+DIASU+OUT+INV+PRI+
SO2+cos(4*pi*IND/365)+sin(4*pi*IND/365)+IND,poisson,obic3)
# armazenamento do resíduo (y-m), média e coeficientes#
#rp- resíduo#
# m- média#
# cf- coeficientes#
rp_as.vector(residuals(inic1,"response"))
m_as.vector(fitted.values(inic1))
cf_as.vector(coefficients(inic1))

# Construção da Matriz de Especificação do modelo - Dimensão n x p #

CCLI_cbind(1, obic3$DIASF, obic3$DIASQ, obic3$DIASS, obic3$DIASU, obic3$OUT,
obic3$INV,obic3$PRI)
X_as.matrix(cbind(CCLI, obic3$SO2,cos(4*pi*obic3$IND/365),sin(4*pi*obic3$IND/365),
obic3$IND))

# Construção do Vetor de Respostas #
y_as.vector(obic3$TOT65)

# Módulo 2 -Processo recursivo#
# Controle do processo recursivo #
# Número máximo de iterações #
iterac_30

#tamanho da série#
n_365

#ordem do vetor de coeficientes #
```

p_ (8+4)

#controle das iterações#

ni_1

repeat

{

#=====

Cálculo de sigma2

#=====

rp2m_rp*rp-m

m2_m*m

srd_t(um)%*%(rp2m)

sm_t(um)%*%(m2

sigm2_srd/sm

print(sigm2)

if (sigm2 < 0) {

rp2_rp*rp

sigm2_(t(um)%*%(rp2)/(n-1)

print("sigma2 negativo=")

print(sigm2)

}

#=====

Cálculo da matriz D = mu+sigma2*mu2

#=====

print("Valor de sigma2 corrente")

print(sigm2)

d_m+sigm2*m2

di_d^(-1)

invrd_sqrt(di)

DELM_m*X

DERIV_as.matrix(invrd* DELM)

#=====

Cálculo de rho

#=====

mat_cbind(rp,m)

lmat_rbind(0,mat)

elmat_lmat[1:n,1:2]

lrp_as.vector(elmat[,1])

lmu_as.vector(elmat[,2])

srpl_t(um)%*%(rp*lrp)

sml_t(um)%*%(m*lmu)

```

rho_srpl/(sml * sigm2)
print(rho)
if (rho>1 | rhoum==1 )
{
    rhoum_1
    warning("RHO maior que um")
    rp2_as.rts(invrd*rp)
    fit_ar(rp2,aic=T,order=1)
    rho_fit$ar
    print( "rho alternativo = ")
    print(rho)
}

```

```

rho_prod(eigen(rho)$values)

```

```

#=====#
#  Cálculo da matriz  $W=L*D-1/2*(delm/delbeta)$   #
#=====#

```

```

LDER_rbind(0, DERIV)
ELDER_LDER[1:n,1:p]
frho_rho/(sqrt(1-rho^2))
L1_DERIV[1,]
LR_(1/sqrt(1-rho^2))*DERIV[2:n,1:p]
LDERIV_rbind(L1,LR)
W_LDERIV- frho*ELDER

```

```

#=====#
#  Cálculo do vetor  $fz=L*D-1/2*z$   #
#=====#

```

```

DELMC_DELM%*%cf
z_DELM%*%cf +rp
dez_invrd*z
LDEZ_rbind(0,dez)
DDEZ_LDEZ[1:(n),1:1]
ddez_as.matrix(DDEZ) *frho
li1_as.matrix(dez[1,])
lres_as.matrix((1/sqrt(1-rho^2))*dez[2:n,1:1])
ldez_rbind(li1,lres)
fz_ldez-ddez

```

```

#=====#
# Regressao de Minimos Quadrados Ordinários #
#=====#

zfit_lsfit(W,fz,intercept=F)
ncf_as.vector(zfit$coef)
rp_(y-exp(X%*%ncf))
m_as.vector(exp(X%*%ncf ))
rp_as.matrix(rp)

#=====#
# Verificacao da Convergencia #
#=====#

if(all(abs((cf-ncf)/ncf) <=1e-4 & ni>1 | ni >= iterac)
    break
cf_ncf
ni_ni+1
print(ni)
}

if (ni>=iterac)
    warning(" O LIMITE DE ITERACAO FOI ATINGIDO")
if (ni<iterac)
{
    warning("Convergiu")

#=====#
# Atualizacao dos valores de m, W, rho para o cálculo do valor final dos coeficientes#
#=====#
    m_as.vector(exp(X%*%ncf ))
    rp_as.vector(y-m)
    rl_as.rts(rp)

#=====#
# Cálculo de sigma2 #
#=====#
    rp2m_rp*rp-m
    m2_m*m
    srd_t(um)%*%(rp2m)
    sm_t(um)%*%m2
    sigm2_srd/sm
    print(sigm2)

```

```

if (sigm2 < 0 ) {
    rp2_rp*rp
    sigm2_(t(um)%*%rp2)/(n-1)
    print(sigm2)
}

sigm2_prod(eigen(sigm2)$values)

#=====#
#  Cálculo de rho                                     #
#=====#
    mat_cbind(rp,m)
    lmat_rbind(0,mat)
    elmat_ lmat[1:n,1:2]
    lrp_as.vector(elmat[,1])
    lmu_as.vector(elmat[,2])
    srpl_t(um)%*%( rp*lrp )
    sml_t(um)%*%(m*lmu)
    rho_srpl/(sml * sigm2)
    print(rho)
    if (rho>1 | rhoum==1 )
        {
            rhoum_1
            warning("RHO maior que um")
            rp2_as.rts(invrd*rp)
            fit_ar(rp2,aic=T,order=1)
            rho_fit$ar
            print( "rho alternativo = ")
            print(rho)
        }

rho_prod(eigen(rho)$values)

#=====#
#  Matriz de Correlacao de Trabalho                                     #
#=====#

    L_matrix(numeric(n^2),n)
    for (i in 1: (n-1))
        for (j in (i+1):n)
            L[i,j]_ abs(rho)^(j-i)

```

```

L_L+L+diag(n)

# Inversa da Matriz de Correlacao de Trabalho#
w_solve(L)

#=====#
# Cálculo da matriz D = mu+sigma2*mu2 #
#=====#

d_m+sigm2*m2
di_d^(-1)
invrd_sqrt(di)
D_diag(invrd)
M_diag(m)
MX_M%*%X

# Calculo de V-1 #
v_D%*%w%*%D

#=====#
# Cálculo da matriz Io #
#=====#

Io_t(MX)%*%v%*%MX # (AX)'V-1(AX)#
print(Io)
pre_solve(Io) # Inversa da Matriz de Informacao de Fisher #

#=====#
# Calculo das Matriz deCovariancia dos coeficientes #
#=====#

V_M+sigm2*(M%*%L%*%M) # V estimado por M+sigm2*MLM #
VV_inv(V)

# Calculo da Matriz I1 #
I1_t(MX)%*%v%*%VV%*%v%*%MX

CVBETA_pre%*%I1%*%pre

# Calculo da Correlação de Y no lag 1 #
rp_as.rts(invrd*rp)
fit_ar(rp,aic=T,order=1)

```

```

yrho_fit$ar
print("Correlacao de Y no Lag1")
print(yrho)

# Calculo da Matriz de COV(beta)
dcoef.beta_sqrt(diag(CVBETA))

CVBETA_pre
dcoefl.beta_sqrt(diag(CVBETA))

# Extracao dos desvios de beta#

print(cbind(coef(inic1),ncf,dcoef.beta,dcoefl.beta))

}

# Calculo da estatística Q do teste de Box-Pierce #

r1.acr_acf(r1,30,"correlation")
r1_as.matrix(r1.acr$acf,31)
r1[1,1]_0
denl_matrix(numeric(31),31)
for (i in 1: 31) denl[i,1]_1/(1095-i)
denl_as.vector(denl)
r1_as.vector(r1)
Q1_(n)*(n+2)*sum(r1*r1*denl)
print(Q1)

```


Referências Bibliográficas

1. Campbell, M. J. (1994). Time series regression for counts: an investigation into the relationship between sudden infant death syndrome and environmental temperature. *Journal of the Royal Statistical Society* **157**, 191-208.
2. Cox, D. R. (1983). Some remarks on over-dispersion. *Biometrika* **70**, 267-74.
3. Dodson, A. J. (1990). *An Introduction to Generalized Linear Models*. London: Chapman and Hall.
4. Dunlop, D. D. (1994). Regression for longitudinal data: bridge from least squares regression. *The American Statistician* **48**, 299-303.
5. Green, P. (1984). Iteratively reweighted least squares for maximum likelihood estimation and some robust and resistant alternatives, (with discussion). *Journal of the Statistical Society, Ser.B*, **46**, 149-162.
6. Hillis, S. L. and Davis, C. S. (1994). A simple justification of the iterative fitting procedure for generalized linear models. *Journal of the American Statistical Association* **48**, 288-289.

7. Liang, K. Y. and Zeger, L.S. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13-22.
8. Lindsey, J. K. (1993). *Models for Repeated Measurements*. New York: Oxford University Press.
9. McCullagh, P. (1983). Quasi-Likelihood functions. *The Annals of Statistics* **11**, 59-67.
10. McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd ed. London: Chapman and Hall.
11. Morettin, P. A. e Toloi, C. M. C. (1987.). *Previsão de Séries Temporais*, 2a ed.. São Paulo: Atual.
12. Murphy, M. F.G. and Campbell, M. J. (1987). Sudden infant death syndrome and environmental temperature: an analysis using vital statistics. *Journal of Epidemiology and Community Health*, **41**, 63-71.
13. Priestley, M. B. (1981). *Spectral Analysis and Time Series*. San Diego: Academic Press.
14. S-PLUS *Guide to Statistical and Mathematical Analysis* Version 3.3 (1995). Seattle: StatSci, a division of MathSoft, Inc.

15. Saldiva, P. H. N., Lichtenfels, A. J. F. C., Paiva, P. S. O. , Barone, I. A., Martins, M. A., Massad, E., Pereira, J. C. R., Xavier, V. P., Singer, J. M. and Böhm, G. M. (1994). Association between air pollution and mortality due to respiratory diseases in children in São Paulo: a preliminary report. *Environmental Research* **65**, 218-225.
16. Saldiva, P. H. N., Pope III, C. A., Schwartz, J., Dockery, D. W., Salge, J., Barone, I., Böhm, G. (1994). Air pollution and mortality in elderly people: a time series study in São Paulo, Brasil. *Archives Environmental Health* **50**, 159-163.
17. Singer, J. M. e Conceição, G. M. S. (1994). *Estudo da Associação entre Mortalidade e Poluição Atmosférica na cidade de São Paulo*. São Paulo: IME-USP. 35p. (RAE-CEA-9407).
18. Sen, P. K. and Singer, J. M. (1993). *Large Sample Methods in Statistics - An Introduction with Applications*. London: Chapman and Hall.
19. Zeger, S. L. (1988). A regression model for time series counts. *Biometrika* **75**, 621-629.
20. Zeger, S. L. and Liang, K. Y. (1986). Longitudinal analysis for discrete and continuous outcomes. *Biometrics* **42**, 121-130.

21. Zeger, S. L. , Liang, K. Y., and Albert, P. S. (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics* **44**, 1049-1060.
22. Wedderburn, R. W. M. (1974). Quasi-Likelihood functions, generalized linear models, and Gauss-Newton method. *Biometrika* **61**, 439-447.
23. Wei, W. W.S. (1990). *Time Series Analysis - Univariate and Multivariate Methods*. Addison-Wesley Publishing Company, Inc.