

Carlos Maurício Cardeal Mendes

**MODELAGEM DE SÉRIES TEMPORAIS EM EPIDEMIOLOGIA:
APLICAÇÃO DOS MODELOS AUTORREGRESSIVOS INTEGRADOS COM
MÉDIA MÓVEL EM CONTEXTOS EPIDEMIOLÓGICOS.**

Salvador - Bahia
Novembro de 2002



**Universidade Federal da Bahia
Instituto de Saúde Coletiva
Programa de Pós-Graduação em Saúde Coletiva**

**MODELAGEM DE SÉRIES TEMPORAIS EM EPIDEMIOLOGIA:
APLICAÇÃO DOS MODELOS AUTORREGRESSIVOS INTEGRADOS COM
MÉDIA MÓVEL EM CONTEXTOS EPIDEMIOLÓGICOS.**

AUTOR: Carlos Maurício Cardeal Mendes

ORIENTADOR: Eduardo Luiz Andrade Mota

Tese apresentada ao Curso de Doutorado em Saúde Coletiva com área de concentração em Epidemiologia do Instituto de Saúde Coletiva da Universidade Federal da Bahia como parte dos requisitos para obtenção do grau de doutor em Saúde Coletiva.

Salvador - Bahia
Novembro de 2002

Ficha Catalográfica
Preparada por: Maria Creuza F. Silva CRB/5-996

M538m Mendes, Carlos Maurício Cardeal.

Modelagem de séries temporais em Epidemiologia: aplicação dos modelos autorregressivos integrados com média móvel em contextos epidemiológicos / Carlos Maurício Cardeal Mendes. --

Salvador, 2002.

72 p.

Tese (Doutorado) - Instituto de Saúde Coletiva. Universidade Federal da Bahia

Orientador: Prof. Eduardo Luiz Andrade Mota

1. Análise Estatística de Fenômenos de Movimento Complexo 2. Séries de Tempo –
ARIMA. 3. Epidemiologia I. Título

CDU 311.17

Carlos Maurício Cardeal Mendes

Modelagem de séries temporais em Epidemiologia

Aplicação dos modelos autorregressivos integrados com média móvel em contextos epidemiológicos.

Banca Examinadora

Prof. Eduardo Luiz Andrade Mota (Orientador) – Instituto de Saúde Coletiva - UFBA

Prof. Nelson Fernandes Oliveira – Instituto de Matemática - UFBA

Prof. Nelson da Cruz Gouveia – Departamento de Medicina Preventiva - USP

Prof^a. Maria da Conceição Nascimento Costa – Instituto de Saúde Coletiva - UFBA

Prof^a. Maria Inês Costa Dourado - Instituto de Saúde Coletiva - UFBA

Salvador
2002

Sumário

Apresentação e justificativa.....	7
Objetivos.....	11
Artigo 1 - Operacionalização da categoria tempo, no desenvolvimento e na aplicação do método epidemiológico em séries temporais: uma breve revisão.....	13
Introdução.....	15
Um método que pouco mudou.....	16
Conclusões.....	25
Referências bibliográficas.....	27
Artigo 2 - Aplicação do método de modelagem de Box e Jenkins, em séries temporais epidemiológicas distintas. Parte I - Internações e regulação da oferta e do acesso: uma avaliação estatística de impacto e mortalidade por diarreia: a descrição dos componentes de uma série temporal.....	30
Introdução.....	34
Metodologia.....	35
As séries.....	35
Fonte de dados.....	36
Fonte de dados – Vitória da Conquista, Bahia.....	36
Fonte de dados de mortalidade.....	38
Resultados.....	39
Internações e regulação.....	39
Descrição da série.....	40
Modelo tentativo proposto.....	43
Mortalidade de menores de cinco anos devida à diarreia.....	44
Descrição da série.....	44
Modelo tentativo proposto.....	48
Discussão e conclusões.....	53
Referências bibliográficas.....	57
Artigo 3 - Aplicação do método de modelagem de Box e Jenkins, em séries temporais epidemiológicas distintas. Parte II - Internações por diarreia e valores pagos: uma análise bivariada de séries temporais.....	60
Introdução.....	62
Metodologia.....	63
Resultados.....	64
A série dependente (saída): valores pagos pelas internações.....	67
A série de entrada: taxa de internações.....	69
A correlação cruzada das séries.....	71
Discussão e conclusões.....	73
Referências bibliográficas.....	76
Considerações sobre limitações do estudo e conclusão (tese).....	77
Apêndice.....	80
Anexos.....	98

謝謝 Xie Xie

你 Ni

學習 Xue Shi

做 Zuo

Apresentação e justificativa

A Epidemiologia na sua constituição enquanto ramo do conhecimento, com objeto e interesses investigativos próprios no campo da saúde-doença em populações humanas, dispõe de um arsenal metodológico que progressivamente vem incorporando novas técnicas e estratégias analíticas à medida que o conhecimento se amplia e se aprofunda. Nesse processo de consolidação do conhecimento científico, tais aportes teóricos e metodológicos, têm contribuído para ampliar a sua base operacional, instrumental e conceitual.

Desde os primeiros trabalhos publicados, considerados como parte da história da Epidemiologia, a dimensão temporal tem sido considerada central ao raciocínio epidemiológico, tanto na perspectiva teórica constitutiva do próprio campo, como pragmaticamente na avaliação das repercussões do que possa ocorrer em populações. Apesar disso, ao longo do tempo, a incorporação de tais aportes metodológicos não se efetivou na abordagem analítica das séries históricas epidemiológicas.

Ao se verificar a produção Epidemiológica através de pesquisa no MEDLINE, constata-se que existem 2.455 artigos publicados entre 1965 e 2000 relacionados a séries temporais. Destes, cinquenta e dois artigos utilizaram a metodologia ARIMA a partir de 1986, trezentos e noventa e um utilizaram técnicas de regressão desde 1966 e vinte e oito aplicaram análise de variância (ANOVA) desde 1983. Dos 2.455 artigos, apenas oito deles são publicações brasileiras, todas a partir de 1994, das quais quatro aplicando técnicas de regressão e apenas uma (DE GOUVEIA et al., 1997) utilizando técnicas de modelagem em séries temporais. Quanto ao uso da metodologia ARIMA em teses ou dissertações no país, no campo da Epidemiologia, apenas duas monografias a utilizaram de maneira explícita, a primeira versa sobre os efeitos da poluição atmosférica sobre a saúde (GOUVEIA, 1997) e a segunda, refere-se à um estudo sobre mortalidade e desnutrição em idosos (OTERO, 2001).

Uma consulta ao Banco de Teses da CAPES/MEC catalogadas até o ano de 2001, através da palavra chave ARIMA, no campo de busca denominado “assunto”, foram selecionadas vinte e duas monografias de pós-graduação, sendo que apenas uma dissertação era na área de Saúde Pública. Dezoito destas eram dissertações de mestrado e quatro teses de doutorado distribuídas nas seguintes áreas: Engenharia, (9); Economia, (8); Física, (1); Estudos

populacionais, (1); Ciências da computação, (1); Saúde Pública, (1); Biologia funcional e molecular, (1).

Pragmática e operacionalmente, uma série temporal corresponde a uma sucessão de mensurações de pelo menos um determinado objeto de investigação no tempo. No âmbito da Epidemiologia, as mensurações são obtidas discretamente e agregadamente no tempo (*clustering in time*) (FORATTINI, 1980). Ou seja, a sua unidade de informação e consequente unidade de análise é o resultado de uma medida agregada no tempo, espaço ou tempo-espaço.

Ainda que o conceito de espaço se relacione explicitamente e externamente à população como uma região, por exemplo, ou implicitamente como sendo o espaço dos próprios sujeitos componentes dessa população, nos quais se manifestam sinais, sintomas e síndromes, ou mesmo como um espaço social e cultural, este é mais evidente e melhor delineado do que o conceito de tempo adotado pela Epidemiologia, que surge como processo ou é entendido como espaço - um tempo fragmentário ou uma temporalidade “espacializada” (ALMEIDA-FILHO, 1989).

A abordagem da metodologia epidemiológica, nesse contexto, mensura e analisa preferencialmente dicotomias (saúde-doença, vida-morte, exposição-não-exposição) cristalizadas no tempo-espaço, mas com interesses inferenciais (mesmo que efetuando pseudopredições) mas, cujo procedimento metodológico pouco contempla a dinamicidade, a possibilidade da emergência radical (ALMEIDA-FILHO, 2000) e a consequente diversidade dos fenômenos fundamentais para o entendimento do processo saúde-doença.

Sabe-se que as técnicas de modelagem estatística em geral requerem independência dos erros como pressuposto fundamental, entretanto esse requisito não é observado amiúde com dados empíricos. Quando são analisados dados em sequência temporal, essa autocorrelação ou dependência temporal, por vezes, encontra-se presente. As técnicas mais difundidas, tais como, a análise de regressão e a análise de variância não são adequadas, e podem produzir resultados completamente equivocados (PANKRATZ, 1983).

Entretanto, na década de 70, foi apresentada uma ferramenta de análise de séries temporais, denominada de modelos autorregressivos integrados com média móvel (ARIMA) (BOX, GEORGE E. P. et al., 1976), que significou um avanço nesse campo - e que foi rapidamente

adotada no âmbito da Economia - de modo a incorporar o aspecto de dinamicidade, com consequente impacto na predição de valores futuros, tornando possível uma melhor especificação dos componentes estruturais de uma série temporal, uma estimativa mais precisa dos parâmetros dos modelos preditivos e de intervenção e a estimativa da autocorrelação como componente histórico dessa série, o que, por seu turno, possibilitou um avanço no conhecimento em relação aos métodos de análise concorrentes.

De acordo com o exposto, apesar da técnica proposta por Box e Jenkins estar sendo empregada com êxito, desde a sua proposição, predominantemente em outros campos do saber, ela é aplicada com pouca frequência em estudos epidemiológicos, sobretudo no Brasil. A utilização de técnicas de predição, através de séries históricas é mais comum nas ciências sociais (ROTHMAN et al., 1998).

Assumindo um dado tempo conceitual, imerso no paradigma vigente, seja aquele real, lógico ou simbólico, conforme (ALMEIDA-FILHO, 1997), esta tese de doutorado apresenta três artigos que se relacionam através da categoria tempo – com enfoque na metodologia de séries históricas no âmbito da Epidemiologia – dada esta lacuna e na perspectiva de contribuir para o suporte ferramental no manejo de séries temporais.

O primeiro artigo, intitulado “*Operacionalização da categoria tempo, no desenvolvimento e na aplicação do método epidemiológico em séries temporais: uma breve revisão*” foi escrito adotando-se um estilo que “conta a história” como um relato sucessivo e sucinto de fatos, encadeados e sumarizados agregadamente em décadas, assemelhando-se à uma série temporal, tal como é conhecida na prática epidemiológica, a partir, principalmente da década de trinta, uma vez que, aquele foi um período no qual as condições econômicas e sociais aliadas à uma ciência em crescente produção de conhecimento, influenciaram na decisão político-institucional e no direcionamento do saber científico como um todo em busca de determinados objetivos históricos, assim como na produção cumulativa dos elementos que culminaram no desenvolvimento atual da modelagem de séries temporais.

O segundo artigo, intitulado “*Aplicação do método de modelagem de Box e Jenkins em séries temporais epidemiológicas distintas*” reúne quatro séries típicas do contexto epidemiológico, visando demonstrar pragmaticamente a aplicação do método de Box & Jenkins no campo da Epidemiologia Geral.

As séries foram construídas a partir de dados de internamentos hospitalares, visando demonstrar o potencial analítico do método ARIMA na avaliação de impacto de uma intervenção, assim como na análise multivariada de series temporais e a partir de dados de mortalidade, no intuito de demonstrar os seus recursos exclusivos de estimativa de parâmetros estruturais de séries históricas e sua capacidade de predição.

Objetivos

- 1) Apresentar uma revisão histórica sobre a operacionalização da categoria tempo no desenvolvimento e aplicação do método epidemiológico em séries temporais (artigo número um).
- 2) Demonstrar a aplicação do método de modelagem de Box e Jenkins, em quatro séries temporais epidemiológicas distintas, salientando o seu potencial, vantagens e limitações, através dos parâmetros de estrutura, de intervenção e correlação multivariada de séries históricas (artigos dois e três), a saber:
 - a) Número de internações no município de Vitória da Conquista de janeiro de 1995 a fevereiro de 2002, mensalmente, incluindo o componente de regulação da oferta e do acesso ocorrido em outubro de 1995 pela Prefeitura do Município (uma série, artigo 2);
 - b) Mortalidade por diarreia, em menores de cinco anos no Estado da Bahia, entre 1979 e 1998, mensalmente (uma série, artigo 2);
 - c) Taxa de internações por diarreia em menores de cinco anos na Bahia e os valores pagos entre julho de 1994 e dezembro de 1998, mensalmente (duas séries, artigo 3).

ARTIGO 1

Operacionalização da categoria tempo, no desenvolvimento e na aplicação do método epidemiológico em séries temporais: uma breve revisão.

Artigo 1 - Operacionalização da categoria tempo, no desenvolvimento e na aplicação do método epidemiológico em séries temporais: uma breve revisão.

Resumo

A Epidemiologia, enquanto campo constituído do saber, tem incorporado historicamente aportes teóricos e metodológicos proveniente de outras ciências, cujas contribuições têm variado conforme a dinâmica das concepções científicas ou políticas ao longo do tempo. Em sua formulação didática clássica tem sido apresentada como uma ciência que estuda os processos de saúde/doença e agravos à saúde, relativos à pessoa, ao lugar e ao tempo. A partir da década de 60 ocorreu uma sistematização dos seus métodos de investigação através de uma tipologia de desenhos de estudo, que essencialmente procuram captar, na sua vertente mais quantitativa, medidas que descrevam a velocidade com que pessoas são acometidas por agravos relativos à saúde, a prevalência desses agravos, possíveis relações associativas ou de causalidade entre grupos considerados expostos e não expostos e a identificação de prováveis fatores de risco.

Com o advento da computação eletrônica, o investimento e desenvolvimento metodológicos aumentaram substancialmente. Entretanto, nesse ímpeto de crescimento metodológico, os estudos epidemiológicos efetuados sobre eventos agregados ao longo do tempo, não acompanharam a evolução e a disponibilidade de técnicas empregadas com sucesso em outras áreas do conhecimento científico desde a década de 70, como por exemplo, na Econometria, deixando com isso de oferecer aos usuários do conhecimento epidemiológico, medidas menos enviesadas e estimativas mais confiáveis dos eventos organizados e analisados como séries históricas dinâmicas.

Descritores: epidemiologia; metodologia; história; séries temporais

Operationalization of the time factor in the development and application of epidemiological methods in time series: a brief review.

Abstract

Epidemiology, while remaining a field constituted from knowledge, has historically incorporated theories and methodologies deriving from other sciences whose contributions have varied according to the dynamics of the scientific or political conceptions throughout time. In its classical didactic formulation, it has been presented as a science that studies the health/sickness processes and health aggravations in relation to the person, the place and the time. From the sixties on, a systematization of its investigative methods occurred by way of a typology of study designs that essentially sought to collect in its more quantitative approach measures that describe the speed at which people are attacked by aggravations to their health, the prevalence of these aggravations, possible associative links or causalities between groups considered to have been exposed and not exposed and the identification of probable risk factors.

With the advent of electronic computing, the investment and methodological development increased substantially. However, in this impetus in methodological growth, the epidemiological studies carried out in relation to events aggregated over time have not accompanied evolution and the availability of techniques successfully used in other areas of scientific knowledge since the seventies, such as Econometry for example, thus denying users of Epidemiological knowledge less oblique measures and more reliable estimates of events organized and analyzed as dynamic historical series.

Key words: epidemiology, methodology, history, temporal series.

Introdução

A medida do tempo e os conceitos de tempo têm preenchido os pensamentos de filósofos e cientistas ocidentais desde os primórdios da filosofia conhecida. Entre esses, há uma referência a um tempo abstrato (o tempo cronológico) e um tempo concreto (a duração, duração vivida pela consciência). O primeiro é um número, portanto, resultado de uma medida, mas que não aprisiona a duração, que flui, escoia e escapa ao confinamento da medida. Tenta reduzi-lo, retirando-lhe o caráter de continuidade, mas com a presunção de medi-la. Esse tempo numerado é geométrico, é uma medida de espaço, que quando justaposta é traduzida como duração. Mas a realidade temporal efetiva não é de fato capturada, é figurativamente *a mão espalmada e molhada que é soerguida após a imersão em água*. É, desse modo, apresentado como espaços justapostos, passado e futuro conscientemente concomitantes e, portanto, não é mais tempo. Então, o futuro não é previsível plenamente, porque se assim o fosse, seria o presente. Mas como é possível se definir algo que dura, permanece, se for constituído de instantes que nunca duram ? Apenas idealmente, não concretamente (Bergson APUD (PIETTRE, 1997)).

Entretanto, a medida desse espaço para dar sentido a uma noção de tempo cronológico e sequenciado (não contínuo), pressupõe a justaposição. Tal justaposição pressupõe a coexistência, que por sua vez pressupõe a interação, a co-influência, a co-relação, a interdependência.

Partindo desses argumentos e remetendo-os especificamente à medida do tempo na Epidemiologia, percebe-se que as contagens de doentes ou mortos relativos a alguma população de origem são mensurados em instantes finitos de tempo. Esses novos números, agora relativos, denominados “co-eficientes”, quando justapostos em “espaços” de tempo, para formarem em seu conjunto as séries históricas, não capturam o tempo duração, ou mesmo negam o tempo real.

Ao longo da história do método epidemiológico na abordagem de séries temporais, na passagem da medida à análise, apesar da compreensão explicitada do princípio de continuidade entre doença e saúde (AYRES, 1997), verifica-se que o escopo ferramental não apenas pouco mudou (apesar da importância que tem tido na disciplina), como deixou de

incorporar a noção básica e fundamental da interdependência entre os espaços justapostos de tempo das séries denominadas históricas.

Modernamente, os estudo de séries temporais em Epidemiologia são aplicados na indicação de potenciais fatores de risco, na monitoração da saúde de populações, na predição da ocorrência de eventos relacionados à saúde, no fornecimento de subsídios para explicações causais, no auxílio ao planejamento de saúde e na avaliação do impacto de intervenções ou programas (PEREIRA, 1995).

Para lidar metodologicamente com essas questões, os epidemiologistas recorrem à análise visual de gráficos (situação que depende da subjetividade e experiência do analista) de eventos relacionados à saúde, à reduções matemáticas que controlam a sazonalidade de séries sem tendência secular (diagrama de controle), à funções matemáticas “suavizadoras” que eliminam as variações sazonais e irregulares (médias móveis) e à modelos matemáticos que não incorporam o componente de autocorrelação (regressão) (SOUNIS, 1985).

Por fim, o método epidemiológico por possuir um forte componente matemático, a sua história, inclusive na abordagem das séries temporais, é muito vinculada à história da Estatística e da Informática, o que pressupõe que determinados avanços instrumentais da Epidemiologia não podem ser entendidos sem uma vinculação histórica com essas outras disciplinas.

Um método que pouco mudou

Mundialmente, a década de 30 foi marcada por uma intensa conturbação econômica. A quebra da Bolsa de Nova York foi apenas o primeiro episódio do que se seguiria com a depressão dos anos 30. A crise financeira desencadeou a crise econômica, de produção e de força de trabalho, o desemprego em massa e um aumento expressivo no índice de suicídios. Diversos países ligados à economia de mercado foram arrastados na crise levando o comércio internacional a declinar em cerca de um terço e os empréstimos internacionais caírem mais de 90%. Ao lado desses acontecimentos, Irving Fisher e Ragnar Frisch, em 1930, fundaram a

Econometric Society e em 1933, deu-se início a publicação da Revista *Econometrica* (MATOS, 1997) que representaram um primeiro esforço científico de confluência entre as correntes teóricas da macro economia com as necessidades pragmáticas de uma economia e de uma ciência da economia em crise (MONTORO FILHO, 1988).

Dadas as contingências inesperadas de uma crise que demonstrava a fragilidade do sistema econômico americano, a indústria passou a utilizar o que hoje se denomina controle estatístico de processo ou Carta de Controle Industrial¹ (Carta Controle de Shewhart) (CORDEIRO, 2001), na tentativa de estabelecer um maior controle sobre custos e a qualidade dos produtos, baseando-se na aleatoriedade dos fenômenos, e nas famílias de modelos de distribuições probabilísticas, visando o estabelecimento da produção considerada “fora de controle”.

Naquele contexto mundial de crise econômica, guerra e “busca de controle”, em 1946, um professor de biologia da *Stanford University*, Willis H. Rich, juntamente com um médico da Califórnia, Terry, M.C., publicaram um artigo sobre a aplicação da Carta de Controle Industrial aos estudos epidemiológicos, com a intenção de determinação mais confiável do início de uma epidemia (RICH et al., 1946). A sua argumentação era de que os gráficos usuais de incidência de doença no tempo não ofereciam critérios com bases mais seguras para o julgamento de quando uma epidemia poderia estar ocorrendo. Tal preocupação refletia a necessidade do controle das doenças transmissíveis que predominavam no quadro de morbimortalidade.

Coincidentemente, naquele mesmo ano, foi apresentado o primeiro computador digital eletrônico, ENIAC (*Electronic Numeric Integrator And Calculator*), construído pelo engenheiro elétrico John Presper Eckert Jr. e pelo físico John William Mauchly, na Escola Moore de Engenharia Elétrica, da Universidade da Pensilvânia, e pelo Laboratório de Pesquisas Balísticas do exército americano. O advento dos computadores viria a ser um marco no desenvolvimento metodológico científico nos anos seguintes, pois além de expandir a rapidez de cálculos matemáticos, o que potencializou a estatística matemática, mas principalmente, pela possibilidade concreta e empírica da construção de bases de dados eletrônicas.

¹ Método estatístico para detectar a ocorrência de causas especiais, isto é, perturbações anormais afetando o comportamento de um processo.

Para a Epidemiologia, foi um período de organização departamental e de esforços para o desenvolvimento de metodologias investigativas, ao lado da reemergência de discussões no campo teórico, como consequência das dificuldades geradas pela teoria monocausalista da doença, propiciando o surgimento de uma tendência ecológica na disciplina (ALMEIDA-FILHO, 1999) e que futuramente repercutiria na construção metodológica de tipologias que dessem conta dessa nova abordagem, incluindo as séries históricas.

Como consequência desse acúmulo de conhecimentos, e dos resultados obtidos com os esforços de guerra, os anos 50 propiciaram avanços nas diversas áreas da ciência. Entre os avanços da física, da química e da biologia, Moran, em 1950, propôs um índice denominado coeficiente de autocorrelação espacial ou Indicador de Moran, que teve um impacto muito grande na análise espacial e posteriormente na análise de séries temporais, sendo ainda hoje um indicador padrão de uso obrigatório e básico em geoprocessamento.

A década de 60 delineou-se por um movimento social, cultural e militar intensos, juntamente com o avanço tecnológico e científico sem precedentes (CUNHA, 2000). Esta foi uma década marcada por movimentos de natureza contestatória e prosperidade econômica, principalmente dos Estados Unidos, inclusive com a fundação, em 1966, da Intel (*Integrated Eletronics*) que lançou o primeiro microprocessador programável do mundo, o 8008. Talvez, tais eventos sociais e culturais tenham influenciado o pensamento de alguns cientistas com repercussões na política e na ética, mas aparentemente não interferiu nos alicerces paradigmáticos de uma ciência cada vez mais compartimentalizada e tecnológica, fazendo com que o 8008 tivesse muito mais impacto na ciência do que questões de outra natureza: fato esse refletido nas décadas seguintes.

Nessa mesma década, Brian MacMahon, Thomas Pugh e Johannes Ipsen, do Departamento de Epidemiologia da Escola de Saúde Pública da Universidade de Harvard, ao publicarem o livro intitulado *Epidemiologic Methods* faziam referência ao conceito de tempo como um elemento necessário para cada uma das medidas epidemiológicas e para o entendimento da associação causal. Porém, a sua mensuração, seria realizada em unidades arbitrárias semelhantes à uma medida de localização, como uma referência espacial, a exemplo das epidemias pontuais: epidemia de cólera em Londres (MACMAHON, 1960).

A medida do tempo também aparece de modo mais diversificado quando os autores citados anteriormente se referem à avaliação de tendências no tempo – mudança na frequência de doenças que ocorrem em longos períodos de tempo (anos ou décadas). Ou ainda quando comparam tendências da doença sob investigação com tendências de outras doenças e comparam a frequência da doença em diferentes períodos. Entretanto, observa-se nessa abordagem uma proposição de inferência temporal, através da predição da tendência no tempo: necessidade de extrapolação para além de dados atuais, com fins administrativos, a partir de um raciocínio sobre estratos de idade, chegando-se a uma predição qualitativa de possível incremento ou decremento.

Os autores admitem que os métodos de análise concernentes à sazonalidade (flutuações cíclicas) são simples, uma vez que os casos de doença são apresentados graficamente de acordo com o seu início em uma unidade de tempo definida (mês, etc.) e a interpretação vincula-se a uma correlação da variação sazonal da frequência da doença de interesse com a variação sazonal de outras circunstâncias possivelmente importantes (mortalidade e temperatura). Finalmente, descrevem-se considerações sobre pessoa-tempo sob risco (*person-years*), que é uma medida interativa de unidades observacionais distintas, do mesmo modo que calculam incidência, risco-relativo e risco atribuível, medidas essas que adotam o tempo secundariamente como um componente não dinâmico no método e na teoria.

Ainda naquele período, mais precisamente em 1961, foram desenvolvidos os testes para sazonalidade (ainda hoje usados em epidemiologia) de Capon e de Edwards. Em 1966, foi proposto por Weiss e Wolfowitz, o estimador de máxima verossimilhança generalizado, de uso abrangente em modelagem estatística, inclusive na análise das séries de tempo.

A década de 70 foi palco para a emergência e continuidade de golpes militares e ditaduras, particularmente na América Latina - mantendo-a em posição desfavorável não apenas em relação às questões de liberdade e bem-estar social, mas também, dependente do desenvolvimento científico e tecnológico estrangeiros - e o surgimento de novas crises econômicas relacionadas ao controle da exploração e da produção de petróleo, levando a uma nova recessão econômica mundial no início da década de 80.

Apesar disso, o microcomputador veio a se tornar popular em 1975, quando a revista americana *Popular Electronics* apresentou "o primeiro kit de minicomputador do mundo",

que chegava para competir com os modelos comerciais de maior porte. A revista apresentava as instruções de como montar um Altair 8800 na residência, que deveria ser comprado por reembolso postal, e que veio a ser o primeiro PC (*personal computer*), fabricado pela MITS (*Micro Instruments and Telemetry Systems*), empresa fundada em 1969 pelo engenheiro Edward Roberts, um ex-oficial da Força Aérea Americana (GONICK, 1984).

O desenvolvimento da microinformática, a partir da década de 70 e a conseqüente produção de computadores em escala industrial a preços mais acessíveis, possibilitou a sua utilização como uma ferramenta que teve um impacto muito grande nas diversas áreas da ciência, inclusive na estatística e na epidemiologia.

Naquele contexto, em 1970, Box e Jenkins propuseram a família de modelos ARIMA (autorregressivos integrados com média móvel) para a análise de séries temporais e Durbin-Watson desenvolveram um teste adequado para verificar autocorrelação de primeira ordem em modelo de regressão. Essa proposição viria metodologicamente dar conta especialmente de problemas relativos às predições em séries temporais, pois oferecia melhores estimativas de dados no futuro e que sempre foram de grande interesse para o campo da Economia.

Nesse ínterim, em 1976, Box e Jenkins propunham o modelo de função de transferência para a análise de séries temporais, o que alargou consideravelmente a aplicação dos modelos ARIMA propostos anteriormente, possibilitando a avaliação de impactos ou descontinuidades de séries históricas econômicas, assim como a correlação multiseriada temporal.

Nesse mesmo período, Abraham M. Lilienfeld, do Departamento de Epidemiologia da Escola de Higiene e Saúde Pública da Universidade *Johns Hopkins*, lançou o livro *Foundations of Epidemiology*, no qual refere-se ao tempo como uma das dimensões de interesse para a Epidemiologia. Entretanto, o autor tão somente faz alusões à distribuição de uma doença no tempo - período de incubação ou a tendência de mortalidade e de incidência no tempo - considerando aspectos, tais como a sazonalidade, ou comparando coortes de mortalidade por estratos de idade (LILIENFELD, 1976), mas não acrescentando conteúdos à abordagem metodológica das séries históricas e de certo modo distanciado das inovações metodológicas ao seu entorno.

No mundo, a década de 80 iniciou-se com as crises política e econômica no bloco socialista, levando a movimentos democráticos e neoliberais, enquanto os países capitalistas voltavam a crescer economicamente, o que voltaria a impulsionar os investimentos metodológicos e teóricos da ciência.

Na Estatística, entre 82 e 86, foi proposta uma classe de modelos, pertencentes à teoria de séries de tempo, denominados modelos de heteroscedasticidade condicional autorregressiva, que tratam de superar as ineficiências estruturais dos modelos financeiros e em 1982 foi desenvolvido o modelo de regressão multivariado de Cox, uma expansão natural dos modelos de riscos proporcionais publicados na década anterior, para a análise de sobrevivência que viria a ser utilizado pela Epidemiologia posteriormente.

No campo da Epidemiologia propriamente dito, a década de 80 foi marcada pelo desenvolvimento metodológico sistematizador acentuado, impulsionado também pela disponibilidade dos microcomputadores. Foi publicada uma sistematização mais ampliada de configurações tipológicas na abordagem investigativa (desenhos de estudo), medidas de associação e de impacto, confundimento, interação e modelagem matemática (KLEINBAUM et al., 1982). As diversas possibilidades de delineamentos de pesquisa passaram a resultar da combinação e integração dos conceitos de temporalidade, unidade de observação, experimentação, fator de estudo e a randomização. No que diz respeito à temporalidade nesses delineamentos, os autores ou estabelecem um ponto de corte em um momento específico e arbitrário de tempo ou definem um acompanhamento ao longo do tempo para a observação dos efeitos de interesse.

Dentre os procedimentos investigativos propostos, os desenhos quase experimentais são aqueles nos quais o conceito de tempo recebe uma diferenciação dos demais desenhos de investigação, pois este passa a fazer parte da unidade de observação (COOK et al., 1979). Entretanto, tal desenvolvimento metodológico ainda não incorpora técnicas que contemplem a dinâmica dos fenômenos historicamente, cumulativamente, temporalmente.

No Brasil, naquele período, epidemiologistas passaram a publicar livros-texto sobre Epidemiologia e Saúde Pública. Dentro da nomenclatura adotada em tais livros, a abordagem metodológica das séries temporais – cunhada de estudo da variação temporal ou distribuição cronológica - foi apresentada como fazendo parte da epidemiologia descritiva. As

características das séries foram denominadas de variação sazonal, tendência secular e movimento cíclico e variação acidental (FORATTINI, 1980; LESER et al., 1985; SOUNIS, 1985; MALETTA, 1988).

A variação sazonal seria o estudo da distribuição das frequências de doenças por meses e por estações do ano que se apresenta graficamente como uma flutuação. Já o estudo da distribuição de uma doença em um longo período de tempo (décadas), seria a tendência secular ou marcha. Flutuações da curva de frequência que se apresentam com uma certa regularidade, seria a variação cíclica. Enquanto que acentuados aumentos da frequência corresponderiam à variação acidental.

Talvez, a formação de estatístico de Sounis o tenha influenciado a preocupar-se em apresentar com mais detalhes matemáticos aspectos relativos à análise de séries temporais, pois se referiu a métodos de ajustamento de dados cronológicos a modelos matemáticos. Segundo o autor, levando-se em consideração o grau de precisão, tais métodos poderiam ser classificados em: aproximados (mão-livre, semi-médias e média móvel) e rigorosos (mínimos quadrados e dos momentos). Porém, nessa abordagem, um dos principais conceitos da análise de séries históricas, a autocorrelação temporal, não foi por ele contemplado.

Vale ressaltar a contribuição valiosa para a área de séries temporais no Brasil de Pedro Alberto Morettin, estatístico da USP, que publicou a partir da década de 80, vários livros sobre o tema. Entretanto, o autor produziu e direcionou a aplicação da análise de séries temporais para as áreas de Economia e Engenharia, sendo por isso homenageado por ocasião da realização da 9ª Escola de Séries Temporais e Econometria em 2000.

Em 1987, uma abordagem sobre a medida do tempo para a epidemiologia foi descrita, através de situações nas quais a observação da frequência de doenças no tempo pode sugerir possíveis relações etiológicas (HENNEKENS et al., 1987). Para tanto, os autores exemplificam com os estudos sobre a associação entre o uso de talidomida e malformação congênita entre 1957 e 1961; a Síndrome toxêmica em mulheres que fizeram uso prévio de uma determinada marca de tampão absorvente entre 1977 e 1981, sendo que com a retirada do produto do mercado houve uma drástica queda do número de casos (estudo de intervenção). Além disso, os autores chamam a atenção para fatores que podem produzir tendência secular em uma série histórica, tais como: mudanças nas técnicas diagnósticas, mudanças na acurácia da contagem

populacional sob um determinado risco, conduzindo a alterações nas taxas; mudanças na estrutura etária de uma população; mudanças na sobrevivência de uma doença devido a melhorias de tratamento e mudanças na incidência real de uma doença em consequência de alterações ambientais ou do estilo de vida.

O contexto internacional da década de 90 revelou uma realidade de continuidade da pobreza e de desigual distribuição de renda, principalmente na América Latina, constantes guerras no Oriente Médio e um período de prosperidade dos países do primeiro mundo, especialmente dos EUA (PETRAS, 2002), permitindo conciliar em uma mesma década fome, violência e avanços da ciência, como por exemplo a clonagem genética e a computação neural (técnica que permite também a análise de séries temporais).

Naquela década, (ROTHMAN et al., 1998) demarcaram os desenhos de tendência temporal, classificando como tendência exploratória no tempo ou simplesmente estudo de séries temporais, aqueles desenhos de investigação nos quais são realizadas comparações de taxas de doenças no tempo em uma população geograficamente delimitada e situações de predição temporal, admitindo o uso de modelagem ARIMA em Epidemiologia. Denotaram a importância do estudo de séries temporais, devido à possibilidade de avaliação de impacto de intervenções populacionais, assim como a estimativa de associações em estudos ecológicos mistos. Finalmente, caracterizaram como estudos analíticos de tendência temporal, aqueles que avaliam associação ecológica entre a mudança no nível de exposição média ou prevalência e mudança na taxa de doença em uma população delimitada.

Como limitação para a inferência, a partir desses delineamentos, os referidos autores enfatizaram que mudanças na classificação de doenças e critérios diagnósticos ao longo do tempo são capazes de produzir distorções nas tendências das taxas de doenças e consequentes vieses nas interpretações a partir delas, assim como a duração do tempo de latência entre a exposição ao primeiro fator de risco e a detecção da doença que, mormente é desconhecido.

Corroborando esses aspectos, um estudo da avaliação do impacto da mudança da Classificação Internacional de Doenças em 1996, efetivado com dados do Sistema de Informação de Mortalidade do Rio Grande do Sul, verificou um aumento artificial de 67,9 % de óbitos classificados no capítulo de doenças infecciosas e parasitárias, através de codificação simultânea pela CID 9 e pela CID 10 (LAURENTI et al., 1998).

Mesmo contemporaneamente, quando da análise de dados epidemiológicos, que subsidiam a avaliação de serviços ou políticas de promoção da saúde, a metodologia adotada opera sobre um conceito de estaticidade, que não incorpora técnicas de identificação dinâmica e consequentemente possui sérias limitações (ALMEIDA et al., 1987).

Outro aspecto a ser considerado como de fundamental importância para a construção e análise de séries temporais, refere-se às escalas de mensuração adotadas para a medida das grandezas sob investigação. A medida do tempo para a Epidemiologia efetiva-se através de uma escala de mensuração definida para tempos de doença e morte em populações humanas. Entretanto, o entendimento do processo saúde-doença em uma perspectiva ecológica - não tradicional -, dinâmica e interativa sugere a adoção de uma escala que mensure para atender tais objetivos. Nesse processo interativo, além das limitações de método, há que se considerar as diferenças escalares de processos no meio ambiente e nos organismos envolvidos, como por exemplo, as árvores têm ciclos de vida, que são muito mais longos do que os ciclos de vida dos produtores primários no oceano, o *phytoplankton*. Analogamente, sociedades possuem um ciclo de vida maior do que o ciclo de vida de cada indivíduo que a compõe. Estas diferenças nas escalas de tempo têm uma relevância imediata no que concerne ao estudo de série temporais, uma vez que as mensurações dos objetos investigativos são completamente dependentes das suas escalas de medição e, portanto, delineiam a priori os resultados da mensuração, emprestando-lhes forma e imprecisão.

As sucessões longas de tempo podem ser bastante diferentes para fenômenos que acontecem em terra, no oceano ou em sociedades, por exemplo, não somente em termos das vidas dos seres, mas também em relação aos processos temporais do ambiente ou do contexto histórico no qual as plantas, os animais, ou os homens estão inseridos em compartilhamento (POWELL et al., 1995).

Considerando, a partir da década de 90, as publicações brasileiras em livros-texto de epidemiologia (PEREIRA, 1995; ROUQUAYROL et al., 1999; MEDRONHO, 2002), no que diz respeito às séries temporais, os autores detiveram-se aos já conhecidos e difundidos métodos de análise.

Conclusões

Desse modo, à Epidemiologia, na operacionalização da categoria tempo, no desenvolvimento e aplicação do seu método em séries temporais, apesar da centralidade e importância da dimensão temporal no que se refere aos interesses pragmáticos do campo, não incorporou substancialmente técnicas que promovessem uma melhor compreensão quantitativa dos eventos contados no tempo, assim como o cálculo de suas probabilidades. Ou seja, nesses tempos atuais em que o discurso da complexidade tornou-se cotidiano (CASTIEL et al., 1992), intuitivo e óbvio, apesar disso, o domínio da análise estatística de fenômenos de movimento complexo, ainda não se deu no âmbito da Epidemiologia, mesmo na visão mais restrita e específica do conceito de complexidade.

Nesse sentido, a modelagem proposta por Box e Jenkins, dadas as suas características, viria a preencher esse domínio, mas a não utilização do método pelos epidemiologistas é apenas uma questão de tempo e tecnologia. Uma vez que sejam lançados no mercado de informática programas de computador de baixo custo que efetuem as seleções de modelos ARIMA e as estimativas dos seus parâmetros automaticamente, ou na medida que sejam implementados algoritmos que realizem tal tarefa nos pacotes estatísticos existentes, ele será o método de escolha para a análise das séries temporais entre os epidemiologistas.

Por outro lado, numerosa tem sido a publicação nacional sobre Epidemiologia e Epidemiologia Aplicada e cujo enfoque tem variado, conforme os contextos sociais vivenciados pelo país e pelo mercado acadêmico, particularmente a partir da década de 70. A princípio, a Epidemiologia contribuiu como um instrumento geral de informação, com caráter de denúncia de desigualdades. Atualmente, caracteriza-se pela criação de centros de excelência em assuntos específicos, pelo estímulo à produção cada vez maior e mais acirrada de comunicações científicas, ponderada monetariamente ou, mais raramente, com realce epistemológico em busca de uma identidade científica própria, aceitando e mesclando a afluência metodológica de outros campos do saber. Estas particularidades, na maioria das vezes, decorrem da lógica da superespecialização, modelada no paradigma vigente da ciência moderna, concretizada por epidemiologistas especializados em grupos de risco ou em temas específicos e de conteúdos como, por exemplo, a epidemiologia ocupacional, a epidemiologia clínica ou medicina baseada em evidências (como se assim já não o fosse), a epidemiologia

molecular (uma ligação semanticamente estranha de conceitos e de domínios), a paleoepidemiologia (cuja validação instrumental é questionável dada a natureza do fenômeno mensurado), etc.

Nesse outro sentido, a defesa da tese do emprego dessa metodologia no contexto Epidemiológico, portanto deve ser efetivada, pelo primado do método no paradigma vigente, no qual o conceito de tempo adotado pela Epidemiologia, ao longo de sua história, é uma medida ecológica de localização, mas que por outro lado ao defendê-la, expõe mais uma vez o paradoxo da superespecialização, numa visão crítica do conceito de complexidade.

Desse modo, muitas séries históricas podem ser vislumbradas a partir da história seriada da humanidade, que fundamentalmente nos auxilia na compreensão do contexto epidemiológico atual e nos aponta possibilidades futuras. O tempo aqui tem o sentido de tempo sequenciado, unidirecional, espacial e, portanto, cartográfico, mas que indubitavelmente encerra as percepções de tempo, tempos dentro de tempos, agregados, interações e o acúmulo contorcido e progressivo de experiências e expectativas, mas que sugere fortemente a incerteza, a instabilidade do projeto humano e que, paralela e humanamente, deixa transparecer a sua essência de espécie dominante. Conceito esse que se situa muito além do conceito de alta valência ecológica ou de predação, que encerra grande capacidade de proliferação, exploração, autofagismo e solidão coletiva, muitas vezes “pisando nas próprias pegadas” e sonhando com o vir-a-ser ou o porvir, mas enquanto isso, exercitando com grande maestria o seu maior dom, a retórica.

Referências bibliográficas

- ALMEIDA, R.M.V.R.;NOBRE, F.F.;PANERAI, R.B. Caracterização dinâmica de séries históricas de determinantes de saúde. *Revista Brasileira de Engenharia*, 5(1): p. 33-43. 1987.
- ALMEIDA-FILHO, N. *Uma breve história da Epidemiologia*. In: Epidemiologia e saúde. Rio de Janeiro: MEDSI, 1999. p. 1-13.
- AYRES, J.R.D.C.M. *Sobre o risco: para compreender a epidemiologia*. São Paulo: Hucitec, 1997. 327 p.
- CASTIEL, L.D.;SCHRAMM, F.R. Processo saúde/doença e complexidade em Epidemiologia. *Cadernos de Saúde Pública*, 8(4): p. 379-90. 1992.
- COOK, T.D.;CAMPBELL, D.T. *Quasi-experimentation*. U.S.A.: Houghton Mifflin Company, 1979.
- CORDEIRO, G.M. *Cronologia de alguns conceitos e fatos importantes da Estatística*. ABE. Disponível em. <http://www.est.ufba.br/historia.html>. Acesso: 2002
- CUNHA, M.H.G. *Anos 60*. Cunha, Moisés H.G. Disponível em. <http://www.angelfire.com/rock/jhendrix70/sixty.htm>. Acesso: 2002
- FORATTINI, O.P. *Epidemiologia geral*. São Paulo: Edgard Blücher, Universidade de São Paulo, 1980. 259 p.
- GONICK, L. *Introdução Ilustrada a Computação - com muito humor*. São Paulo: Harper & Row do Brasil Ltda, 1984.
- HENNEKENS, C.H.;BURING, J.E.;MAYRENT, S.L. *Epidemiology in medicine*. 1st. ed. Boston: Little Brown, 1987. xv, 383 p.
- KLEINBAUM, D.G.;KUPPER, L.L.;MORGENSTERN, H. *Epidemiologic research*. Belmont, Calif.: Lifetime Learning Publications, 1982. 529 p.
- LAURENTI, R.;GRASSI, P.R. Implicações da introdução da 10ª revisão da classificação internacional de doenças em análise de tendência da mortalidade por causas. *IESUS*, VII(3), Jul/Set: p. 43-7. 1998.
- LESER, W., et al. *Elementos de epidemiologia geral*. Rio de Janeiro: Atheneu, 1985. 175 p. (Medicina preventiva, epidemiologia, saúde pública).
- LILIENFELD, A.M. *Foundations of epidemiology*. New York: Oxford University Press, 1976. 283 p.
- MACMAHON, B. *Epidemiologic methods*. Boston,: Little Brown, 1960. 302 p.
- MALETTA, C.H.M. *Epidemiologia e saúde pública*. Rio de Janeiro: Atheneu, 1988. 184 p. (Epidemiologia, saúde pública).
- MATOS, O.C.D. *Econometria básica*. 2. ed. São Paulo: Atlas, 1997. 246 p.

- MEDRONHO, R.A. *Epidemiologia*. São Paulo: Atheneu, 2002. 493 p.
- MONTORO FILHO, A.F. *Manual de Economia*. São Paulo: Saraiva, 1988. 443 p.
- PEREIRA, M.G. *Epidemiologia*. Rio de Janeiro: Guanabara Koogan S.A., 1995. 583 p.
- PETRAS, J. *El menemismo: el contexto internacional de la década del '90*. Revista Herramienta. Disponível em. <http://www.herramienta.com.ar/12/12-2.html>. Acesso: 2002
- PIETTRE, B. *Filosofia e ciência do tempo*. Bauru: EDUSC, 1997. 223 p.
- POWELL, T.M.;STEELE, J.H. *Ecological time series*. New York: Chapman & Hall, 1995. xvi, 491 p.
- RICH, W.H.;TERRY, M.C. The industrial "Control Chart" applied to the study of epidemics. *Public Health Reports*, 61(42): p. 1501-11. 1946.
- ROTHMAN, K.J.;GREENLAND, S. *Modern epidemiology*. 2. ed. Philadelphia, PA: Lippincott-Raven, 1998. xiii, 737 p.
- ROUQUAYROL, M.Z.;ALMEIDA-FILHO, N. *Epidemiologia & Saúde*. 5. ed. Rio de Janeiro: MEDSI, 1999. 600 p.
- SOUNIS, E. *Epidemiologia geral*. 1. ed. São Paulo: Universidade Federal do Paraná, 1985. 114 p. (Medicina preventiva, epidemiologia, saúde pública).

ARTIGO 2

Aplicação do método de modelagem de Box e Jenkins, em séries temporais epidemiológicas distintas.

Parte I

Internações e regulação da oferta e do acesso: uma avaliação estatística de impacto.

Mortalidade por diarreia: a descrição dos componentes de uma série temporal.

Artigo 2 - Aplicação do método de modelagem de Box e Jenkins, em séries temporais epidemiológicas distintas. Parte I - Internações e regulação da oferta e do acesso: uma avaliação estatística de impacto e mortalidade por diarreia: a descrição dos componentes de uma série temporal.

Resumo

A análise de séries de tempo em Epidemiologia, historicamente, tem privilegiado o emprego de técnicas descritivas visuais. Após a proposição de Box e Jenkins para a análise de séries temporais que considera a autocorrelação, avalia a estacionariedade, estima parâmetros de intervenção e efetiva análise multivariada de séries através das funções de transferência, diversas ciências a têm utilizado, em particular a Econometria.

No intuito de se averiguar o potencial de aplicação do método ARIMA em contextos epidemiológicos, foram analisadas duas séries temporais: a primeira série foi construída a partir dos dados de internações hospitalares totais em Vitória da Conquista/Bahia (de janeiro de 1995 a junho de 2002) com o objetivo de se estimar e testar o parâmetro de intervenção, produzido pela adoção da regulação sobre as internações efetuada pela prefeitura daquela cidade em outubro de 1999. A estimativa do parâmetro de intervenção revelou um impacto significativo no número de internações em Vitória da Conquista, da ordem de 30 %, representando uma redução média absoluta de cerca de 1068 internações em apenas um mês.

Para a construção da série temporal mensal de mortalidade de crianças menores de cinco anos, no período de 1979 a 1998, na Bahia foram considerados os Códigos Internacionais de Doenças 009 (infecções intestinais mal definidas) e A09 (diarreia e gastroenterite de origem infecciosa presumível) para diarreia, 9ª e 10ª revisões, respectivamente. A análise dessa série objetivou descrever a sua estrutura paramétrica temporal e efetuar predições. Para tanto, ajustou-se um modelo ARIMA(0,1,1)(1,0,1)₁₂, que inclui uma diferenciação serial, um termo de média móvel serial trimestral remanescente (-0,15), um termo autorregressivo sazonal (0,91) e um termo de média móvel sazonal (- 0,73), parâmetros esses, todos significantes, e um resíduo de 0,23, obtendo-se desse modo, o seguinte modelo de predição: Número de óbitos previstos_t=0,91 Y_{t-12}-0,15 a_{t-4}-0,73 a_{t-12}+0,23 a_t.

Desse modo, fica evidente a aplicabilidade da proposição metodológica de Box e Jenkins em contextos epidemiológicos, desde que sejam atendidos os seus pressupostos, podendo-se realizá-la com números absolutos ou taxas, não apenas modernizando o arsenal metodológico na abordagem de séries históricas dinâmica e com isso ampliando as possibilidades analíticas no campo, assim como, ao visar primordialmente o diagnóstico de possíveis dependências seriais e sazonais, obter conseqüentemente melhores estimativas para previsões e avaliações de impacto em fenômenos de movimento complexo.

Descritores: fenômenos de movimento complexo; séries temporais; ARIMA; Epidemiologia

Article 2 - Use of the Box and Jenkins model method in different epidemiological time series. Part I: Hospitalization and regulation of offer and access: a statistical evaluation of impact and mortality due to diarrhea: a description of the components of a time series.

Abstract

In epidemiology, the time series analysis has historically given preference to the use of visual descriptive techniques. Following Box and Jenkins' proposal for time series analyses, which considers auto-correlation, evaluates stationarity, estimates intervention parameters and carries out series multivariate analysis using transference functions, various branches of science, particularly Econometry, have made use of it.

The purpose of this study was to verify the potential of the application of the ARIMA method in epidemiological contexts. Two temporal series were analyzed: the first series was based on data from all hospital admissions in the city of Vitória da Conquista, Bahia, from January 1995 to June 2002 with the aim of estimating and testing the parameter of intervention produced by the adoption of regulations regarding hospital admissions established by the city council in October 1999. The estimate of the intervention parameter showed a significant impact in the number of hospital admissions in Vitória da Conquista, around 30%, representing an absolute mean reduction of about 1068 admissions in just one month.

The second monthly temporal series on mortality in children under five years of age in Bahia between 1979 and 1998 was constructed using the International Classification of Diseases, codes 009 (ill-defined intestinal infections) and A09 (diarrhea and gastroenteritis of presumed infectious origin) for diarrhea, 9th and 10th editions, respectively. The analysis of this series was aimed at describing its temporal parametric structure and carrying out predictions. An ARIMA model (0,1,1) (1,0,1)₁₂ was therefore adjusted to include a serial differentiation, a three-month serial moving average remainder term (-0.15), a seasonal auto-regressive term (0.91) and a seasonal moving average term (-0.73), all of these parameters being significant, and a residue of 0.23, thus obtaining the following predictive model:
 Number of expected deaths_t = 0.91Y_{t-12} - 0.15a_{t-4} - 0.73a_{t-12} + 0.23a_t.

The applicability of the Box and Jenkins methodological proposal in epidemiological contexts is therefore evident as long as their assumptions are met. This proposal can be used with either absolute numbers or rates, not only to modernize the methodological arsenal in dealing with dynamic historical series, thus increasing the analytical possibilities within the field, but primarily in aiming to diagnose possible serial and seasonal dependencies, to consequently obtain better estimates for the forecasts and evaluations of impact in complex motion phenomena.

Key words: complex motion phenomena, time based series, ARIMA, epidemiology

Introdução

A constituição de séries históricas tem sido uma estratégia de interesse para o acompanhamento de eventos relacionados às condições de vida e morte desde as primeiras contagens de “almas”, terras e bens na história da humanidade e com propósitos variados. A perspectiva de um desenvolvimento metodológico que permitisse a possibilidade da predição de acontecimentos futuros, sempre foi de grande interesse para a ciência. Historicamente, particularmente no campo da Epidemiologia, o método empregado para alcançar tais objetivos tem enfatizado procedimentos gráficos, e mais raramente, com a adoção da teoria das probabilidades, vem progressivamente incorporando técnicas analíticas mais complexas.

Um dos usos da Epidemiologia se dá na avaliação de impacto de programas ou intervenções através da realização de estudos do tipo antes-depois (COOK et al., 1979; KLEINBAUM et al., 1982). Entretanto, se houverem dados disponíveis para a construção de uma série histórica, tal procedimento avaliativo pode ser melhor efetivado, uma vez que a estimativa do parâmetro do impacto ou intervenção será realizada a partir de um modelo dinâmico e probabilístico (ALMEIDA et al., 1987).

Do mesmo modo, de grande interesse para a Epidemiologia é o acompanhamento de casos de doenças em populações, a observância e monitoramento de situação de saúde de uma comunidade. Porém, mesmo modernamente, as técnicas empregadas variam desde a inspeção visual de gráficos de casos de doença no tempo e a aplicação de diagramas de controle, até a suavização de curvas ou modelos de regressão linear (PEREIRA, 1995). Entretanto, todas as técnicas empregadas ou possuem um forte componente subjetivo, ou controlam aspectos importantes do fluxo de uma série histórica (como a sazonalidade ou a tendência secular), ou não consideram o fenômeno da autocorrelação como característica primordial da dinamicidade de uma série temporal.

Entretanto, após a divulgação dos modelos autorregressivos integrados com média móvel (ARIMA) (BOX et al., 1976), a disponibilidade e acesso de microcomputadores potentes, a análise de séries históricas passou a ser efetivada com maior objetividade e a contemplar a dinamicidade e a minimizar as estimativas paramétricas assim como as predições no tempo,

não sendo admissível que a Epidemiologia deixe de incorporá-los efetivamente ao seu arsenal metodológico.

Assim, esse artigo objetiva demonstrar a aplicação do método de modelagem de séries temporais dinâmicas, através da proposição de Box e Jenkins, visando dois contextos de interesse epidemiológico, na descrição da estrutura, na avaliação estatística de impacto e na predição de séries históricas.

Metodologia

Para a análise das séries temporais epidemiológicas empregou-se a modelagem ARIMA (autorregressivos integrados com média móvel), através da qual, foram estimados os parâmetros estruturais, após a verificação da estacionariedade das séries. Para tanto, adotou-se a abordagem em estágios (vide apêndices), estratégia sugerida por Box e Jenkins para a identificação dos modelos.

As séries

A série temporal de internações hospitalares de Vitória da Conquista foi constituída com o número total de internações agregadas mensalmente com início em janeiro de 1995 até junho de 2002. Para a modelagem da série, criou-se uma variável sequencial iniciada com o número um, correspondendo ao mês de janeiro de 1995 e uma variável indicadora com o objetivo de avaliar a significância estatística da intervenção, cujo valor zero foi a ela imputada para antes do ponto temporal número 58 (outubro de 1999 – mês da intervenção/regulação) e atribuiu-se o valor um, daquele ponto em diante, por considerar-se o efeito da intervenção como tendo um caráter abrupto e permanente.

Obteve-se inicialmente a estatística descritiva univariada para o número de internações antes e após outubro de 1999. Em seguida, gerou-se um correlograma² para os primeiros atrasos

² Gráfico de autocorrelações

(lags) da série e calculou-se a estatística de teste de Bartlett (*Cumulative periodogram white noise test*) (MCDOWELL, 2002) com o intuito de averiguar a aleatoriedade da série. Após a fase inicial de identificação descrita anteriormente, efetuou-se a verificação dos pressupostos do modelo tentativo e transformação logarítmica quando necessário (visando estabilização da variância ao longo do tempo). Posteriormente, estimou-se a medida da intervenção, assumindo um caráter abrupto e permanente para o seu efeito, através de uma variável indicadora (COOK et al., 1979).

Foram considerados os Códigos Internacionais de Doenças 009 (infecções intestinais mal definidas) e A09 (diarreia e gastroenterite de origem infecciosa presumível) para diarreia, 9ª e 10ª revisões, respectivamente, para a construção da série temporal de mortalidade de crianças menores de cinco anos, no período de 1979 a 1998, agregados mensalmente, na Bahia a partir do CD-ROM de mortalidade divulgado no Brasil pelo DATASUS/FUNASA. Nessa série foram imputados valores médios de meses adjacentes para o mês de agosto de 1995, janeiro e fevereiro de 1998, por serem valores muito diferentes (inferiores) do padrão médio da série.

Fonte de dados

Fonte de dados – Vitória da Conquista, Bahia

Os dados necessários para a elaboração da série temporal sobre internações hospitalares de Vitória da Conquista foram obtidos mediante acesso ao “*site*” do DATASUS/FUNASA (BRASIL, 2002), enquanto que as informações pertinentes ao período de regulação³ foram conseguidas junto à Prefeitura Municipal de Vitória da Conquista (VITÓRIA DA CONQUISTA, 2002).

Confiabilidade dos dados nos formulários de autorização hospitalar

³ Serviço de controle, avaliação e auditoria do SUS no município, avalia a oferta de serviços, autoriza as internações hospitalares (as eletivas são autorizadas previamente) e investiga as denúncias e reclamações formuladas pelos usuários do SUS.

Em 1994, um estudo foi realizado sobre a qualidade das informações contidas no banco de dados dos formulários de Autorização de Internação Hospitalar (AIH) (VERAS et al., 1994). O estudo efetuou-se sobre uma amostra de 1.934 formulários AIH preenchidos pelos 10 hospitais privados contratados da cidade do Rio de Janeiro, em 1986. Entre os vários objetivos pretendidos pelo estudo, a análise da validade da informação nos prontuários médicos efetivou-se tendo como referência o diagnóstico principal.

Para as variáveis demográficas e administrativas, as autoras concluíram por uma alta confiabilidade de uma maneira geral. Para as variáveis clínicas (diagnóstico) houve uma tendência de maior confiabilidade para diagnósticos com maior nível de agregação (três dígitos) do que com menor nível de agregação.

(Lebrão APUD Veras & Martins, 1994) obteve um nível de concordância para os diagnósticos de 82,5 % (codificação de três dígitos), resultado esse, segundo as autoras, semelhante ao obtido em seu estudo, apesar de diferentes metodologias empregadas. As autoras também observaram uma maior confiabilidade para os diagnósticos com alta frequência no banco de dados. Verificaram que os diagnósticos anotados nos formulários AIH são codificados mais frequentemente nos dígitos 8 ou 9 – classificações geralmente inespecíficas para a categoria diagnóstica. Na maioria dos prontuários analisados (98,5 %) havia um diagnóstico principal, sendo que em 85,4% destes a equipe da pesquisa avaliou esse diagnóstico como sendo correto.

Quanto aos procedimentos realizados, as autoras encontraram uma alta concordância. Segundo elas, *“A análise do impacto das discordâncias (341 observações) no reembolso dos hospitais apontou para a ocorrência, com uma frequência considerada alta, de discordâncias que resultam no aumento deste valor em favor do hospital.”*

Em um estudo de morbidade relacionada à esquistossomose em quatro estados do Nordeste brasileiros, (CARMO et al., 1995), no qual foram revistos 2.030 prontuários de pacientes hospitalizados, identificados através do formulário de autorização de internamento hospitalar (AIH), observou-se uma concordância de 90 % entre os dados da AIH e os registros hospitalares de 178 pacientes do Estado da Bahia, para a referida morbidade.

Fonte de dados de mortalidade

Os dados necessários para a construção da série de mortalidade foram provenientes do CD-ROM de mortalidade divulgado no Brasil pelo DATASUS/FUNASA.

Confiabilidade dos dados de mortalidade

De acordo com o Ministério da Saúde, admite-se que a cobertura dos óbitos pelo sistema é de 80% e há uma pequena diferença de acordo com o sexo: 85 % para o sexo masculino e 79 % para o feminino (PAES et al., 1999). Algumas regiões do País apresentam um maior sub-registro, como as regiões Norte, Nordeste e Centro-Oeste, assim como a proporção de óbitos por Causas Mal Definidas (OTERO, 2001).

Portanto, segundo o Ministério da Saúde, o total de óbitos para o Brasil gerado pelo sistema não expressa, quantitativamente, a magnitude da mortalidade do país, não refletindo, igualmente, o padrão geral da mesma, por estarem as regiões Sul e Sudeste, melhor representadas, com características diferentes das demais. Daí, não ser recomendável o cálculo de coeficientes que usem população como denominador, sugerindo-se, como alternativa, o emprego de indicadores de mortalidade proporcional (BRASIL, 2001).

Admite-se que as inconsistências relacionadas a algumas variáveis, tais como sexo, idade e causa básica, foram em grande parte corrigidas, ficando um resíduo inferior a 0,5% do total de óbitos. Entretanto, a causa básica da morte, continua a ser um problema que depende do correto preenchimento do atestado de óbito por parte dos médicos e da assistência médica disponível.

Resultados

Internações e regulação

O Brasil dispunha em junho de 2002 de 5.714 hospitais e 439.577 leitos hospitalares, a Bahia de 462 e 28.582 e Salvador de 44 e 6.985, respectivamente. De janeiro de 1984 a junho de 2002, foram realizadas 216.292.800 internações no Brasil, 14.147.091 na Bahia, e 14.600.075 em Salvador a um valor de R\$ 86.788.681.918.446,91, R\$ 4.610.489.859.437,08 e R\$ 1.210.534.854.479,78 , respectivamente (BRASIL, 2002).

Na Bahia foram pagas 661.572 internações devidas a infecções intestinais mal definidas (4,67 % do total de internações), a um valor de R\$ 223.678.717.473,22, correspondendo a 4,85 % do montante total pago no período (BRASIL, 2002).

O município de Vitória da Conquista na Bahia ocupa o segundo lugar quanto ao número de internações efetuadas no período e o quarto quanto aos valores pagos por aquelas mesmas internações (Tabela 1).

Tabela 1 – Número de internações hospitalares e valores pagos em Reais de acordo com o município - Bahia, 1995 a 2002.

Município	Número de internações	Valor em Real
Salvador	1.314.040	616.325.258,55
Feira de Santana	265.369	88.432.006,91
Itabuna	255.262	81.423.981,39
Vitória da Conquista	293.115	66.901.022,28
Outros	5.536.785	1.093.622.696,27
TOTAL	7.664.571	1.946.704.965,40

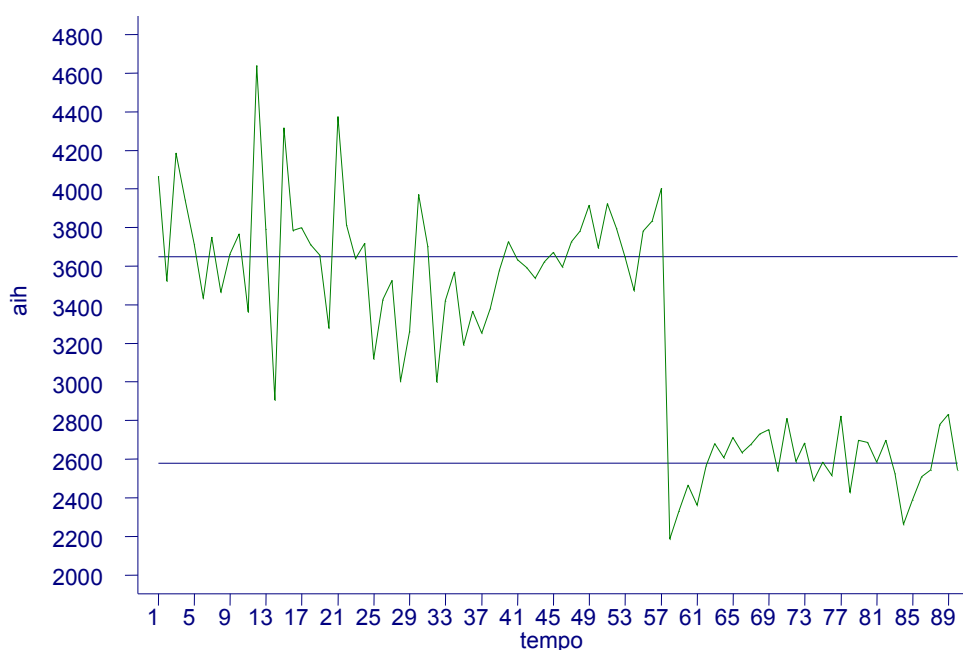
Nota: até Junho de 2002.

Descrição da série

Ao observar-se o número de internações hospitalares em Vitória da Conquista ao longo do período de estudo, estas se apresentam como uma série temporal interrompida, agregada mensalmente, na qual pode-se perceber uma composição em dois níveis distintos (Figuras 1, 2 e Tabela 2): o primeiro, anterior a outubro de 1999 (ponto 58) e o segundo após aquele ponto, inclusive.

Nota-se, à inspeção visual, haver estacionariedade na média (ausência de tendência secular), mas não na variância ao longo do tempo, com um coeficiente de variação de 8,93 % antes da regulação e 6,20 % após, sendo que a maior variância situa-se entre os pontos 13 e 30. Também não se verificam movimentos sazonais (Figura 1).

Figura 1 – Número total de internações hospitalares efetuadas em Vitória da Conquista – Bahia, janeiro de 1995 a junho de 2002.



Nota: as linhas horizontais azuis correspondem às médias em cada período.

AIH – Autorização de internação hospitalar.

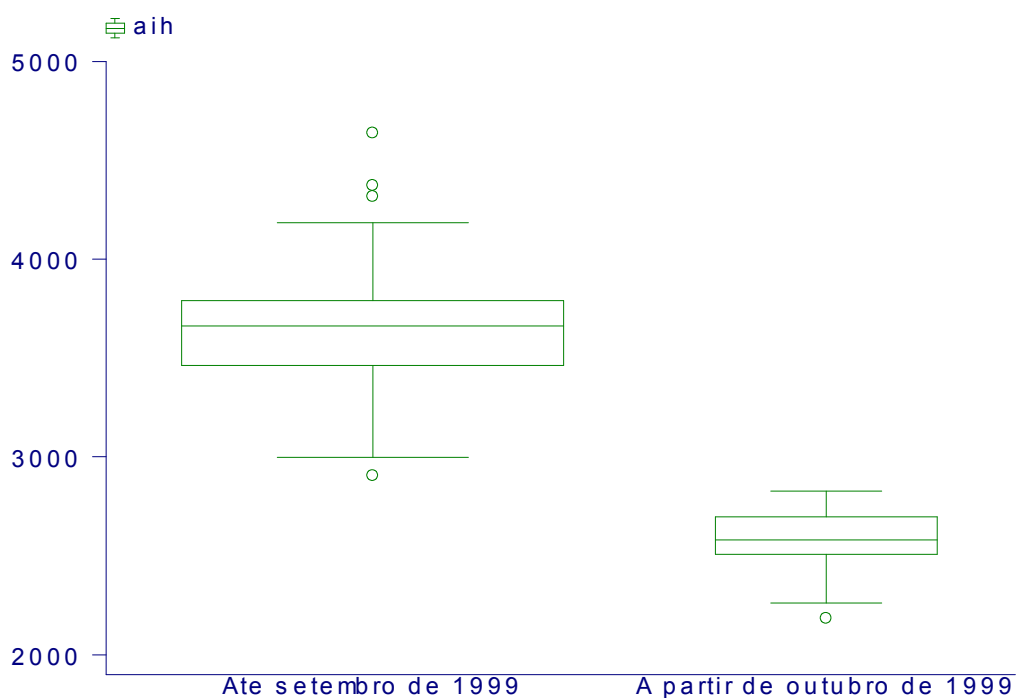
Tempo – O mês, janeiro de 1995 corresponde ao número 1 na série.

Tabela 2 - Estatísticas descritivas do número total de interações mensais efetuadas em Vitória da Conquista – Bahia, no período de janeiro de 1995 a junho de 2002, antes e após a regulação.

Período	N	Media	DP	Min	Q1	Md	Q3	Max
Até setembro de 1999	57	3648,35	325,67	2905	3464	3662	3791	4637
A partir de outubro de 1999	33	2580,58	160,11	2185	2508	2582	2696	2829

Nota: DP – Desvio-padrão; Min – valor mínimo; Q1 – 1º quartil; Md – mediana; Q3 – terceiro quartil e Max – valor máximo.

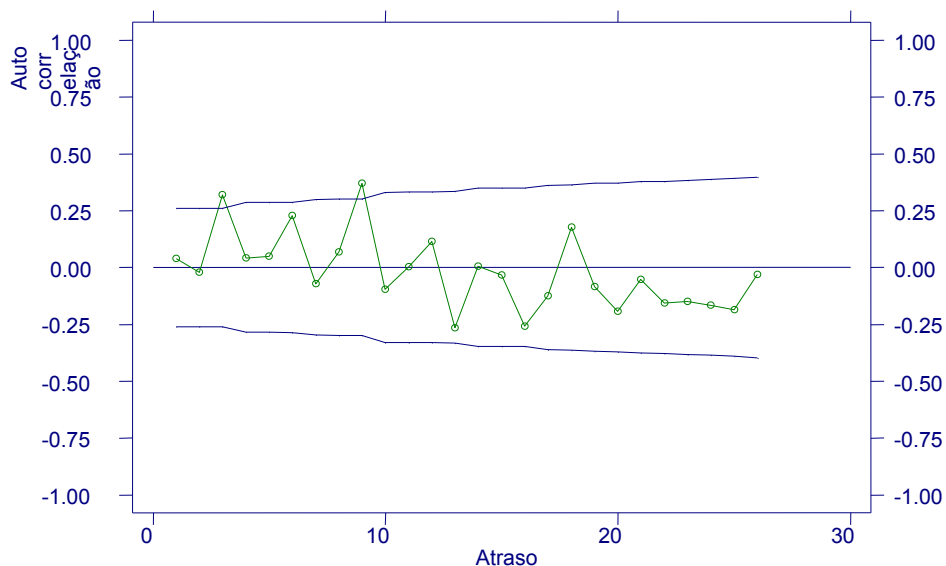
Figura 2 – Distribuição do número total de interações mensais efetuadas em Vitória da Conquista – Bahia, no período de janeiro de 1995 a junho de 2002, antes e após a regulação.



De acordo com a Figura 3, os valores das correlações (auto e parcial) antes da intervenção são relativamente pequenos com algumas poucas exceções (estatística Q significativa), e os mesmos, na sequência dos *lags* (atrasos) sugerem haver estacionariedade na série. A ausência de altas correlações no início do correlograma, indica a possibilidade de tratar-se de uma série do tipo ruído branco (variação aleatória), sem termo autorregressivo. O teste cumulativo de

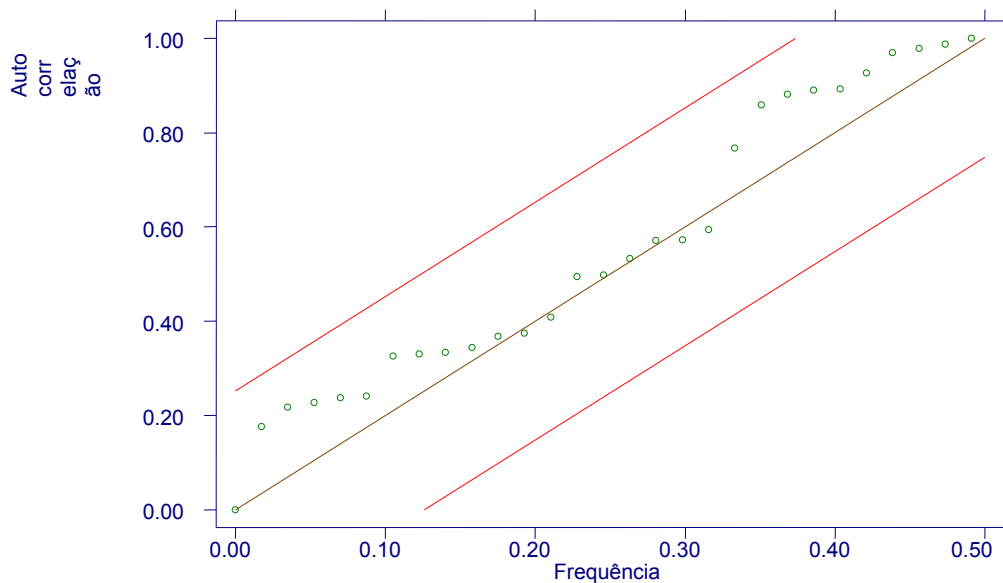
bartlett aceita a hipótese nula para ruído branco ($p = 0,4780$). O teste de Dickey-Fuller ($p = 0,0056$) rejeita a hipótese de não estacionariedade da série, confirmado pelo teste de Phillips-Perron ($p = 0,0172$) (MARGARIDO et al., 2001).

Figura 3 – Correlograma da série histórica de Vitória da Conquista.



Nota: Bandas de confiança de Bartlett a 95% .

Figura 4 – Periodograma cumulativo para teste de ruído branco.



Nota: Teste de Bartlett (B) para ruído branco = 0,84 Prob > B = 0,4780

Modelo tentativo proposto

Apesar de haver uma não estacionariedade na variância ao longo do tempo, verifica-se entretanto, que antes e após a intervenção (ponto 58) as variâncias são homogêneas. Nesse caso, como o objetivo é avaliar o impacto da intervenção ocorrido após o ponto 58 e ainda considerando que a série antes do ponto de intervenção é uma série do tipo ruído branco, então, optou-se por estimar o parâmetro de intervenção a partir do modelo ARIMA(0,0,0) pré-impacto - ruído branco (conforme diagnóstico anterior). Para conseguir tal intento, adicionou-se o termo de intervenção (série composta por uma variável indicadora assumindo valores 0 antes do impacto e valores 1 a partir do impacto), caracterizado como um efeito do impacto de natureza abrupta e permanente (COOK et al., 1979). Portanto, estimou-se, através do método de Huber/White/Sandwich (estimador robusto da variância) (MCDOWELL, 2002), dada uma possível heterocedasticidade dos resíduos e obteve-se um coeficiente para o parâmetro de intervenção igual a $-1067,775$ (erro-padrão = 51,09) estatisticamente significativa ($z = -20,90$; $p < 0,0001$) e uma variância residual de 274,1 (erro-padrão = 28,9) (Tabela 3).

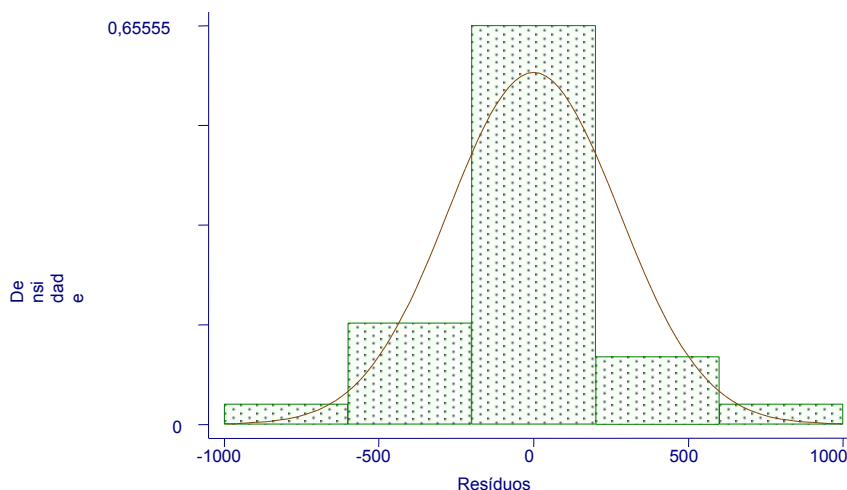
Após o diagnóstico do atendimento aos pressupostos do modelo, através da análise de resíduos pelo teste de normalidade conjunta dos erros (teste de *Skewness/Kurtosis*) ($p = 0,0210$) (Figura 5) - pode-se concluir que a regulação ocorrida em outubro de 1999 em Vitória da Conquista produziu um impacto de natureza abrupta e permanente e estatisticamente significativa na redução do número de internações totais, da ordem de 1068 internações, correspondendo a uma redução da ordem de 29,27 %.

Tabela 3 – Modelo ARIMA proposto e coeficientes estimados.

Parâmetros	Coefficiente.	Erro-padrão	z	Valor de p
Intervenção	-1067,775	51,09164	-20.90	< 0,0001
Constante	3648,351	42,99572	84.85	< 0,0001
resíduo	274,0594	28.93682	9.47	< 0,0001

Nota: 90 observações; Wald $\chi^2(1) = 436,78$; Prob > $\chi^2 < 0,00001$; Log likelihood = -632,9055.

Figura 5 – Histograma dos resíduos da serie logaritimizada do número de internações.



Mortalidade de menores de cinco anos devida à diarreia.

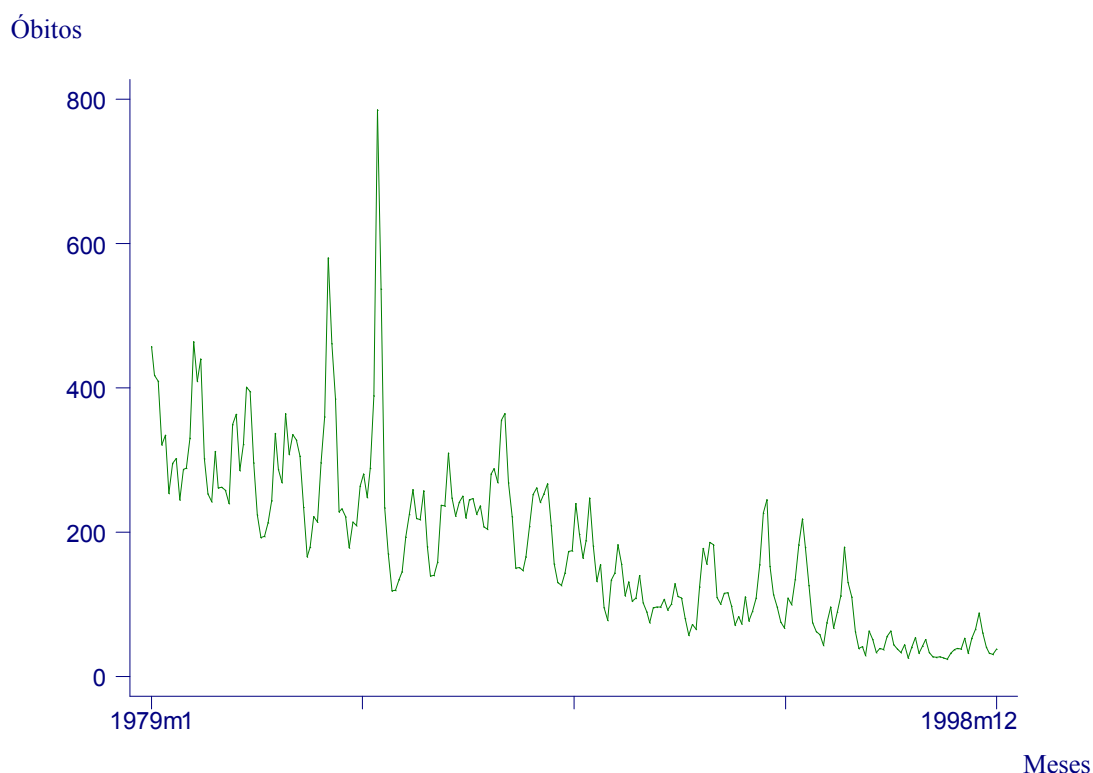
Na Bahia, de janeiro de 1978 a dezembro de 1998, registraram-se 1.022.767 óbitos, dos quais 232.209 (22,7 %) foram de menores de cinco anos. Desses, 43.379 óbitos (18,7 %) se deveram à diarreia inespecífica.

Descrição da série

A série histórica do número absoluto de óbitos demonstra à inspeção visual (Figura 6) um aparente declínio uniforme e suave desde 1979, salvo nos meses de março (580), maio (785) e junho (537) do ano de 1984 cujos valores ultrapassaram os valores habituais: em média 181 óbitos mensais. Verifica-se que devido a esses valores fora do comum a variabilidade em torno da média (181 óbitos mensais) é muito expressiva ($CV = 65,27\%$). Até dezembro de 1988, metade da série, a média era de 267 óbitos mensais e um coeficiente de variação de

37,26 %, enquanto que a partir dali a média caiu para 94 óbitos por mês, com um coeficiente de variação de 57,9 %. Desse modo, pode-se afirmar que a série é não estacionária na sua média (tendência secular) e na sua variância ao longo do tempo.

Figura 6 – Número de óbitos mensais de menores de cinco anos devidos à diarreia, Bahia, 1979 a 1998.



Após transformação logarítmica para estabilizar a variância, verifica-se de acordo com o padrão do correlograma da série (Figuras 7 e 8) a sugestão de sua não estacionariedade temporal através da função de autocorrelação que decresce suavemente e confirmada pela estatística Q ($p < 0,0001$) e pelo periodograma acumulado para séries com ruído branco (Figura 9), necessitando desse modo de diferenciação para torná-la estacionária.

Figura 7 – Autocorrelação do número de órbitas.

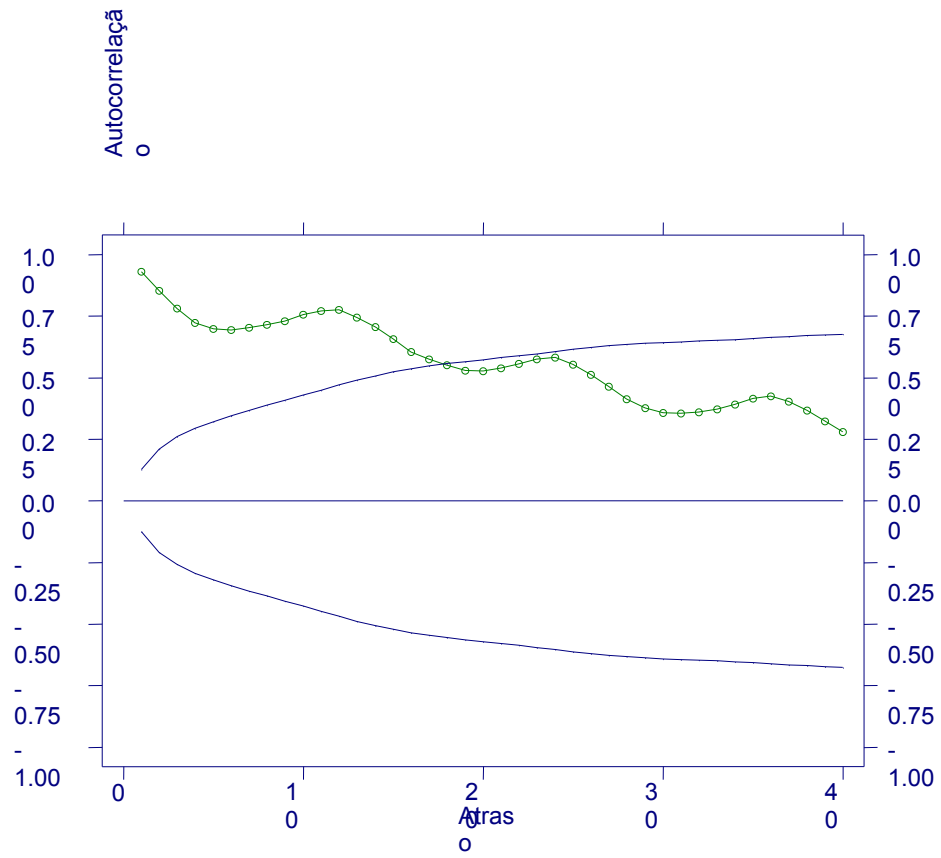


Figura 8 – Autocorrelação parcial do número de órbitas.

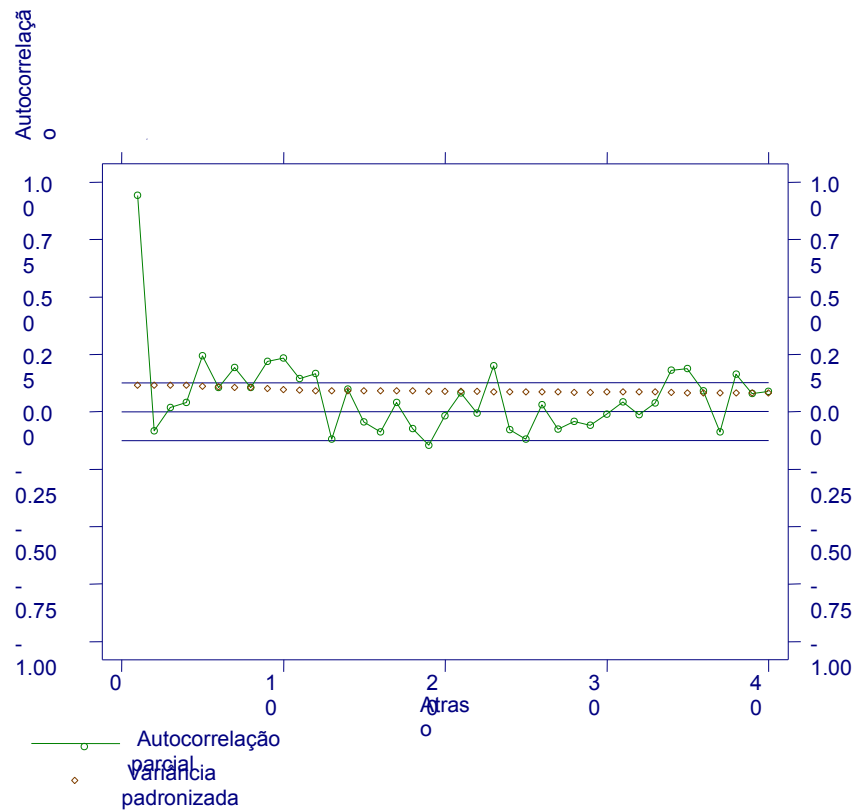
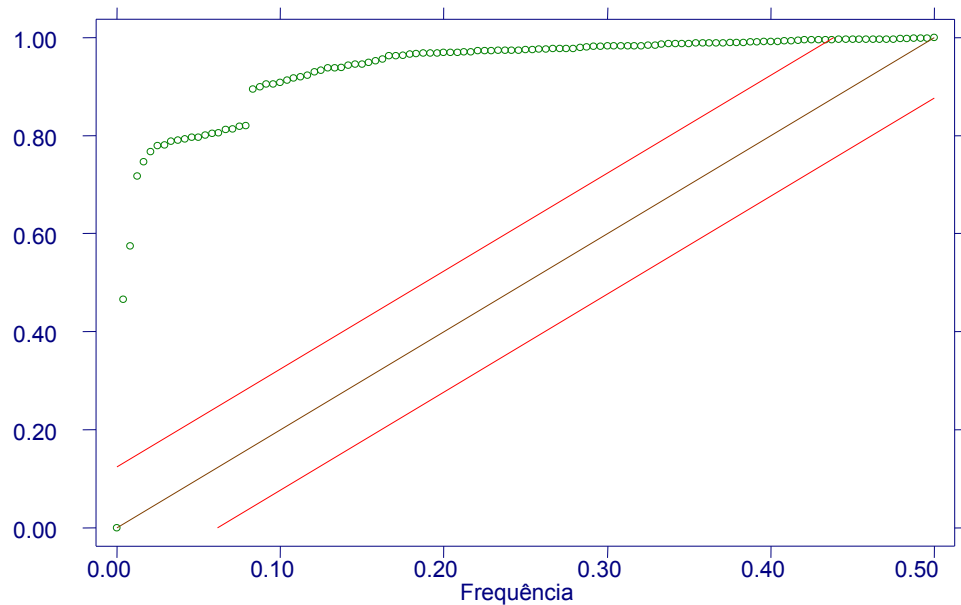


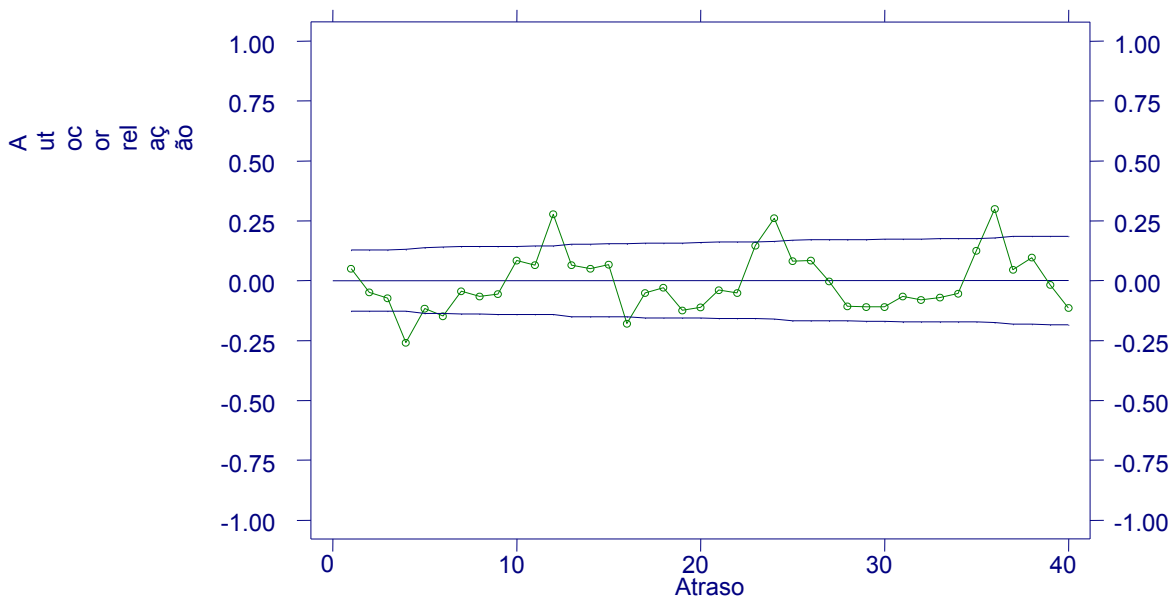
Figura 9 – Periodograma do número de órbitas.



Nota: Teste de Bartlett = 8.02 Prob > B = 0.0000

Na Figura 10, pode-se observar picos (*spikes*) significantes e correlações persistentemente altas nos *lags* 12, 24 e 36 na função de autocorrelação, assim como no *lag* 4, o que sugere um modelo ARIMA tentativo que inclua termos autorregressivos sazonais.

Figura 10 – Autocorrelação do número logaritimizado de óbitos.



Modelo tentativo proposto

Observa-se que após a estimação do modelo $ARIMA(0,1,1)(1,0,1)_{12}$, que inclui uma diferenciação serial, um termo de média móvel serial complementar, um termo autorregressivo sazonal e um termo de média móvel sazonal, os parâmetros respectivos são - 0,15, 0,91 e -0,73 (todos significantes), e um resíduo de 0,23. Portanto, a série possui um forte componente autorregressivo positivo a cada doze meses o que significa que a variação do número de óbitos que ocorre a cada doze meses é fortemente explicada pelo número de óbitos que ocorreram doze períodos anteriores. A equação que define esse modelo misto é:

$$\text{Número de óbitos previstos}_t = 0,91 Y_{t-12} - 0,15 a_{t-4} - 0,73 a_{t-12} + 0,23 a_t$$

Tabela 4 – Modelo ARIMA proposto e coeficientes estimados*.

Componentes		Coefficiente	Erro-padrão	z	Valor de p
Autorregressão sazonal	L12	0,9118008	0640652	14,23	0,0001
	Média móvel				
	L4	-0,1543592	0,0465881	-3,31	0,001

L12	-0,7298182	0,094913	-7,69	0,0001
/resíduo	0,2302456	0,0103969	22,15	0,0001

Nota: 239 observações; Wald $\chi^2(3) = 407,01$; Prob > $\chi^2 < 0,00001$; Log likelihood = 9,482491.

* Número logaritimizado de órbitos e diferenciados uma vez.

O correlograma dos resíduos (Figura 11) do modelo tentativo proposto apresenta uma série tipo *ruído branco*, com variações devidas ao acaso. Os resíduos possuem uma distribuição aproximadamente normal ($p = 0,8391$) (Figuras 12 e 13), com média aproximadamente zero (-0,009) e variância aproximadamente constante (0,054). Além disso, o diagrama de dispersão dos valores preditos versus os resíduos não apresenta nenhuma tendência, sugerindo a distribuição aleatória dos resíduos (Figura 14). Não existem picos significantes de acordo com a estatística Q ($p = 0,4871$).

Figura 11- Função de autocorrelação dos resíduos do modelo tentativo.

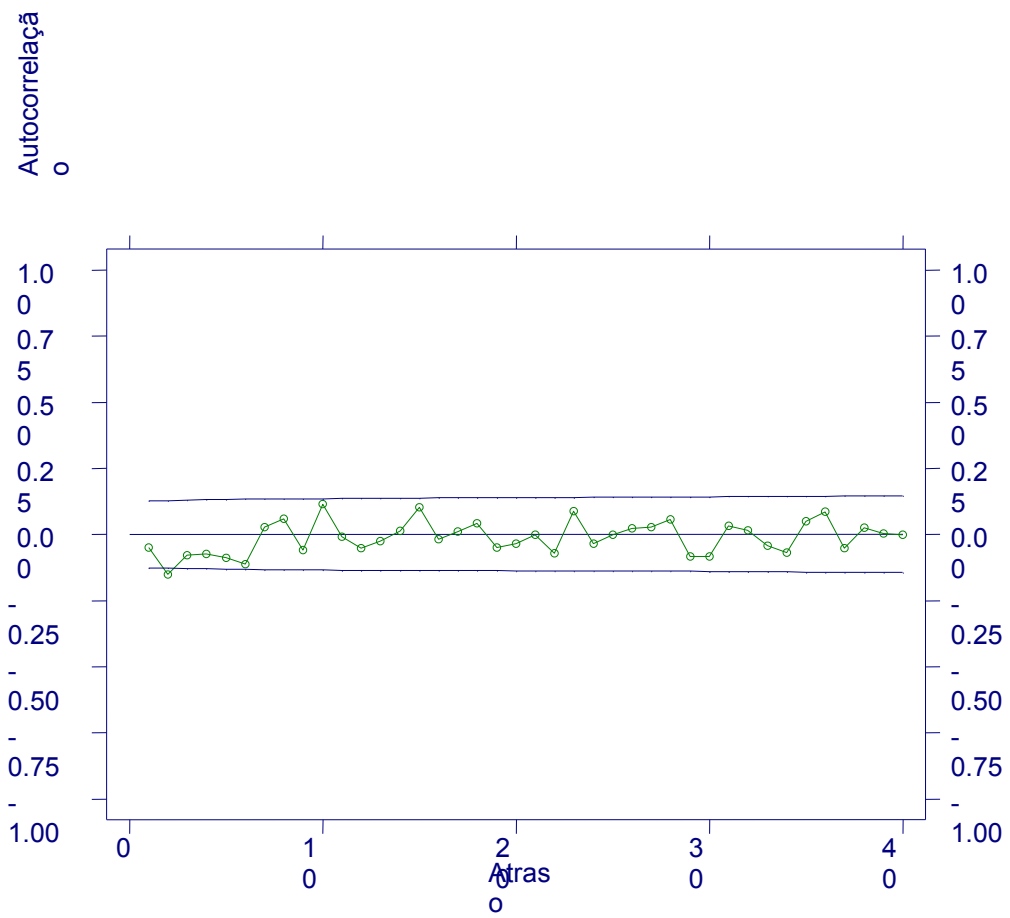
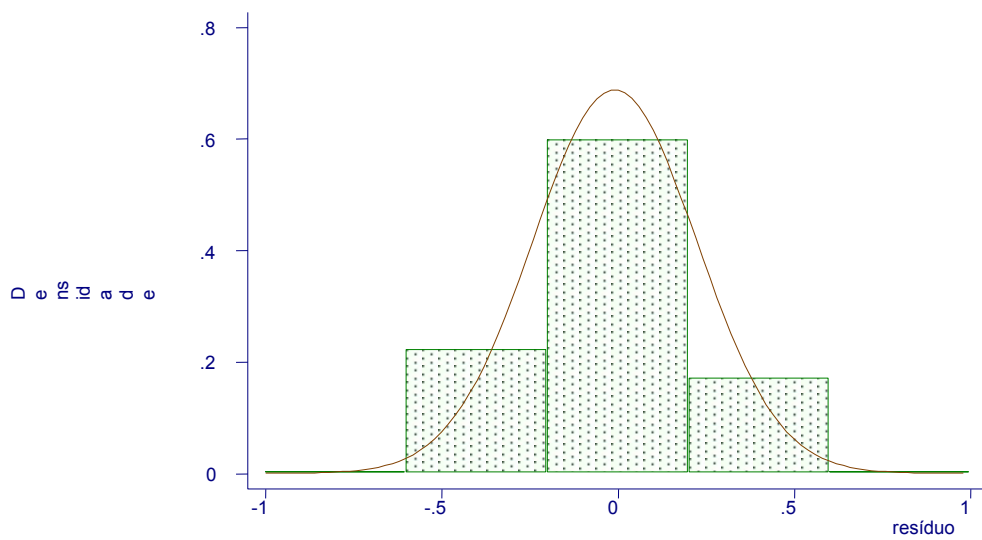
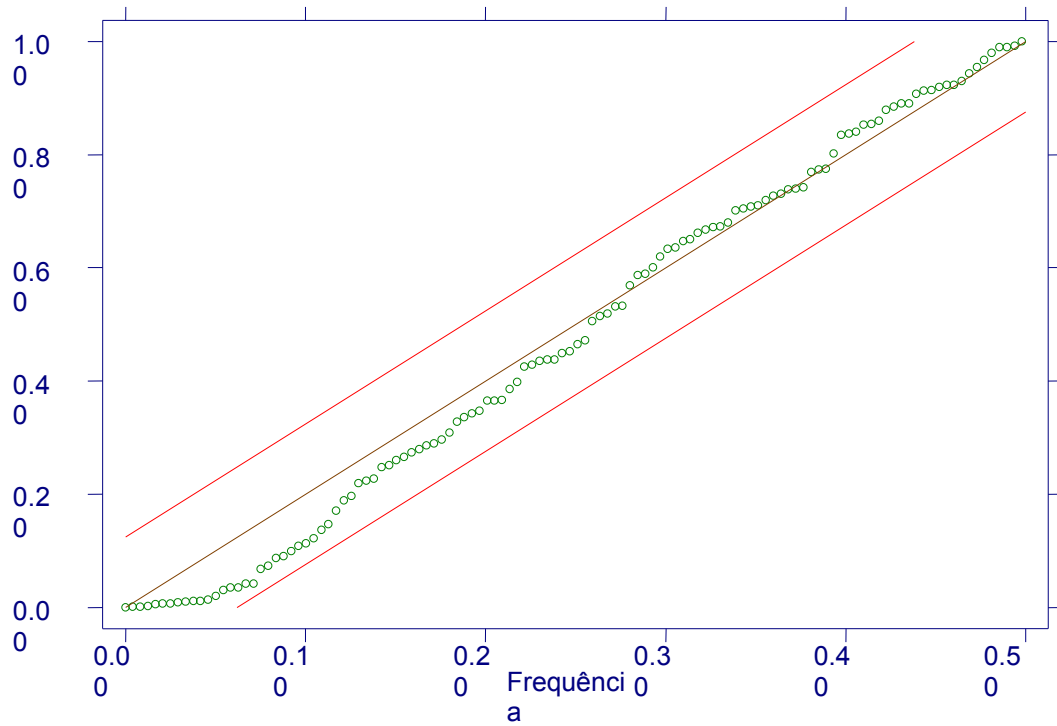


Figura 12 - Histograma dos resíduos do modelo tentativo.



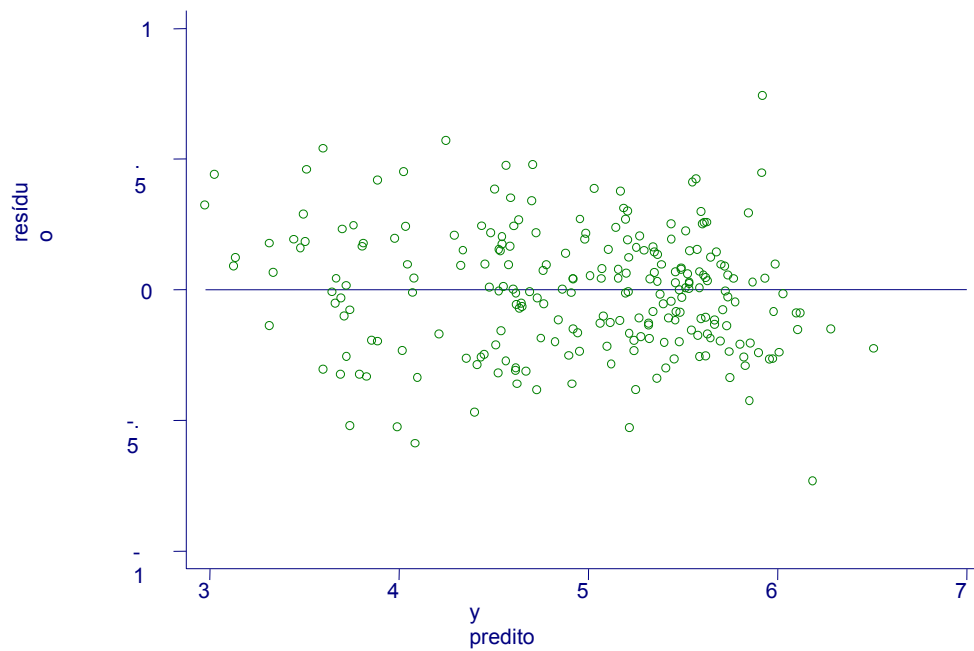
Nota: $p = 0,83906$

Figura 13 – Periodograma cumulativo para o teste de aleatoriedade (ruído branco).



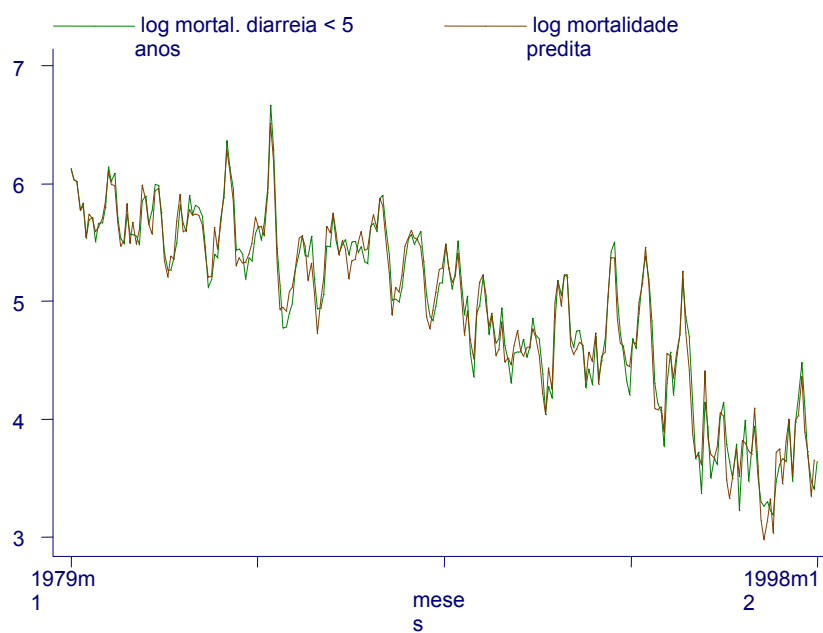
Nota: Teste de Bartlett = 1.10 Prob > B = 0.1772

Figura 14 – Distribuição dos resíduos versus predito a partir modelo proposto.



Finalmente, a Figura 15 apresenta a série original logaritimizada de óbitos de menores de cinco anos devidos à diarreia e os valores preditos pelo modelo. Percebe-se a quase superposição das séries, o que revela a bondade do modelo proposto.

Figura 15 – Número de óbitos observados e preditos pelo modelo ARIMA proposto de janeiro de 1979 a dezembro de 1998 na Bahia.



Discussão e conclusões

Os dados de interesse para a Epidemiologia, quando se trata de quantidades, são construídos a partir de contagens com referências espaciais, ou de grupos específicos de indivíduos e frequentemente sob a forma de séries ditas históricas ou temporais. Tais séries são delineadas compondo agregados ecológicos, compartimentados sequencialmente e contigüamente sem perder a noção de continuidade, segundo unidades temporo-espaciais ou ecológicas.

Entretanto, as séries temporais epidemiológicas, sejam elas de números absolutos de doentes ou casos de uma dada doença, sejam proporções, ou ainda coeficientes, é possível que as suas medições apresentem graus diversos de autocorrelação. Essa medida de dependência de natureza probabilística dos valores de uma série epidemiológica cria uma situação peculiar que constringe as técnicas de análise usualmente empregadas, particularmente quando o propósito é a análise de intervenção (número de internações no município de Vitória da Conquista, incluindo o componente de regulação da oferta e do acesso), ou a predição temporal (mortalidade por diarreia em menores de cinco anos na Bahia).

As séries aqui estudadas foram projetadas em unidade de informação e análise mensal. Reúnem aspectos distintos, porém fazem parte do campo de investigação e de atuação da epidemiologia, seja na sua vertente avaliativa de interesse no subsidio às políticas de saúde ou no seu caráter exploratório do entendimento dos fenômenos de morbimortalidade de populações humanas.

A mortalidade de crianças, sendo estas consideradas como grupo especial de atenção e de cuidados, reflete de algum modo as condições de vida de uma população. Estudos atuais demonstram um decréscimo na mortalidade infantil no Brasil (PAIM et al., 1993). Por outro lado, a ocorrência da diarreia está vinculada a fatores sociais, econômicos, culturais, nutricionais e ambientais, que atinge preferencialmente crianças menores de cinco anos (SILVA et al., 1988). Segundo a OMS, nos países em desenvolvimento, a diarreia é uma das cinco principais causas (pneumonia, diarreia, malária, desnutrição e sarampo) de mortalidade de crianças, chegando a matar cerca de dois milhões de crianças por ano. Entretanto, cerca de 90 % das crianças que morrem, em consequência da diarreia, podem ser salvas através de um adequado manejo. Em 1993, um relatório do Banco Mundial descreve que o manejo integrado das doenças da infância provavelmente possui o maior impacto na redução da carga de doença global (OMS, 2002).

A mortalidade proporcional de menores de cinco anos na Bahia, em 1982, era da ordem de 36,2 % (SILVA et al., 1988), quase duas vezes maior do que a mortalidade proporcional (18,7 %) acumulada entre 1978 e 1998, deixando clara a redução da mortalidade proporcional ao longo do tempo.

Os estudos sistemáticos de séries históricas de mortalidade de crianças, devido a sua importância, através de bases de dados confiáveis e empregando a técnica proposta por Box e Jenkins, possibilita não apenas um diagnóstico epidemiológico mais adequado da sua evolução temporal, como também a avaliação estatística mais acurada de possíveis intervenções efetuadas e sua consequente predição estimada quantitativamente, assim como permite a inclusão dos resultados de sua análise no raciocínio de modelos explicativos e causais sobre esse fenômeno. Portanto, na série de mortalidade estudada nessa tese, os parâmetros estimados (resultados obtidos) devem ser incorporados para aprimorar qualquer uma dessas alternativas analíticas citadas.

A observação da série de mortalidade apresentada, “ao passar dos olhos”, não sugere a existência de fenômenos sazonais, e chama a atenção apenas a tendência secular decrescente e os pontos extremos no início da mesma. Entretanto, após a aplicação da técnica de modelagem ARIMA, verificou-se a existência de um forte componente sazonal anual da ordem de 91 %, demonstrando assim, a importância do método em desvendar aspectos estruturais mascarados por outros movimentos mais visíveis.

Por outro lado, ainda como consequência do vetor resultante dos modelos assistenciais existentes no Brasil, do paradigma científico fragmentador, do modelo econômico baseado no ganho de capital, na tecnologização crescente, entre outros aspectos, permanece uma ênfase e um incentivo na hospitalização dos indivíduos com problemas relacionados à falta de saúde, fomentando uma lógica de obtenção de lucro, que no caso específico da diarreia, resulta em um número de internações desnecessárias de crianças, a um custo crescente.

Em 18 anos, de 1984 a junho de 2002, foram realizadas 216 milhões de internações no Brasil e 14 milhões na Bahia, pelas quais pagou-se um valor de aproximadamente 87 trilhões e cerca de 4,6 trilhões de reais, respectivamente. Enquanto que em 3 anos e meio (entre julho de 1994 e dezembro de 1998) foram internadas 662 mil crianças com diagnóstico de infecção intestinal mal definida, correspondendo a 4,67 % do total de internações na Bahia, e por elas pagou-se o montante de cerca de 224 bilhões de reais, correspondendo a 4,85 % do valor total pago no período. Todavia, verificou-se naquele período, um declínio aproximado na população estudada de Um milhão e trezentas mil para Um milhão e duzentas mil crianças, sendo que a taxa de internação variou de 8,13 a 33,89 por dez mil crianças.

Desse modo, políticas racionalizadoras são fundamentais para a implementação de mudanças que reorientem as ações de saúde dentro dos princípios elaborados para o Sistema Único de Saúde. Em Vitória da Conquista o impacto da intervenção (regulação), um ato de natureza fundamentalmente política, produziu uma redução observada de 1816 internações com uma economia de 412 mil reais em apenas um mês. De acordo com o modelo ARIMA proposto, a estimativa probabilística de redução no número de internações totais foi da ordem de 30 % ($p < 0,0001$), quase um terço das internações, em média, levando-se em conta a série histórica de internações no município, a partir de janeiro de 1995. Entretanto, se fosse calculada a redução das internações a partir de um estudo do tipo antes-depois, usando como meses de referência setembro de 1999 e outubro de 1999, aquela teria sido de 45 %. Ou seja, uma superestimação da ordem de 15 % na proporção da redução das internações. Por conseguinte, a redução esperada, ou o impacto da regulação sobre o número de internações totais, nas condições de Vitória da Conquista é de 30 %. Trata-se de uma estimativa menos enviesada por basear-se em um modelo probabilístico, obtida a partir de uma metodologia que considera a dinamicidade da série temporal (BOX et al., 1976).

Entretanto, esse é um exemplo de que quando a série temporal antes da intervenção é estacionária e não apresenta autocorrelação serial ou sazonal e o efeito do impacto é de natureza abrupta e permanente sobre uma série também estacionária após a intervenção, as estimativas dos parâmetros de intervenção se aproximam da diferença simples entre as médias de antes e após o ponto de interrupção. Caso contrário, os valores estimados tendem a ser bastante diferentes (PANKRATZ, 1983). Mas, para a decisão da utilização do cálculo do impacto baseando-se nas médias, antes se faz necessária a avaliação da estacionariedade das séries antes e após o impacto, procedimento esse obrigatório da modelagem ARIMA.

Como se pode constatar, a adoção da metodologia de Box e Jenkins em contextos epidemiológicos torna-se uma alternativa importante, uma vez que além de produzir estimativas menos enviesadas dos parâmetros estruturais das séries históricas, descrever parâmetros que não são produzidos por outras técnicas e que são úteis para uma maior compreensão dos fenômenos epidemiológicos, ainda permite efetuar previsões com maior precisão.

Referências bibliográficas

ALMEIDA, R.M.V.R.;NOBRE, F.F.;PANERAI, R.B. Caracterização dinâmica de séries históricas de determinantes de saúde. *Revista Brasileira de Engenharia*, 5(1): p. 33-43. 1987.

BOX, G.E.P.;JENKINS, G.M. *Time series analysis : forecasting and control*. Rev. ed. San Francisco: Holden-Day, 1976. xxi, 575 p.

BRASIL. *Estimativas de impacto da vinculação constitucional de recursos para a saúde : Emenda Constitucional n.º 29/2000*. Brasília: Ministério da Saúde-Secretaria de Gestão de Investimentos em Saúde, 2001.36 p.

_____. *DATASUS*. Ministério da Saúde. Departamento de Informática do SUS, DATASUS. Disponível em. <http://www.datasus.gov.br>. Acesso: 2002

CARMO, E.H.;SANTOS, C.A.T.;BARRETO, M.L. *Characteristics of schistosomiasis mansoni cases admitted to hospitals in the state of Bahia, 1994*. In: Simpósio Internacional de Esquistossomose, 5ª Reunião Nacional de Esquistossomose. Resumos. Salvador (Brasil): Fundação Oswaldo Cruz, 1995.

COOK, T.D.;CAMPBELL, D.T. *Quasi-experimentation*. U.S.A.: Houghton Mifflin Company, 1979.

KLEINBAUM, D.G.;KUPPER, L.L.;MORGENSTERN, H. *Epidemiologic research*. Belmont, Calif.: Lifetime Learning Publications, 1982. 529 p.

MARGARIDO, M.A.;ANEFALOS, L.C. Testes de co-integração utilizando o SAS: teoria e aplicação. *Informações Econômicas*, 31(1): p. 21. 2001.

MCDOWELL, A. From the help desk: transfer functions. *The Stata Journal*, 2(1): p. 71-85. 2002.

OMS. *Reducing mortality from major killers of children*. OMS. Disponível em. <http://www.who.int/inf-fs/en/fact178.html>. Acesso: 2002

OTERO, U.B. *Estudo da mortalidade por desnutrição em idosos na Região Sudeste do Brasil, 1980-1997*. Rio de Janeiro: ENSP/FIOCRUZ; 2001.

PAES, N.A.;ALBUQUERQUE, M.E.E. Avaliação da qualidade dos dados populacionais e cobertura dos registros de óbitos para as regiões brasileiras. *Rev. Saúde Públ*, 33(1): p. 33-43. 1999.

PAIM, J.S.;COSTA, M.D.C.N. Decline and unevenness of infant mortality in Salvador, Brazil, 1980-1988. *Bulletin of the Pan American Health Organization*, 27(1): p. 1-14. 1993.

PANKRATZ, A. *Forecasting with univariate Box-Jenkins models*. USA: John and Wiley & Sons, 1983. 562 p.

PEREIRA, M.G. *Epidemiologia*. Rio de Janeiro: Guanabara Koogan S.A., 1995. 583 p.

SILVA, L.R., et al. *diarreia aguda na criança*. Rio de Janeiro: MEDSi, 1988. 158 p.

VERAS, C.M.T.;MARTINS, M.S.A. Confiabilidade dos dados nos formulários de Autorização de Internação Hospitalar (AIH). *Cad Saúde Públ*, 10 (3):: p. 339-55. 1994.

VITÓRIA DA CONQUISTA. *Vitória da Conquista*. Prefeitura Municipal. 2002:43 p.

ARTIGO 3

Aplicação do método de modelagem de Box e Jenkins, em séries temporais epidemiológicas distintas.

Parte II

Internações por diarreia e valores pagos: uma análise bivariada de séries temporais.

Artigo 3 - Aplicação do método de modelagem de Box e Jenkins, em séries temporais epidemiológicas distintas. Parte II - Internações por diarreia e valores pagos: uma análise bivariada de séries temporais.

Resumo

É de interesse no campo da Epidemiologia a comparação de séries temporais, com o propósito de se verificar a possível relação entre elas. Entretanto, tais comparações limitam-se a sobreposições figurativas que dependem exclusivamente da subjetividade do analista. No campo da Econometria, a análise multivariada de séries históricas é frequente e incorpora, através de funções de transferência de Box e Jenkins para a análise covariada de séries temporais, critérios objetivos no seu estudo.

Com o propósito de se demonstrar a aplicação da técnica de Box e Jenkins denominada modelos autorregressivos integrado com médias móveis (ARIMA) em um contexto epidemiológico específico e o seu potencial na averiguação do grau de interdependência cruzada entre séries históricas, foram selecionadas, a série da taxa de internação hospitalar por 1.000 hab e a série dos valores pagos em Reais correspondentes, entre julho de 1994 e dezembro de 1998. Ajustou-se um modelo ARIMA(0,1,0) para a série de internações, a qual foi filtrada em relação à série de valores pagos diferenciada, obtendo-se a equação de predição $v \hat{p}_t = 0,0074794 + 1,52 \cdot txi_t + 0,0645743$, onde $v \hat{p}_t$ é o valor pago predito e txi_t , a taxa de internações, ambos no tempo. A correlação cruzada entre as séries foi significativa no atraso (*lag*) 0 e foi da ordem de 99,24 %, representando uma concordância ou dependência quase perfeita, significando haver uma correspondência efetivamente mensal no pagamento das internações hospitalares.

Desse modo, fica patente a aplicabilidade da proposição metodológica de Box e Jenkins em contextos epidemiológicos que visem analisar mais de uma série histórica concomitantemente e obter critérios objetivos no julgamento de suas interdependências.

Descritores: fenômenos de movimento complexo; séries temporais; ARIMA; Epidemiologia

Article 3 – Use of the Box and Jenkins model method in different epidemiological time series. Part II – Hospital admissions due to diarrhea and costs incurred: a bivariate time series analysis.

Abstract

In the field of Epidemiology, there is great interest in comparing temporal series in order to establish a possible relationship among them. Such comparisons are, however, limited to figurative overlays that depend exclusively on the analyst's subjectivity. In the field of Econometry, the multivariate analysis of historical series is common and incorporates objective criteria into its study by means of Box and Jenkins transference functions for the covariate analysis of time series.

In order to demonstrate the application of the Box and Jenkins technique, known as autoregressive integrated moving average (ARIMA) models, in a specific epidemiological context, and its potential in the verification of the degree of crossed interdependence between historical series, the following criteria were selected: the series of hospital admission rates per 1,000 inhabitants and the series of costs incurred in corresponding Brazilian Reais between July 1994 and December 1998. An ARIMA model (0,1,0) was adjusted for the series of hospital admissions, which was filtered in relation to the differentiated series of costs incurred, obtaining the prediction equation $v\hat{p}_t = 0.0074794 + 1.52 \cdot txi_t + 0.0645743$, in which $v\hat{p}_t$ is the predicted cost incurred and txi_t the hospital admission rate, both in time. The crossed correlation between the series was significant in the lag 0 and at around 99.24%, demonstrated an almost perfect agreement or dependence, showing evidence of a clearly monthly correspondence in the costs incurred with hospital admissions.

The applicability of the Box and Jenkins methodological proposition in epidemiological contexts, in which the aim is to analyze more than one concomitant historical series and to obtain objective criteria in the judgment of their interdependencies, is therefore obvious.

Key words: complex motion phenomena, time series, ARIMA, epidemiology

Introdução

É de interesse no campo da Epidemiologia a comparação de séries temporais, com o propósito de se verificar a possível relação entre elas. Embora, a maior perspectiva seja no sentido de elucidar relações de causalidade, a descrição do comportamento simultâneo de séries históricas tem grande importância, até mesmo como um indicativo de possíveis associações, mesmo que não necessariamente causais (ROTHMAN et al., 1998). Entretanto, tais comparações limitam-se a sobreposições figurativas que dependem exclusivamente da subjetividade do analista.

O usual nessa abordagem, como se pode observar nos livros-texto de Epidemiologia, é a realização da análise a partir de uma representação gráfica de linhas simultâneas, correspondentes às séries em um mesmo plano cartesiano, através de processos cognitivos baseados no conhecimento e na experiência do epidemiologista, com o objetivo de reconhecer padrões que sejam reconhecíveis dentro do campo paradigmático.

Entretanto, esse procedimento comparativo de séries de tempo é mais ou menos empregado na dependência da área especializada da Epidemiologia. Por exemplo, é mais comum esse interesse particular entre epidemiologistas que lidam com questões ambientais do que na pesquisa epidemiológica em saúde mental.

Porém, no campo da Economia, particularmente na Econometria⁴ (MATOS, 1997), a situação multivariada de séries históricas é frequente, despertando interesses principalmente de natureza preditiva (MONTORO FILHO, 1988), e para isso utiliza-se de métodos mais avançados de modelagem que produzem resultados com maior objetividade (parâmetros), permitindo que outros analistas possam não apenas validar os resultados, como utilizá-los com maior confiabilidade e conseqüentemente menos viés. Dentre esses métodos, a modelagem ARIMA⁵ com funções de transferência de Box e Jenkins, é o mais empregado (MORETTIN et al., 1986).

4 Método de análise quantitativa a Economia

5 Modelos autorregressivos integrados com médias móveis

Assim, esse artigo objetiva demonstrar a aplicação do método de modelagem de séries temporais dinâmicas, através da proposição de Box e Jenkins, visando demonstrar o seu potencial na averiguação do grau de interdependência cruzada entre séries históricas epidemiológicas, nesse caso, em uma situação bivariada.

Metodologia

Optou-se pela escolha do número de internações por diarreia em menores de cinco anos na Bahia, e os valores pagos dada a importância daquela morbidade como fonte de internação de um grupo de risco de interesse epidemiológico e os altos custos hospitalares pagos por uma condição sabidamente evitável. As séries foram constituídas entre julho de 1994 e dezembro de 1998, mensalmente, e obtidas diretamente do *site* do DATASUS/FUNASA (BRASIL, 2002).

Foram considerados os Códigos Internacionais de Doenças 009 (infecções intestinais mal definidas) e A09 (diarreia e gastroenterite de origem infecciosa presumível) para diarreia, 9ª e 10ª revisões, respectivamente, para a construção da série temporal de internações.

A partir dos dados do censo de 1991 e da contagem populacional de 1996 da população de menores de 5 anos na Bahia fornecidos pelo IBGE (IBGE, 2002) foi construído um algoritmo (MICROSOFT, 2001) (apêndice) para o cálculo das estimativas populacionais entre 1991 e 1998, mensalmente, através do método de crescimento geométrico (SANTOS, 1980).

A taxa de internação foi então criada como sendo uma razão entre o número de internações devidas à diarreia de crianças menores de cinco anos e a população de menores de cinco anos estimada para o mês correspondente e expressa em número de internações por 1.000 habitantes.

A série histórica com 54 meses para os valores pagos pelas internações hospitalares devidas à diarreia em menores de cinco anos foi construída a partir de julho de 1994 (valores disponibilizados em Reais pelo DATASUS/FUNASA) até dezembro de 1998. Nessa série

foram imputados valores médios de meses adjacentes para o mês de agosto de 1995, janeiro e fevereiro de 1998, devido a presença de valores muito abaixo da média.

Para a análise das séries empregou-se a modelagem ARIMA (autorregressivos integrados com média móvel) (BOX et al., 1976; PANKRATZ, 1983), através da qual, foram estimados os parâmetros estruturais, após a verificação da estacionariedade das séries. Para tanto, adotou-se a abordagem em estágios (vide apêndice), estratégia sugerida por Box e Jenkins para a identificação dos modelos.

Particularmente, no procedimento analítico das séries em uma perspectiva multivariada, considerou-se como série de entrada (ou série independente) a taxa mensal de internações e a série de saída (ou série dependente) os valores pagos mensalmente em reais.

Como estratégia descritiva inicial, gerou-se um correlograma das séries para se verificar a existência de atraso na correlação entre o número de internações em um mês e o correspondente valor pago. Em seguida, efetuou-se a proposição da função impulso-resposta que permitisse a estimativa da natureza e do grau de correlação entre as duas séries ao longo do tempo.

As séries temporais foram analisadas através de pacote estatístico apropriado (STATACORP, 2001), instalado em computador IBM compatível, contendo 512 Mb de memória de acesso aleatório, processador de 900 Mz e memória permanente de 100 Gb.

Resultados

Foram efetuadas de julho de 1994 a dezembro de 1998 cerca de 150 mil internações (em torno de 21 % do total: 718.970) de crianças menores de cinco anos a um custo total de cerca de R\$ 17 milhões (cerca de 13,0 % do gasto total na faixa etária – R\$ 134 milhões). Em média foram realizadas naqueles 54 meses 2.778 internações por mês (variando entre 1.023 e 4.588) a um custo médio mensal de R\$ 321.224,00, com uma amplitude entre R\$ 510.518,30 em 1994 e R\$ 117.977,00 em 1998 (Figuras 1 e 2).

Verificou-se no período, um declínio na população estudada de 1.369.609 para 1.237.381 crianças. A taxa de internação variou de 8,13 a 33,89 por dez mil crianças (Figura 1).

Figura 1 – Taxa de internação e respectivos valores pagos mensais de menores de cinco anos devidos à diarreia, Bahia, jul. 1994 a dez. 1998.

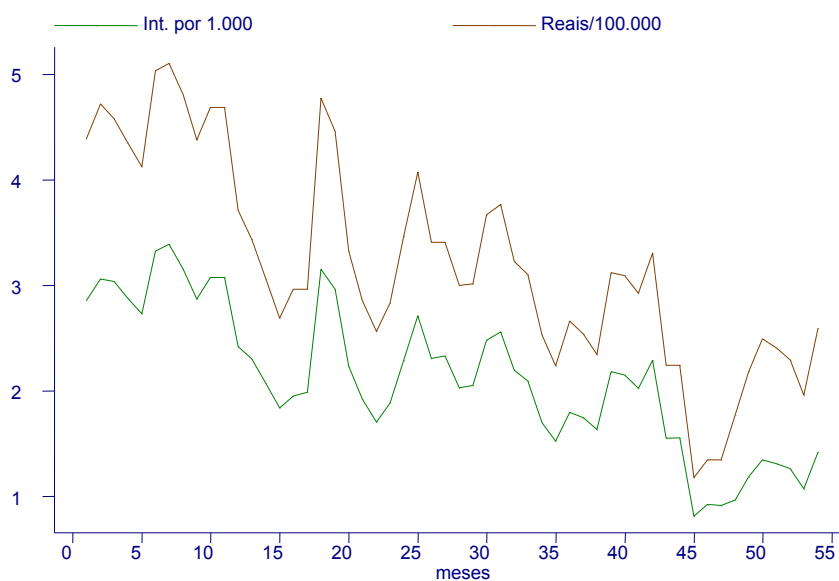
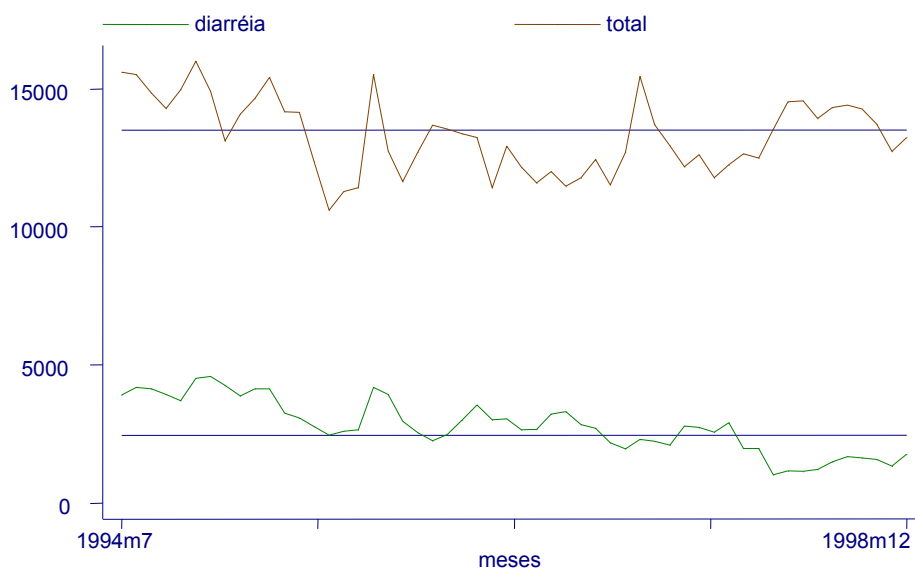


Figura 2 – Número de internações de menores de cinco anos devidas à diarreia e totais, Bahia, jul. 1994 a dez. 1998.



Nota: Séries imputadas. As linhas paralelas azuis representam as médias de internações.

De acordo com a Figura 3, verifica-se um aparente decréscimo dos valores pagos pelas internações de menores de cinco anos devidas à diarreia na Bahia, de julho de 1994 a dezembro de 1998. Enquanto que a Figura 4 sugere uma certa estabilização até o início de 1998, e um aumento a partir daquele momento.

Figura 3 – Valores pagos em reais pelas internações mensais de menores de cinco anos devidas à diarreia, Bahia, jul. 1994 a dez. 1998.

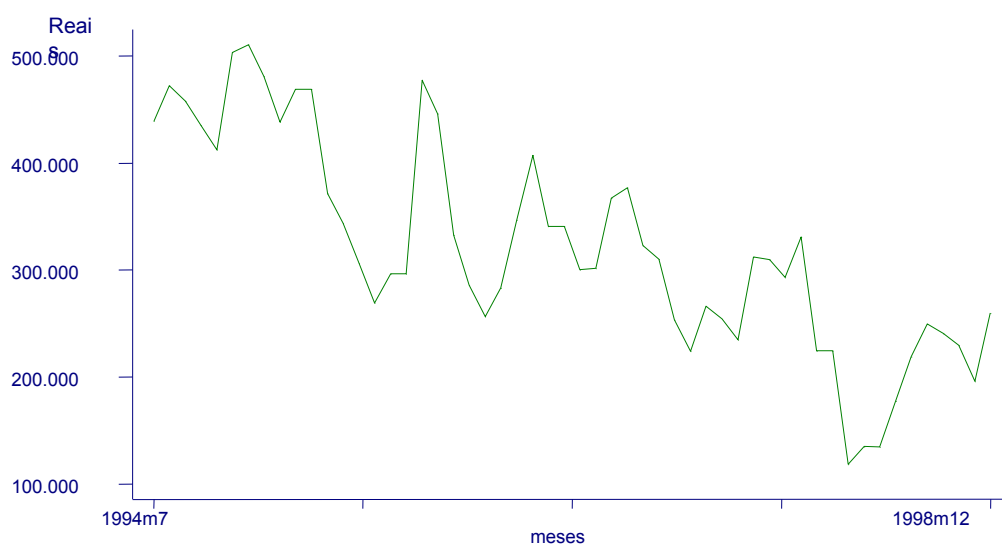
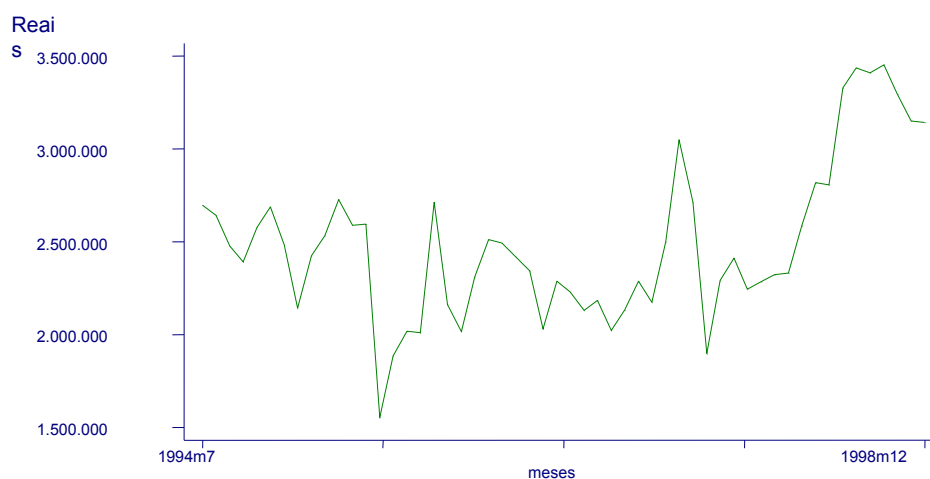


Figura 4 – Valores pagos em reais por todas as internações mensais de menores de cinco anos, Bahia, jul. 1994 a dez. 1998.

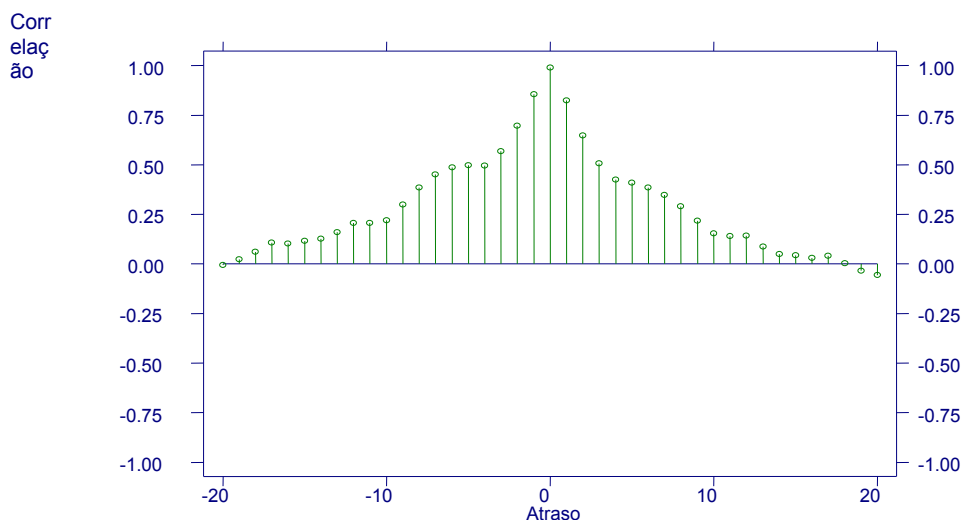


Observa-se no correlograma cruzado⁶ da taxa de internação por diarreia e valores pagos em reais (Figura 5) a existência de um pico diferenciado e marcante no *lag 0*, o que sugere descritivamente a não existência de atraso na correlação entre o número de internações em um mês e o correspondente valor pago. Além disso, verificam-se correlações positivas, o que

⁶ Figura de autocorrelações e correlações entre séries

sugere uma associação direta entre a taxa de internação e os valores pagos. É preciso então, efetuar a proposição da função impulso-resposta que permita a estimativa da natureza e do grau de correlação entre as duas séries ao longo do tempo.

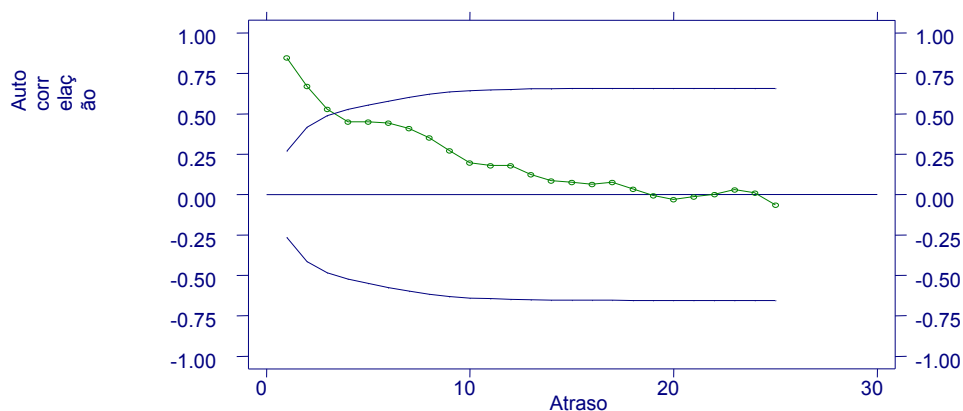
Figura 5 – Correlograma cruzado da taxa de internação por diarreia e dos valores pagos pelas internações mensais de menores de cinco anos devidas à diarreia.



A série dependente (saída): valores pagos pelas internações

Conforme a Figura 6, a função de autocorrelação para a série dependente apresenta um declínio suave e que considerando a estatística Q de Portmanteau ($Q = 154,18$; $p < 0,0001$), indica a necessidade de sua transformação, através de diferenciação, devido a não estacionariedade serial.

Figura 6 – Correlograma da série de valores pagos pelas internações mensais de menores de cinco anos devidas à diarreia.



Após a primeira diferenciação (Figuras 7 e 8), percebe-se ($Q = 24,76$; $p = 0,4197$) que isto foi suficiente para torná-la uma série estacionária. Do mesmo modo, não há evidências de tendências sazonais.

Figura 7 – Correlograma da série diferenciada uma vez dos valores pagos pelas internações mensais de menores de cinco anos devidas à diarreia.

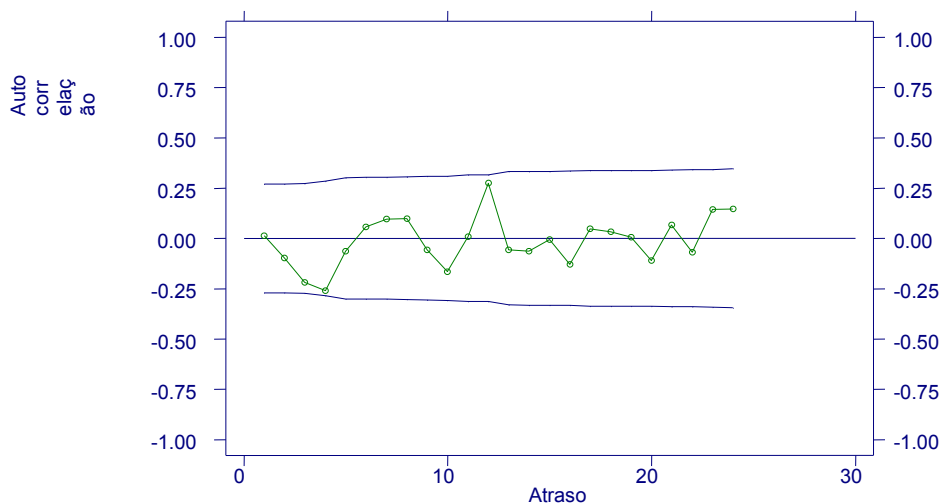
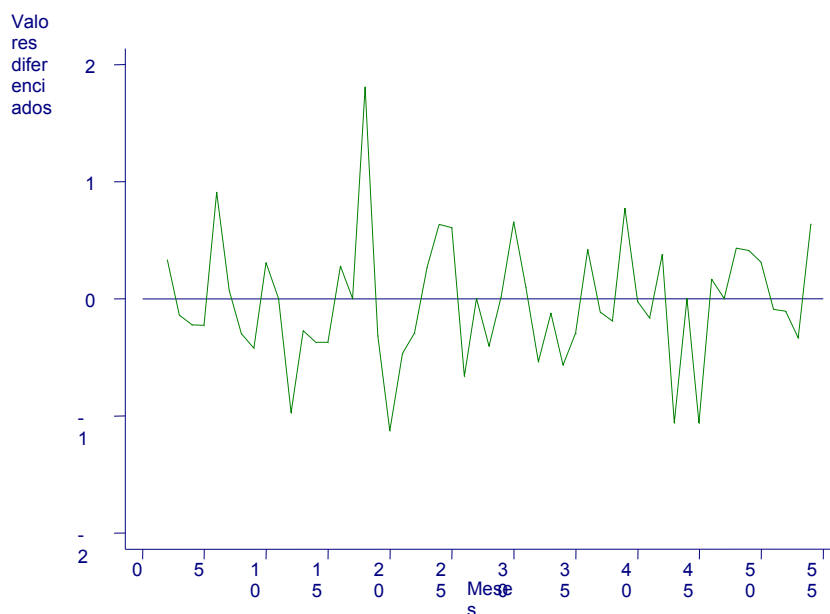


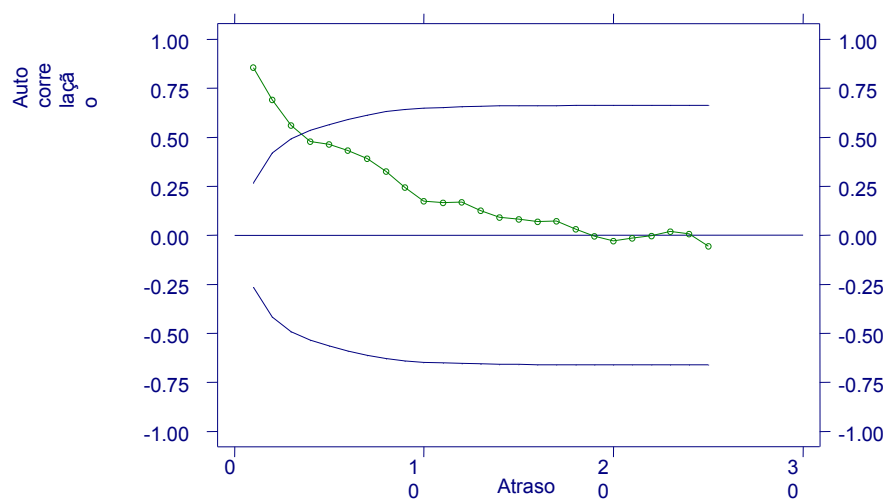
Figura 8 – Série dos valores pagos pelas internações mensais de menores de cinco anos devidos à diarreia, diferenciada uma vez.



A série de entrada: taxa de internações

Pode-se observar na Figura 9 que, de acordo com a função de autocorrelação e da estatística Q (156,75; $p < 0,0001$), trata-se de uma série temporal não estacionária, cuja média varia descendentemente ao longo do tempo, necessitando, portanto de diferenciação para transformá-la em estacionária.

Figura 9 – Correlograma da taxa de internação mensal de menores de cinco anos devidas à diarreia.



Após a primeira diferenciação (Figuras 10 e 11), a série não apresenta evidências de não estacionariedade ($Q = 23,05$; $p = 0,5168$). Como está é a série de entrada, como reforço na averiguação da sua estacionariedade, os testes de Dickey-Fuller e de Phillips-Perron para raiz unitária (MARGARIDO et al., 2001) concordam em serem significantes, rejeitando a hipótese nula de que a série é não estacionária (valor de p aproximado pelo algoritmo de MacKinnon $< 0,0001$).

Figura 10 – Correlograma da série diferenciada uma vez da taxa de internação mensal de menores de cinco anos devidas à diarreia.

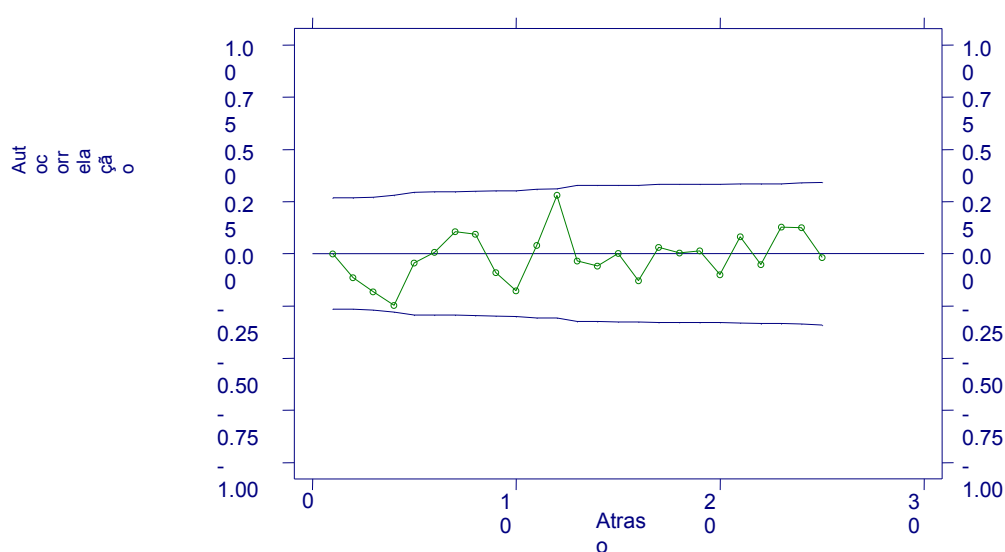
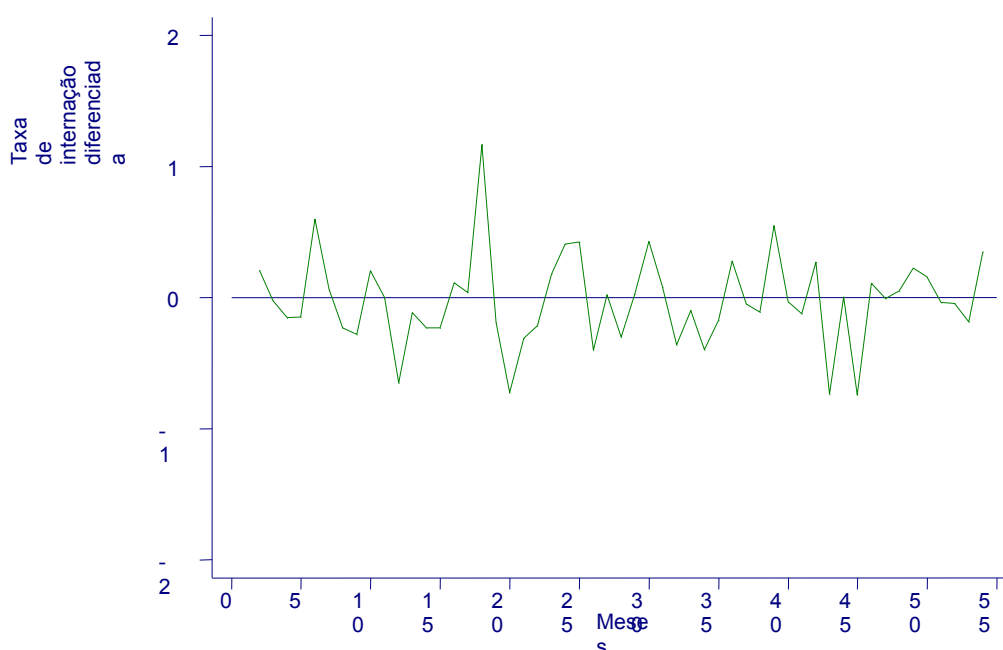


Figura 11 – Taxa de internações diferenciada uma vez.

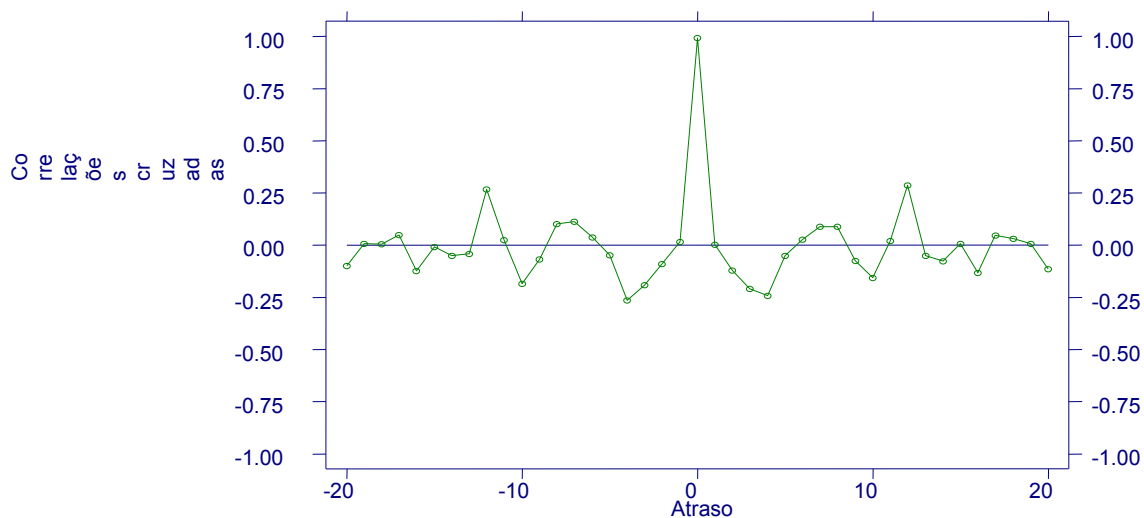


A correlação cruzada das séries

Após a estimativa dos resíduos a partir do modelo tentativo ARIMA(0,1,0) – uma diferenciação serial - sobre a série de entrada (taxa de internação), e a aplicação do modelo estimado (filtro) sobre a série saída (valores pagos), obteve-se o correlograma cruzado dos resíduos (Figura 12). Neste, evidencia-se um único pico marcante (correlação cruzada = 0,9924) no *lag* 0. Esse correlograma será proporcional à função impulso-resposta, de acordo com (MCDOWELL, 2002). Ao se adotar o critério de zero menos dois desvios-padrão como ponto de corte para a identificação de correlações cruzadas significativamente diferentes de zero, apenas a correlação cruzada no *lag* 0 satisfaz o ponto de corte, configurando-se esta, como a estrutura de atraso da série de entrada em relação à de saída. Ou seja, o efeito nos valores pagos deverá ocorrer no mesmo mês da internação, confirmando o que foi obtido anteriormente pelo correlograma descritivo da correlação cruzada entre as séries (Figura 5).

Finalmente, ao realizar-se o modelo ARIMA proposto, incluindo a série *prewhitened* (diferenciada e convertida em *ruído branco*) de entrada e filtrada em relação à série de valores pagos diferenciada, obtêm-se a equação de predição $\hat{v}_t = 0,0074794 + 1,52 \cdot x_{t-1} + 0,0645743$ (Tabela 1), onde \hat{v}_t é o valor pago predito. Nesse caso, a equação reduziu-se a uma equação linear simples porque não houve termos autorregressivos ou de médias móveis. A estatística Q é não significativa ($p = 0,9890$), sugerindo tratar-se de um modelo tipo ruído branco ou contendo apenas variação aleatória no tempo (*White noise*), não correlacionada com o filtro proposto (Figura 13), atendendo a esse pressuposto para correlacionamentos cruzados de séries temporais.

Ao se efetuar predições a partir da equação obtida pelo modelo, como prova final da bondade de ajuste do modelo tentativo, observa-se na Figura 14, uma quase superposição de valores preditos pagos e os valores observados pagos, com uma correlação da ordem de 99,24 %, representando uma concordância quase perfeita entre elas.

Figura 12 – Correlograma cruzado após aplicação de filtro.**Tabela 1 – Modelo ARIMA proposto e coeficientes estimados.**

Filtro sobre a série valores pagos	Coefficiente	Erro-padrão	z	Valor de p
Taxa de internação	1,52358	0,0418544	36,40	< 0,0001
Constante	0,0074794	0,0128646	0,580	0,561
resíduo	0,0645743	0,0033527	19,26	< 0,0001

Nota: 53 observações; Wald $\chi^2(1) = 1325,10$; Prob > $\chi^2 < 0,00001$; Log likelihood = 70,01298.

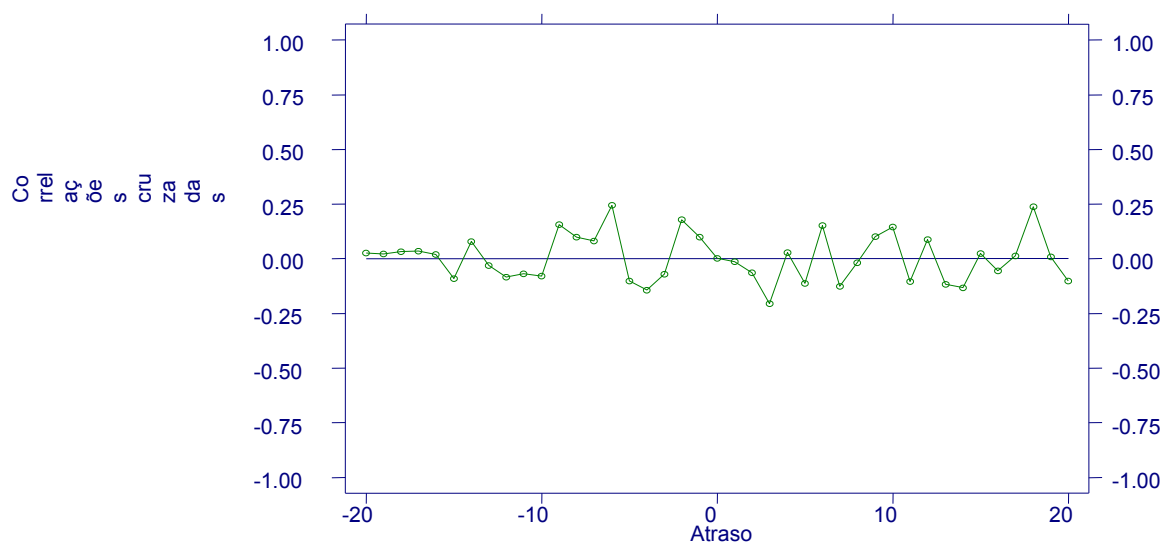
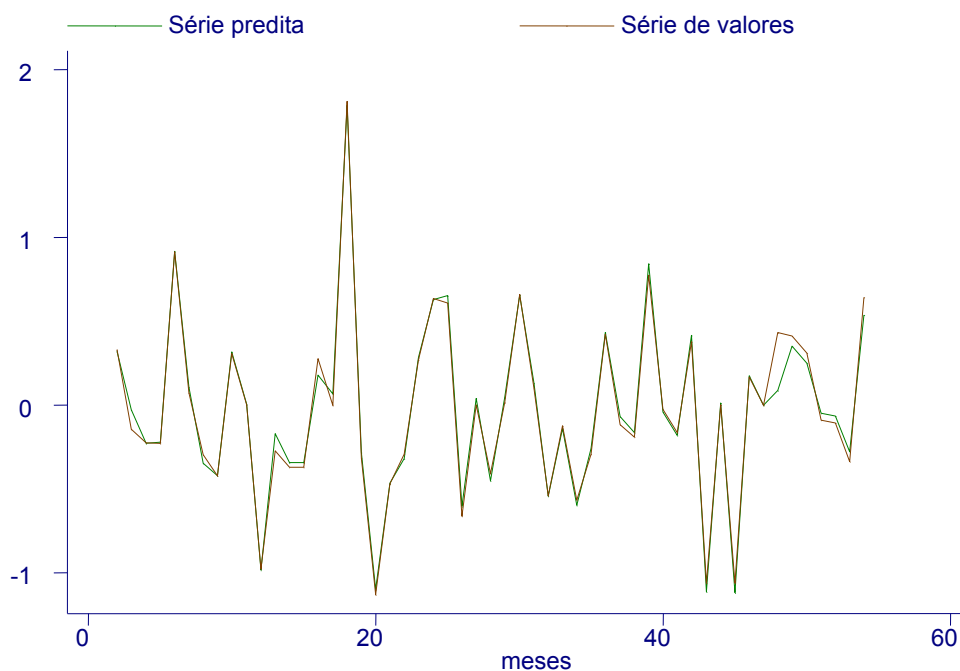
Figura 13 – Correlograma cruzado dos resíduos da série de entrada com os resíduos do filtro aplicado.

Figura 14 – Valores preditos pagos e os valores observados pagos.



Discussão e conclusões

O Sistema de Autorização de Internação Hospitalar é a forma pela qual são remunerados os prestadores de serviços de saúde no Brasil. Essa remuneração baseia-se em uma tabela de pagamentos que a princípio pretendia corresponder aos custos de produção. Porém, o contexto político, as corporações envolvidas e interessadas, a situação econômica - a inflação marcante na década de 80 e parte da década de 90 - a crise fiscal, entre outras possibilidades definem os valores dessa remuneração.

Especificamente, em relação às internações de crianças menores de cinco anos com diagnóstico de diarreia na Bahia, o que se pode observar a partir do modelo ARIMA proposto é que no período estudado, os valores pagos ocorreram no mesmo mês da internação e com um grau de correlação da ordem de 99 %, ou seja, sem parâmetro de atraso (*delay*). Tal fato suscita uma pergunta sobre a direção dessa co-relação: seria, como definida no modelo, a série dependente o pagamento pelas internações, ou seriam as internações consequência dos interesses lucrativos dos prestadores de serviços de saúde ?

O emprego da proposição de Box e Jenkins em situações cujo objetivo é a comparação de séries temporais, pressupõe a adoção de uma estratégia que não apenas leve em conta a autocorrelação, a estacionariedade das séries e a escolha de uma função de transferência adequada. Nesses termos, uma função de transferência corresponderia a um filtro linear ponderado e cujos pesos correspondem à função impulso-resposta do sistema (BOX et al., 1976; MCDOWELL, 2002).

Efetivamente, das diversas situações em que se vislumbra o emprego da modelagem ARIMA, a comparação de séries é a mais complexa, uma vez que todos os pressupostos necessários de serem contemplados em uma análise univariada são pertinentes na análise multiserial, além de outros (PANKRATZ, 1983).

Na aplicação apresentada no presente artigo, em primeiro lugar, teve-se que garantir a estacionariedade das duas séries. No caso da série de saída (valores pagos) havia uma não estacionariedade que foi resolvida facilmente através de uma diferenciação de primeira ordem, enquanto que a série de entrada já era naturalmente estacionária. Avaliação essa efetivada através do exame das funções de autocorrelação e dos testes estatísticos empregados para tal fim. Em seguida, através de um procedimento denominado *pre-whiten*, encontrou-se um melhor modelo ARIMA para a série de internações e este foi utilizado como filtro na série de valores pagos (série de saída), produzindo um resíduo que atendeu aos pressupostos modelares (Figuras 6 a 11 e Tabela 1).

A importância desse procedimento é exatamente estabelecer a relação entre as séries, levando-se em conta a natureza dinâmica das séries (MCDOWELL, 2002). A partir desses passos foi possível se determinar a função de correlação cruzada entre as séries.

Por seu turno, a função de correlação cruzada permitiu a identificação da estrutura de atraso da função de transferência, que no caso, foi simplesmente zero (Figura 12).

A validade desse procedimento (Figura 14), diferentemente de uma correlação simples entre as séries, pode ser demonstrada através da comparação dos valores pagos preditos resultantes do modelo ARIMA proposto a partir da taxa de internações, com os valores realmente pagos em cada mês. Percebe-se uma quase superposição das séries.

Desse modo, a adoção dessa metodologia em séries temporais epidemiológicas garante a contemplação do componente dinâmico das séries, permitindo uma estimativa paramétrica mais confiável e objetiva, assim como previsões menos enviesadas. Além disso, é possível o emprego de outras variáveis preditoras combinadas em contextos multi-séries.

Referências bibliográficas

BOX, G.E.P.;JENKINS, G.M. *Time series analysis : forecasting and control*. Rev. ed. San Francisco: Holden-Day, 1976. xxi, 575 p.

BRASIL. *DATASUS*. Ministério da Saúde. Departamento de Informática do SUS, DATASUS. Disponível em. <http://www.datasus.gov.br>. Acesso: 2002

IBGE. *Instituto Brasileiro de Geografia e Estatística*. IBGE. Disponível em. <http://www.ibge.gov.br>. Acesso: 2002

MARGARIDO, M.A.;ANEFALOS, L.C. Testes de co-integração utilizando o SAS: teoria e aplicação. *Informações Econômicas*, 31(1): p. 21. 2001.

MATOS, O.C.D. *Econometria básica*. 2. ed. São Paulo: Atlas, 1997. 246 p.

MCDOWELL, A. From the help desk: transfer functions. *The Stata Journal*, 2(1): p. 71-85. 2002.

MICROSOFT. Microsoft Visual FoxPro.Versão: 7.0. Microsoft. 2001.

MONTORO FILHO, A.F. *Manual de Economia*. São Paulo: Saraiva, 1988. 443 p.

MORETTIN, P.A.;TOLOI, C.M. *Séries temporais*. São Paulo: Atual Editora, 1986.

PANKRATZ, A. *Forecasting with univariate Box-Jenkins models*. USA: John and Wiley & Sons, 1983. 562 p.

ROTHMAN, K.J.;GREENLAND, S. *Modern epidemiology*. 2. ed. Philadelphia, PA: Lippincott-Raven, 1998. xiii, 737 p.

SANTOS, L.F.J. *Principais técnicas de análise e projeção*. In: Dinâmica da população. São Paulo: BPCS, 1980. p. 103-55.

STATACORP. Stata Statistical Software.Versão: 7. College Station, TX. Stata Corporation. 2001.

Considerações sobre limitações do estudo e conclusão (tese)

A constituição de séries temporais na Epidemiologia representa a possibilidade do encadeamento, acúmulo e registro continuado dos eventos quantificados de interesse, sejam aqueles relacionados à morbidade, à mortalidade, ou fatores de risco, dando uma perspectiva de uma certa dinamicidade e historicidade quantitativa e que apesar disso não deve ser entendida isoladamente, fora dos contextos sociais onde acontecem efetivamente.

Nessa breve versão da história da operacionalização da categoria tempo, pode-se perceber a importância que aquele tem tido nos domínios investigativo e intervencionista da Epidemiologia, uma vez que, em se tratando de doenças infecciosas ou não, ocupa posição de centralidade, chegando a ser fator classificatório das tipologias dos delineamentos de pesquisa ou critério fundamental para a inferência causal.

Mas, mormente, a fonte das séries epidemiológicas no Brasil tem sido bases de dados secundários e que, por seu turno, são geralmente motivo de preocupação quanto à sua disponibilidade, qualidade e confiabilidade. Apesar disso, a base de dados do DATASUS acredita-se ser uma fonte que, mesmo com as limitações possíveis, não pode ser desprezada. Entretanto, urge uma avaliação mais acurada do seu conteúdo, através de outros estudos de validação e confiabilidade da mesma. Esse seria um desafio necessário e primordial, a ser enfrentado através de projetos específicos de pesquisa. Além disso, as bases de dados do DATASUS e outras bases nacionais merecem uma melhor apreciação na perspectiva mais geral de compatibilidade, relevância e adequação às necessidades informacionais do país que, por sua vez precisam ser também discutidas. Porque a questão não é apenas acumular dados, mas acumular dados com quais propósitos.

Na perspectiva mais restrita da construção de séries históricas, as bases de dados do DATASUS são de qualidade questionável, mesmo considerando as dificuldades operacionais de informática devidas à introdução da 10ª revisão da Classificação Internacional de Doenças, pois há uma falta de padronização em relação à estrutura das mesmas, além de períodos onde estranhamente os dados estão bastante fora do comportamento geral das séries. Esse aspecto prejudica a construção de séries temporais confiáveis, demandando do analista o conhecimento de linguagens de programação e a imputação de dados arbitrariamente.

No que tange ao campo de atuação e investigação da Epidemiologia, este se expandiu muito além de sua primeira incursão histórica, incorporou técnicas e métodos de outras disciplinas, tecnologia de ponta e adotou referenciais teóricos distintos. Entretanto, especificamente em relação a abordagens metodológicas de series históricas ainda têm-se muito o que explorar, levando-se em conta o grande potencial tecnológico da atualidade e a incorporação de conceitos tais como autocorrelação, dinamicidade e retroalimentação.

Diante disso, a aplicação da modelagem ARIMA de séries temporais epidemiológicas não tem limites, a não ser o atendimento dos pressupostos de sua formulação, oriundos da teoria das probabilidades e conseqüentemente da matemática e da disponibilidade de dados sequenciados temporalmente e confiáveis. Uma desvantagem de sua aplicação é a necessidade de tecnologias de informática capazes de processar os seus algoritmos de estimação paramétrica, experiência do analista com o manejo de séries temporais, a falta de programas estatísticos de baixo custo que automaticamente realizem tais estimativas e um dos seus produtos, a predição de valores futuros que, embora menos enviesados e mais precisos do que métodos concorrentes, só possui uma confiabilidade maior a curto prazo, mas que a rigor são pseudo-predições. Esse argumento da disponibilidade tecnológica facilitadora, talvez seja uma das explicações do baixo investimento e da pouca exploração, por parte dos epidemiologistas brasileiros, no uso desse procedimento metodológico no estudo de séries históricas.

Desse modo, a família de modelos ARIMA pode ser utilizada na epidemiologia para a obtenção de estimativas mais precisas para predições de fenômenos quantitativos, para a avaliação de impacto de políticas e tecnologias e a observância de associações entre múltiplas séries temporais. As séries podem ser construídas com números absolutos, proporções, médias, coeficientes de incidência, taxas, desde que correspondam a medidas ecológicas mensuradas a intervalos constantes de tempo, ainda que em níveis diferenciados de agregação (multinível). Entretanto, as séries precisam ser estacionárias naturalmente ou transformadas em estacionárias e terem sua variabilidade conhecida e estabilizada. Há ainda a utilização dos modelos ARIMA como uma nova proposta de construção de cartas de controle de qualidade, aprimorando os Gráficos de controle de Shewhart (SAMOBYL et al., 2001), e que pode facilmente ser transposta para o contexto dos diagramas de controle da vigilância epidemiológica, com reais possibilidades de melhores estimativas na detecção do início de uma epidemia ou surto epidêmico e o seu conseqüente acompanhamento e monitoramento.

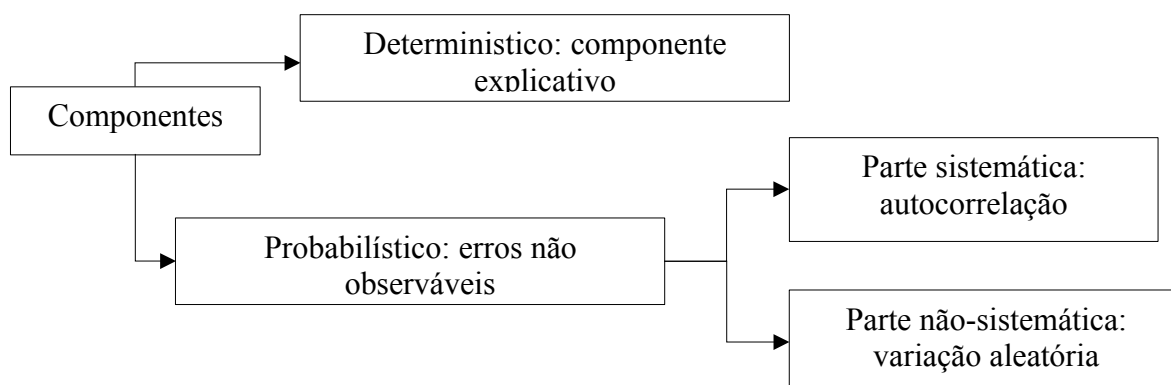
Infelizmente, a utilização de ferramentas analíticas complexas, como o próprio nome sugere, faz parte daquilo que agora é uma norma: a de busca do conhecimento subatomizado, facilmente descartável, fragmentado, e que talvez nos remeta a paradoxos explícitos entre o coletivo e o indivíduo, entre o acesso para todos e muitos para terem acesso, entre o pluralismo de super especialistas em microcosmos, coabitando contiguamente e solitários em um cosmo que cada vez mais desconhecemos.

Apêndice

A - O método de Box-Jenkins

Os modelos de séries temporais possuem dois componentes: um, que descreve o comportamento sistemático da série que é o componente determinístico - este componente é representado por todos os parâmetros da série que não dependem da estrutura do erro e tais parâmetros não predizem com perfeição os valores observados da série; o outro, que descreve um processo subjacente de erros “não observáveis”, conhecido como componente probabilístico do modelo ou “ruído” (Figura 1). Esse componente segue certas leis de probabilidade e é composto de duas partes: uma parte sistemática que é responsável pela autocorrelação da série e uma não-sistemática. É de fundamental importância conhecer a estrutura de autocorrelação da série para o posterior cálculo de estimativas não enviesadas dos parâmetros do modelo (COOK et al., 1979).

Figura 1 – Componentes dos modelos de séries temporais.



Ainda na análise, outros aspectos são considerados, como o seu caráter estacionário, a autoregressão, as médias móveis, e os componentes de sazonalidade (MORETTIN et al., 1986).

Além dos componentes descritos anteriormente, os componentes de intervenção denominados por Box & Jenkins (1976) de “funções de transferência”, são modelos que se ajustam à série a

dependem da natureza da intervenção e do efeito esperado da mesma. A escolha do modelo resulta da natureza do efeito hipotético esperado após cada intervenção ou “replicação”.

Modelo ARIMA (p,d,q)

Os parâmetros estruturais p , d e q , caracterizam o modelo ARIMA. Tais parâmetros podem assumir valores específicos e definir modelos ARIMA particulares. Cada um desses parâmetros se associa com uma propriedade da série temporal: a estacionariedade se associa com o parâmetro d , a autoregressão se associa com o parâmetro p e a média móvel se relaciona com o parâmetro q .

Modelos ARIMA são apropriados apenas para séries temporais que são “naturalmente” estacionárias ou são transformadas em estacionárias através de diferenciação. São consideradas estacionárias as séries que não apresentam uma tendência sistemática de incremento ou decremento no nível da série – tendência secular. De uma maneira geral uma série temporal precisa ser diferenciada mais de uma vez. É importante ressaltar que a diferenciação não afeta os parâmetros da série, principalmente os de intervenção, afetando apenas a sua representação no modelo (PANKRATZ, 1983).

A média aritmética de uma série temporal é definida como sendo $\bar{z} = \frac{1}{n} \sum_{t=1}^n z_t$ e a variância, como $s_z^2 = \frac{1}{n} \sum_{t=1}^n (z_t - \bar{z})^2$. Se z_t observações estiverem próximas em torno de \bar{z} , então a variância será pequena, ou a série será homogênea em torno de média.

No âmbito da econometria, constatou-se que, apesar de os resultados de um processo estocástico independente apresentarem resultados significativos para o teste t e de o coeficiente de determinação (R^2) ser elevado, ainda assim, esses resultados não tinham significado em termos econômicos. De acordo com GUJARATI (1995, p.509) APUD (MARGARIDO et al., 2001) ao se fazer uma regressão entre duas séries de tempo “*esse problema surge porque se ambas séries temporais envolvidas exibirem forte tendência (movimentos sustentados tanto com inclinação positiva, quanto negativa) o alto valor*

observado de R^2 será devido à presença da tendência, e não ao verdadeiro relacionamento entre as duas séries. Por essa razão torna-se muito importante descobrir se o relacionamento entre variáveis econômicas é verdadeiro ou espúrio”.

Segundo as hipóteses da estatística “tradicional” tanto a média como a variância são constantes ao longo do tempo. No entanto, quando ambas deixam de ser constantes ao longo do tempo, isso implica a presença de raiz unitária na série e, conseqüentemente, os pressupostos da estatística “tradicional” perdem validade e os resultados da regressão podem ser espúrios. Para verificar se uma série individualmente é estacionária ou não (isto é, possui raiz unitária) é necessária a execução dos testes de raiz unitária. Quando uma variável individualmente tem raiz unitária, pode acontecer que a sua combinação linear com outra variável seja estacionária e, nesse caso, as duas séries são ditas co-integradas, ou seja, possuem um relacionamento de equilíbrio no longo prazo.

O procedimento de verificação de presença ou não de raiz unitária, tendo como base os resíduos da equação de co-integração, testa a hipótese nula de que a série de resíduo tem raiz unitária ($\rho = 0$) contra a hipótese alternativa de que os resíduos são estacionários ($\rho \neq 0$), ou seja, testa a hipótese nula de que não há co-integração *versus* a hipótese alternativa de que existe co-integração entre as variáveis y_t e x_t .

Diferenciação

A diferenciação consiste em subtrair a primeira observação da segunda, a segunda da terceira, e assim por diante. Em notação matemática: $Y_t - Y_{t-1}$

Para ilustrar, consideremos a seguinte sequência de inteiros: 1,2,3,4,5, ..., n-1, n

Se tratarmos como uma série temporal, ela será não-estacionária. Entretanto, se ela for diferenciada, obteremos:

$$2-1=1$$

$$3-2=1$$

$$4-3=1$$

$$5-4=1 \dots n-n-1=1$$

E teremos uma nova série temporal estacionária, a saber: 1,1,1,1,...,1

Ao se retirar a tendência de uma série temporal (através de diferenciação), isso não afeta qualquer parâmetro determinístico da série, incluindo parâmetros que representem efeitos de intervenção. Para demonstrar isso, consideremos: 2,4,6,8,10,...2N

Observa-se uma nítida tendência secular. Tal tendência poderia ser modelada facilmente através da equação $Y_t = 2 \cdot t$, onde a inclinação é 2. Mas poderíamos representar essa mesma tendência através da equação $Y_t = Y_{t-1} + 2$. Ou seja:

$$Y_2 = Y_1 + 2 = 2 + 2 = 4$$

$$Y_3 = Y_2 + 2 = 4 + 2 = 6$$

$$Y_4 = Y_3 + 2 = 6 + 2 = 8,$$

e assim por diante.

Temos, portanto duas equações que expressam a mesma coisa, mas de formas diferentes. Podemos observar que o processo de diferenciação resulta sempre em uma constante, ou seja:

$$Y_t - Y_{t-1} = \theta_0, \text{ ou } Y_t = Y_{t-1} + \theta_0, \text{ ou } Y_t = \theta_0 t$$

$$Y_t - Y_{t-1} = \theta_0$$

Tal constante, θ_0 , corresponde à inclinação da série.

Em um modelo ARIMA o parâmetro estrutural d indica o número de vezes que a série precisa ser diferenciada para se tornar estacionária. A segunda diferenciação é quadrática, podendo-se chegar até a enésima diferenciação (MORETTIN et al., 1986). Entretanto, de um modo geral, obtém-se a estacionariedade da série com duas diferenciações. Um modelo ARIMA($p,0,q$) significa que a série já é estacionária e não precisa ser diferenciada. Um modelo ARIMA($p,1,q$) indica que a série não é estacionária, mas poderá assim tornar-se caso seja diferenciada uma vez.

Autoregressão

Algumas séries se caracterizam por possuírem uma relação de dependência entre observações adjacentes. O parâmetro estrutural p indica a ordem de autoregressão do modelo. Então, um modelo ARIMA $(1,d,0)$ terá um único termo autorregressivo ($p=1$) e poderá ser escrito como sendo:

$$Y_t = \varphi_1 Y_{t-1} + a_t$$

φ_1 é o coeficiente de correlação (varia entre -1 e $+1$ – limites de estacionariedade) que descreve o grau de dependência entre observações adjacentes na série. Ou seja, o tamanho de φ_1 indica o quanto um valor prévio na série permite prever o valor atual. Se apenas uma observação prévia é usada para prever a observação atual, existe apenas um termo autorregressivo com o seu correspondente peso φ_1 . O termo a_t é o erro da predição chamado “white noise” e corresponde à parte não-sistemática do componente probabilístico, que tem a propriedade de ser normal e identicamente distribuído, com média zero e variância constante σ^2 . Tal como:

$$a_t \sim \text{NID}(0, \sigma^2)$$

Se φ_1 for positivo, grandes valores da série serão seguidos por grandes valores e pequenos valores também por pequenos valores. Se φ_1 for negativo, grandes valores serão seguidos por pequenos valores e vice-versa.

Um modelo ARIMA $(2,0,0)$ possui dois termos autorregressivos. Ou seja, a observação Y_t atual é descrita em parte por duas observações adjacentes anteriores e pode ser escrita como:

$$Y_t = \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + a_t$$

Teoricamente um processo ARIMA $(p,d,0)$ poderá ter N termos autorregressivos, entretanto, na prática, processos ARIMA $(3,d,0)$ são raros.

Algumas séries temporais apresentam um erro persistente de uma observação para a seguinte. Tais séries são bem descritas através de modelos de médias móveis, indicadas pela ordem do

parâmetro estrutural q . Em um processo ARIMA(0,d,1), por exemplo, a atual observação Y_t é descrita a partir do erro anterior, a_{t-1} . Tal modelo pode ser escrito como:

$$Y_t = a_t - \theta_1 a_{t-1}, \text{ onde } \theta_1 \text{ é um coeficiente de correlação.}$$

Do mesmo modo, um modelo ARIMA(0,0,2) pode ser apresentado como:

$$Y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$$

Todos os processos ARIMA apresentados até o momento consideram apenas um termo estrutural da série temporal, porém teoricamente modelos mistos podem ser formulados. Um modelo ARIMA(1,0,1) descreve uma série temporal na qual a observação atual, Y_t , é predita a partir da observação precedente, Y_{t-1} , e o antecedente erro aleatório, a_{t-1} .

$$Y_t = \varphi_1 Y_{t-1} - \theta_1 a_{t-1} + a_t$$

Nesse processo deve-se identificar a parte sistemática do componente probabilístico através das funções de autocorrelação (FAC) e autocorrelação parcial (FACP). Uma vez que um provável modelo seja identificado, os parâmetros φ e θ são estimados. Em seguida, através da análise de resíduos verifica-se a bondade do modelo. Tal procedimento em três estágios – identificação, estimação e diagnóstico – deve ser repetido até que se encontre um modelo adequado (BOX, G. E. P. et al., 1976).

Modelo ARIMA (P,D,Q)_s

A sazonalidade no contexto das séries temporais é definida como mudanças periódicas sistemáticas que apresentam um padrão de comportamento que se repete a intervalos mais ou menos regulares de tempo. Como visto anteriormente, a autocorrelação implica na dependência estrutural entre observações adjacentes. Se o componente sazonal é forte, observações que estão separadas por um ciclo mantêm entre si uma estrutura de dependência: é a denominada autocorrelação sazonal. Nessa perspectiva, a estrutura sazonal pode ser incorporada no componente estocástico da série. Assim, parâmetros para as estruturas autoregressivas, média móvel e mista podem ser estabelecidos, respectivamente, P, D e Q. O parâmetro P especifica o número de termos autorregressivos sazonais, o parâmetro D especifica o número de vezes que a série deve ser diferenciada sazonalmente e o parâmetro Q especifica o número de termos de médias móveis sazonais. O quarto parâmetro S denota o comprimento do ciclo. Este é usualmente igual a 4 para dados trimestrais, 12 para dados mensais, 52 para dados semanais, e assim por diante.

O método convencional para especificar um modelo sazonal é: ARIMA (p,d,q) (P,D,Q)_s. Aqui as estruturas regulares e sazonais são multiplicadas, o que significa a existência de um produto-termo no modelo. Entretanto, um modelo ARIMA sazonal só deve ser usado quando se têm evidências de substancial sazonalidade (PANKRATZ, 1983).

Considere-se um modelo ARIMA (1,00)(1,00)₁₂. Esse modelo implica que uma série temporal mensal possui um componente de autorregressão regular e um componente de autorregressão sazonal. Se esses componentes são modelados aditivamente, temos:

$$Y_t = \phi_1 Y_{t-1} + \phi_{12} Y_{t-12} + a_t$$

Entretanto, esses componentes são modelados multiplicativamente, ou seja:

$$Y_t = \phi_1 Y_{t-1} + \phi_{12} Y_{t-12} - \phi_1 \phi_{12} Y_{t-13} + a_t$$

A diferença entre sazonalidade aditiva ou multiplicativa é o produto-termo $\phi_1 \phi_{12} Y_{t-13}$.

Algumas vezes o produto-termo é tão pequeno que pode ser tratado como sendo zero. Nesse caso, não existe uma real diferença entre sazonalidade aditiva ou multiplicativa. De fato, o que faz a diferença é a inclusão da observação correspondente ao *lag-13* (treze observações antecedentes). Modelos de sazonalidade aditiva predizem a observação atual, Y_t , a partir da observação precedente, Y_{t-1} , e a partir da observação de um período anterior, Y_{t-12} . Modelos de sazonalidade multiplicativa, acrescentam a observação Y_{t-13} , que é a observação imediatamente precedente a observação de um período anterior para predizer a atual observação Y_t .

Desse modo, modelos multiplicativos sazonais são mais informativos (usam mais informação) do que os modelos aditivos sazonais, e conseqüentemente são melhores para modelar a sazonalidade. Entretanto, um modelo ARIMA sazonal só deverá ser utilizado se houver uma substancial sazonalidade na série temporal evidenciada através da FAC e da FACP.

Não-estacionariedade sazonal

Uma série pode ser não-estacionária de ciclo para ciclo, assim como de observação para observação. Do mesmo modo que as séries temporais podem apresentar uma tendência em seu componente regular, também o podem em seu componente sazonal. Se ambas as tendências se manifestarem conjuntamente, a diferenciação se faz necessária levando-se em conta os dois componentes.

$$Y_t = \phi_{12} Y_{t-1} + a_t \text{Componente de intervenção}$$

Ao ajustar um modelo ARIMA (p,d,q) à uma série temporal, pode-se descrevê-la como “*noise*”. Tal modelo proporciona algum grau de predição e corresponde à hipótese nula: $Y_t = \text{noise}$. Ao se adicionar um componente de intervenção, obtêm-se: $Y_t = \text{noise} + \delta$ intervenção. Se o componente de intervenção aumenta a capacidade de predição do modelo, os seus parâmetros serão estatisticamente significantes. E essa é a hipótese a ser testada.

Assim, a questão principal é estabelecer se a intervenção contribui significativamente para prever o comportamento de uma série temporal.

Neste ponto é importante salientar que tal análise por ela mesma não testa causalidade. Apenas, se uma mudança estatisticamente significativa ocorre em um ponto específico da série.

Os componentes de intervenção de Box-Jenkins são denominados “funções de transferência”, termo esse derivado da Engenharia (PANKRATZ, 1983). As três principais funções de transferência modelam uma abrupta, porém constante mudança na série; uma gradual e constante mudança na série; e uma abrupta e temporária mudança na série.

O mais simples efeito devido à intervenção sobre uma série temporal ocorre quando uma mudança no nível da série quando do momento da intervenção (Figura 2), tal como:

$Y_t = \omega I_t + \text{noise}$, onde ω é um parâmetro que mede a magnitude da mudança abrupta e constante ocorrida em uma dada série; *noise* é um processo ARIMA (p, d, q) e I_t representa a variável independente “intervenção no momento t”, do tipo *dummy*.

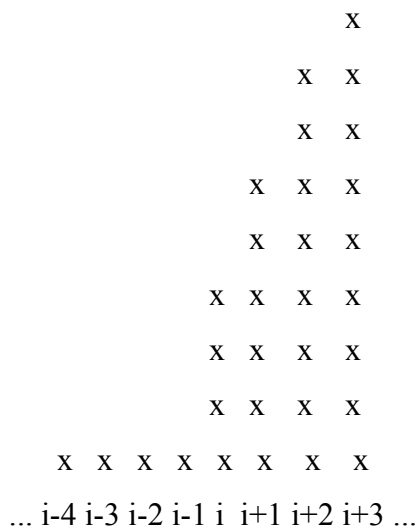
Figura 2 – Padrão do efeito da intervenção.

<p>Padrão do efeito da intervenção quando $\omega > 0$:</p> <p style="text-align: center;">Pós-tratamento</p> <p style="text-align: center;">oooooooooooooooooooo</p> <p style="text-align: center;">o</p> <p style="text-align: center;">oooooooooooooooooooo</p> <p>Pré-tratamento</p>	<p>Padrão do efeito da intervenção quando $\omega < 0$:</p> <p style="text-align: center;">Pré-tratamento</p> <p style="text-align: center;">oooooooooooooooooooo</p> <p style="text-align: center;">o</p> <p style="text-align: center;">oooooooooooooooooooo</p> <p style="text-align: center;">Pós-tratamento</p>
--	--

$I_t = 0$, antes da intervenção e $I_t = 1$, após a intervenção.

A equação que define um efeito de uma intervenção de modo gradual e constante sobre uma série temporal (Figura 3) pode ser definida assim: $Y_t = \delta Y_{t-1} + \omega I_t + \text{noise}$, cujo padrão de intervenção é o seguinte:

Figura 3 – Efeito de intervenção gradual e constante.



Ou seja, o nível da série muda gradualmente a partir da intervenção. O incremento de um período de tempo para outro, no pós-tratamento se torna cada vez menor até ser aproximadamente zero. A mudança no nível da série é dada por:

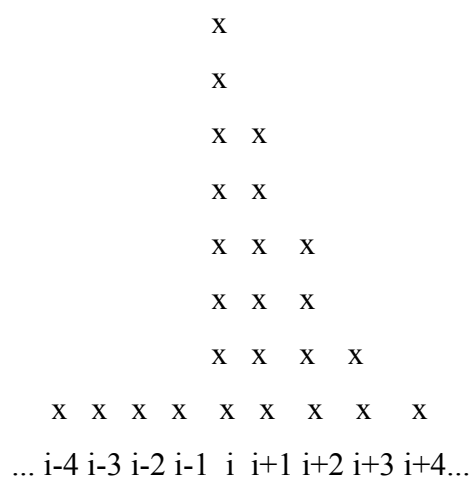
$$\text{Mudança no nível} = \frac{\omega}{1 - \delta}$$

A mudança no nível no momento da intervenção é agora ω e a taxa de mudança de um momento para o outro é $\delta^n \omega$. Portanto, o parâmetro δ determina o quanto a série irá mudar o seu nível.

Um exemplo desse tipo de efeito pode ser encontrado em Box e Tiao (1975) ao analisarem uma série de níveis de poluição do ar em Los Angeles, antes e após ações de controle da poluição. Nesse caso, não é de se esperar uma mudança brusca nos níveis de poluição, mas sim uma redução gradual ao longo do tempo após o controle.

Finalmente, quando o padrão do efeito possui um caráter pulsátil (*pulse function*) (Figura 4), temos: $I_t=0$, antes da intervenção e $I_t=1$, no momento da intervenção e $I_t=0$, após a intervenção. Portanto, no n -ésimo momento após a intervenção, o nível da série será $Y_{i+n}=\delta^n \omega$, com o seguinte esquema representativo (denominado *decaying spike*):

Figura 4 – Efeito de intervenção com caráter pulsátil.



A intervenção provoca uma profunda alteração no nível da série e a sua magnitude corresponde a ω , altura do pico (*spike*). E logo em seguida há um retorno aos níveis anteriores à intervenção. O parâmetro δ corresponde à taxa momentânea de decremento.

Estratégia de modelagem

A modelagem ARIMA é efetuada em estágios bem definidos (BOX, G. E. P. et al., 1976). O primeiro estágio, ou fase de identificação do modelo, compreende o cálculo das funções de autocorrelação e autocorrelação parcial com a intenção de verificar não-estacionariedade da série e conseqüentemente a necessidade de sua diferenciação. O produto desse estágio é um modelo ARIMA, dito, tentativo.

No segundo estágio, ou fase de estimação, os valores dos parâmetros do modelo proposto são estimados, φ e θ . Na fase três, ou estágio de diagnóstico, funções de autocorrelação e autocorrelação parcial são calculadas para os resíduos do modelo tentativo com a intenção de

averiguar a bondade do modelo. Se não existirem picos no lag 1 e nos lags sazonais, e o teste Q for não significativo, o modelo será considerado adequado.

Se o modelo for adequado, passa-se para o quarto estágio, ou fase de teste de hipótese sobre a intervenção. Primeiramente, um modelo de função de transferência é proposto e em seguida são estimados os parâmetros φ , θ , ω e δ (phi, teta, ômega e sigma, respectivamente). Se o tamanho do efeito da intervenção for grande, faz-se o diagnóstico da adequabilidade do modelo conjunto, função de transferência-ARIMA e sua interpretação.

Características de um bom modelo:

De acordo com COOK (1979), para a escolha de um bom modelo ARIMA, critérios estatísticos e experiência do analista devem ser combinados em um processo iterativo e recorrente:

1. Ser parcimonioso – usar o menor número de coeficientes necessários para explicar os dados;
2. Ser estacionário;
3. Ter coeficientes estimados de alta qualidade - φ 's e θ 's não altamente correlacionados e valores absolutos da estatística t maiores ou iguais a 2;
4. Possuir resíduos não-correlacionados;
5. Ajustar os dados disponíveis (passado) o suficiente para satisfazer o analista – erro quadrático médio e erro percentual absoluto médio aceitáveis;
6. Predizer o futuro (dados futuros) satisfatoriamente – pequeno erro de predição após monitoramento.

B – Algoritmos computacionais**Algoritmo *xbase* para estimativa populacional mensal geométrica**

```
CLEAR
USE d:\aih\bames
zap
GO top
nmes=49
nmes2=84
pba0491=1450624
pba0496=1320774
batx04=(1/nmes)*(log(pba0496/pba0491))
APPEND BLANK
replace pba04 WITH pba0491
a=pba0491
b=0
i=1
replace n WITH i
i=i+1
skip
DO WHILE RECNO()<=nmes2
  APPEND BLANK
  replace n WITH i
  b=a+(a*batx04)
  replace pba04 WITH b
  a=b
  SKIP
  i=i+1
ENDDO
```

Algoritmo final de geração da base total de óbitos de 1979 a 1998.

```

* 2002
* Carlos Mauricio Cardeal Mendes
clear
set memory 250m
use d:\aih\dorba79
append using d:\aih\dorba80
append using d:\aih\dorba81
append using d:\aih\dorba82
append using d:\aih\dorba83
append using d:\aih\dorba84
append using d:\aih\dorba85
append using d:\aih\dorba86
append using d:\aih\dorba87
append using d:\aih\dorba88
append using d:\aih\dorba89
append using d:\aih\dorba90
append using d:\aih\dorba91
append using d:\aih\dorba92
append using d:\aih\dorba93
append using d:\aih\dorba94
append using d:\aih\dorba95
append using d:\aih\dorba96
append using d:\aih\dorba97
append using d:\aih\dorba98
keep do tipobito dataobit estcivile sexo idade muniocor ocupacao instruca
causabas
gen str2 m=substr(dataobit,3,2)
gen str4 a=substr(dataobit,1,2)
gen str2 d=substr(dataobit,5,2)
gen ano=real(a)
gen dia=real(d)
gen mes=real(m)
replace dia=1 if dia==0 | dia==.
gen obito=mdy(mes,dia,ano+1900)
format obito %dD/N/CY
gen ida=real(substr(idade,2,2)) if real(substr(idade,1,1))==4
replace ida=real(substr(idade,2,2))+100 if real(substr(idade,1,1))==5
replace ida=real(substr(idade,2,2))/12 if real(substr(idade,1,1))==3
replace ida=real(substr(idade,2,2))/30.125 if real(substr(idade,1,1))==2
replace ida=real(substr(idade,2,2))/8766 if real(substr(idade,1,1))==1
replace ida=. if real(substr(idade,1,1))==0 | real(substr(idade,1,1))==.
drop idade
rename ida idade
gen str3 cid=substr(causabas,1,3)
count if cid=="009"
drop causabas m d a ano dia mes dataobit
save d:\aih\ob7998

```

Algoritmo final para acesso à base de dados de óbitos de 1979 a 1998, seleção de CIDs e faixa etária, agregação mensal e modelagem ARIMA.

* 2002

* Carlos Mauricio Cardeal Mendes

```

set logtype text
set more off
clear
set memory 200m
set matsize 100
use "d:\AIH\ob7998.dta"
*log using "d:\AIH\ob7998.doc",replace
keep if idade>0 & idade<5
gen total=1
gen m9=1 if cid=="009" | cid=="A09"
gen mp=1 if cid=="485" | cid=="486" | cid=="J12" | cid=="J13" | cid=="J14"
| cid=="J15" | cid=="J18"
keep if cid=="009" | cid=="A09" | cid=="485" | cid=="486" | cid=="J12" |
cid=="J13" | cid=="J14" | cid=="J15" | cid=="J18"
tab1 cid
gen mdata = mofd(obito)
*collapse (count) total, by(mdata)
collapse (count) m9 (count) mp, by(mdata)
format mdata %tm
tsset mdata
gen n=_n
drop if n==241
gr m9 mp mdata,c(11) s(..) ylabel sort
gr mp mdata,c(11) s(..) ylabel sort
gr m9 mdata,c(11) s(..) ylabel sort xlabel
univar m9
corrgram m9
corrgram D1.m9
set more off
*drop lm9
gen lm9=ln(m9)
gen sqm9=sqrt(m9)
corrgram D1.lm9
set matsize 300
*arima D1.lm9,ar( 12) ma( 12) nocons
* ou esse?
arima D1.lm9, ar( 12) ma(4 12) nocons
drop mpred
drop r
predict mpred,y
predict r,resid
corrgram r
label var lm9 "log mort. diarreia < 5 anos"
label var mpred "log mort. diarreia predita"
gr lm9 y mdata,c(11) s(..) ylabel xlabel
swilk r
summ r
gr r,normal xlabel ylabel (0 .2 to .8)
wntestq r
wntestb r,table
gen mpred=F1.y
gr lm9 mpred mdata,c(11) s(..) ylabel xlabel

```

Algoritmo final de acesso à base de dados de interações de Vitória da Conquista de 1995 a 2002 e modelagem ARIMA para avaliação de impacto.

```

2002
Carlos Mauricio Cardeal Mendes
clear
set memory 4m
set more off
use "D:\CMCM\Tese\vc952002.dta"
preserve
tsset tempo
set matsize 200
univar aih
univar aih if tempo<58
univar aih if tempo>57
gen z=1 if tempo<58
replace z=2 if tempo>=58
label define o 1 "Ate setembro de 1999" 2 "A partir de outubro de 1999"
label values z o
set logtype text
*log using "D:\CMCM\Tese\correlol.doc",replace
corrgram aih if tempo<58,lags(20)
ac aih if tempo<58
*ac aih if tempo<58
wntestq aih if tempo<58,lags(10)
wntestb aih if tempo<58
dfuller aih
pperron aih
gen x=0
replace x=1 if tempo>=58
replace x=0 if x==.
arima aih if tempo<58,arima(0,0,0)
predict r,res
swilk r if tempo<58
gr r if tempo<58
gr r tempo if tempo<58
univar r if tempo<58
arima aih x ,arima(0,0,0) robust
gr aih tempo,c(11) xlabel(1,5 to 90) s(.) ylabel(2000, 2200 to 4800)
yline(2580.576 3648.351)
drop r
gr r,normal xlabel
gr r tempo
univar r
sktest r

```

Algoritmo final de acesso à base de dados de internações devidas à diarreia, imputação, cálculo de taxa e proporção, modelos ARIMA e correlação cruzada.

2002

Carlos Maurício Cardeal Mendes

```

clear
use "D:\CMCM\Tese\iradia.dta"
set more off
set matsize 100
set logtype text
tsset n2
gen i9im=i9
replace i9im=(3081+2451)/2 if n2==44
replace i9im=(2899+1023)/2 if n2==73 | n2==74
gen valorm=valor
replace valorm=(343522.52+268922.02)/2 if c1==14
replace valorm=(330562.24+117976.99)/2 if c1==43 | c1==44
gen valore=valorm/100000
gen itotalim=itotal
replace itotalim=(14157+10602)/2 if n2==44
replace itotalim=(13698+12175)/2 if n2==68
gen pi9im=i9im*100/itotal
gen pg9=valorm*100/gastot
label var valore "Reais/100.000"
gen txint9im=i9im*1000/pbam5
label var txint9im "Int. por 1.000"
* ponto 55 corresponde a jan de 98 - mudança de cid
gr txint9im valore c1 if c1<55,c(111) s(...) sort ylabel xlabel(0 5 to 55)
*browse txint9im incm9 valore c1 n n2 pbam5 m9 mdata
*xcorr txint9im valore if c1<55,needle
*xcorr txint9im valore,table
*univar txint9im valorm if c1<55
keep c1 valore txint9im
drop if c1==. | c1>54
tsset c1
corrgram valore
wntestq valore
gen val=D1.valore
corrgram val
wntestq val
gr val c1 ,c(1) s(.) sort ylabel(-2 -1 to 2) xlabel(0 5 to 55) yline(0)
corrgram txint9im
wntestq txint9im
gen txi=D1.txint9im
corrgram txi
wntestq txi
dfuller txi
pperron txi
arima txi
predict alpha,resid
corrgram alpha
wntestq alpha
rename txi tx1
rename val txi
predict beta,resid
rename txi val
rename tx1 txi
summ alpha
scalar s_alpha=r(sd)
summ beta in 2/1

```



```
scalar s_beta=r(sd)
xcorr alpha beta in 2/1,saving(alpha_beta,replace)
xcorr alpha beta in 2/1,generate(cross) table
gen lag=_n-21
gen std=-2/sqrt(54)
gen zero=0
gen ir=(s_beta/s_alpha)*cross
list lag cross ir std if ir !=. & lag>=0 & cross<std
set matsize 200
arima val l(0).txi
predict nr,resid
predict yhat
corrgram nr
wntestq nr
xcorr alpha nr in 2/1,saving(alpha_nr,replace)
gr yhat val c1,c(111) s(...) xlabel ylabel
```

Anexos

A - Estrutura da base de dados de internações hospitalares – DATASUS.

Documentação do arquivo RD<UF><AAMM>.DBF, contido no arquivo RD<UF><AAMM>.ARJ onde UF = sigla do Estado e AAMM = ano/mes com 4 posições.

Arquivo reduzido com o movimento das AIH's processadas no mês de competência.

I) Layout

Campo	Nome	Tipo/Tam	Descrição
01	AIH	C 10	Numero da AIH.
02	IDENT	C 01	Identificador da AIH (1=normal 5=continuação 8=pagamento suplementar)
03	ESTADO	C 02	UF SINPAS de Subordinação.
04	PRCD	C 08	Procedimento realizado.
05	CGC_HOSP	C 14	CGC do Hospital no formato RRRRRRRRFFFFDD, onde os R são o radical, os F são o numero da filial e os D os dígitos de controle.
06	VALOR_TOT	N 15,2	SH+SP+SADT+ORTP+SANG+RN
07	VALOR_SH	N 15,2	Serviços Hospitalares.
08	VALOR_SP	N 15,2	Serviços Profissionais.
09	VALOR_SADT	N 15,2	Serviços Auxiliares de Diagnose e Terapia.
10	VALOR_ORTP	N 15,2	Ortese e Protese.
11	VALOR_SANG	N 15,2	Sangue
12	VALOR_RN	N 12,2	Atendimento a recém-nato
13	DIAS_UTI	N 02	Total de dias em UTI.
14	MORTES	N 01	Identificação óbito (0=nao;1=sim)
15	PERMANEN	N 05	Tempo de permanência do paciente
16	CEP_PACIEN	C 08	CEP do endereço do paciente
17	SEXO	C 01	Sexo (1=masculino;3=feminino)
18	DATA_NASC	C 06	Data de nascimento do paciente no formato AAMMDD.
19	DATA_INT	C 06	data de internação no formato AAMMDD
20	DATA_ALTA	C 06	Data da saída no formato AAMMDD.
21	DATA_APRES	C 04	Data da apresentação da AIH no formato AAMMDD.
22	CID	C 06	código CID (diagnostico principal).
23	NATUREZA	C 02	Natureza do Hospital.
24	SEMIPLN	C 01	Indica se o hospital foi pago em semi-plena ou não. 0 = não foi pago em semi-plena; 1 = foi pago em semi-plena.
25	ESPECIALID	C 01	Especialidade.
26	COBRANCA	C 02	Motivo de cobrança.
27	UF_LOC	C 02	UF SINPAS de localização do Hospital.
28	MUNIC_LOC	C 03	Município SINPAS de localização do Hospital.
29	DV_LOC	C 01	Digito verificador da localização.
30	UF_ZI	C 02	UF SINPAS de pagamento do Hospital.
31	MU_IBGE_LO	C 04	Município IBGE de localização do Hospital.
32	UF_IBGE_LO	C 02	UF IBGE de localização do Hospital.
33	COD_IDADE	C 01	0=ignorada, 2=dias, 3=meses e 4=anos.
34	IDADE	N 02	Idade do paciente na data da internação, segundo o campo COD_IDADE.
35	COD_ARQ	C 01	Código identificador de registro reduzido de AIH.

36	CONT	N 07	Contador utilizado nos aplicativos, valor default=1
37	SH_DOLAR	N 09,2	Serviços Hospitalares, em dólar.
38	SP_DOLAR	N 09,2	Serviços Profissionais, em dólar.
39	SADT_DOLAR	N 09,2	Serv. Auxiliares de Diagnose e Terapia, em dólar.
40	ORTP_DOLAR	N 09,2	Ortese e Protese, em dólar.
41	SANG_DOLAR	N 09,2	Sangue em dólar.
42	RN_DOLAR	N 09,2	Recem-nato em dólar.

II) OBSERVAÇÕES:

1) O valor 5 para o campo IDENT corresponde a internações de longa duração (notadamente na área de psiquiatria) ou casos FPT (Fora de Possibilidade Terapêutica).

2) O valor 8 para o campo IDENT corresponde a pagamentos suplementares para AIH do tipo 1 que foram pagas em competências anteriores. Tal fato ocorreu em FEV/92.

3) CID = Código Internacional de Doenças

4) A NATUREZA , caracterizada da seguinte forma:

- 10 = Hospital Próprio
- 20 = Hospital Contratado
- 30 = Hospital Federal
- 31 = Hospital Federal que não recebe via AIH
- 40 = Hospital Estadual
- 50 = Hospital Municipal
- 60 = Hospital Filantrópico
- 70 = Hospital Universitário
- 80 = Hospital de Sindicato
- 90 = Hospital Universitário com Pesquisas

5) Especialidade

- 1 = Cirurgia
- 2 = Obstetrícia
- 3 = Clínica Médica
- 4 = Crônico ou FPT (Fora de Possibilidade terapêutica)
- 5 = Psiquiatria
- 6 = Fisiologia
- 7 = Pediatria
- 8 = Reabilitação
- 9 = Psiquiatria/dia

6) Motivo de Cobrança

Alta:

- 11: Curado
- 12: Melhorado
- 13: Inalterado
- 14: A pedido
- 15: Internado para diagnostico
- 16: Administrativa
- 17: Indisciplina
- 18: Alta por evasão
- 19: Complementação de tratamento em ambulatório

Permanência maior que 30 dias:

- 21: Características próprias da doença
- 22: Intercorrência
- 23: Motivo social
- 24: Doença crônica
- 25: Impossibilidade de vivência sócio-familiar

Transferência:

- 31: Tisiologia
- 32: Psiquiatria
- 33: Clínica médica
- 34: Cirurgia
- 35: Obstetrícia
- 36: Berçário
- 37: Pediatria
- 38: Isolamento
- 39: Outros (UTI, Medicina física, Radioterapia etc)

Óbito com autópsia:

- 41: Ocorrido nas primeiras 48 horas após a internação em estado agônico ou pré-agônico
- 42: Ocorrido nas primeiras 48 horas após a internação em estado não agônico ou pré-agônico
- 43: Ocorrido a partir de 48 horas de internação

Óbito sem autópsia:

- 51: Ocorrido nas primeiras 48 horas após a internação em estado agônico ou pré-agônico
- 52: Ocorrido nas primeiras 48 horas após a internação em estado não agônico ou pré-agônico
- 53: Ocorrido a partir de 48 horas de internação

Alta por recuperação:

- 61: Em politraumatizado com menos de 24 horas da 1ª cirurgia
- 62: Em politraumatizado 24 a 48 horas após 1ª cirurgia
- 63: Em politraumatizado 48 a 72 horas após 1ª cirurgia
- 64: Em politraumatizado acima de 72 horas após 1ª cirurgia
- 65: Em cirurgia de emergência com menos de 24 horas da 1ª cirurgia
- 66: Em cirurgia de emergência 24 a 48 horas após 1ª cirurgia
- 67: Em cirurgia de emergência 48 a 72 horas após 1ª cirurgia
- 68: Em cirurgia de emergência acima de 72 horas após 1ª cirurgia

7) Algoritmo de cálculo da idade:

```
x(em dias) = data_internacao - data_nasc
Se ano-alta < 84
  Idade <-- 0
  Cod_idade <-- 0
Senao
  Se x > 364
    Idade <-- int(x/365)
    Cod_idade <-- 4
  Senao
    Se x > 29
      Idade <-- int(x/30)
      Cod_idade <-- 3
    Senao
      Se x > 0
        Idade <-- x
        Cod_idade <-- 2
      Senao
```

```

Se x = 0 e data_internacao = data_nasc e ano_internacao > 0
  Idade      <-- 0
  Cod_idade  <-- 1
Senao
  Se ano_nasc = ano_internacao e x < 0
    Idade      <-- 99
    Cod_idade  <-- 4
  Senao
    Se x < 0
      Idade      <-- 100 - ( ano_nasc - ano_internacao)
      Cod_idade  <-- 4
    Senao
      Idade      <-- 0
      Cod_idade  <-- 0

```

8) Campos em dólar: e' utilizado o dólar comercial de venda, do ultimo dia do mês da apresentação, ou seja, para o mês de competência x e utilizado o dólar do ultimo dia do mês x + 1. Por exemplo, para AAMM=9302 use-se Cr\$25.125,00.

9) O valor do sangue esta somado ao do sadt nos arquivos de 1992.

10) Tabela de US\$ de 1994 Tabela de US\$ de 1993 Tabela de US\$ de 1992

Tabela de US\$ de 1994			Tabela de US\$ de 1993			Tabela de US\$ de 1992		
Jan	CR\$	458,66	Jan	Cr\$	15.720,00	Jan	Cr\$	1.319,20
Fev	CR\$	637,23	Fev	Cr\$	19.858,00	Fev	Cr\$	1.630,45
Mar	CR\$	913,17	Mar	Cr\$	25.125,00	Mar	Cr\$	1.197,90
Abr	CR\$	1.302,20	Abr	Cr\$	32.268,10	Abr	Cr\$	2.395,80
Mai	CR\$	1.875,28	Mai	Cr\$	41.700,00	Mai	Cr\$	2.849,10
Jun	CR\$	2.750,00	Jun	Cr\$	54.328,00	Jun	Cr\$	3.446,50
Jul	CR\$	2.585,00	Jul	Cr\$	71.154,00	Jul	Cr\$	4.204,60
Ago	R\$	0,89	Ago	CR\$	94,65	Ago	Cr\$	5.130,00
Set	R\$	0,85	Set	CR\$	128,08	Set	Cr\$	6.400,10
Out	R\$	0,85	Out	CR\$	174,00	Out	Cr\$	8.034,20
Nov	R\$	0,85	Nov	CR\$	235,99	Nov	Cr\$	9.949,90
Dez	R\$	0,85	Dez	CR\$	320,92	Dez	Cr\$	12.343,00

Tabela de US\$ de 1995 Tabela de US\$ de 1996

Tabela de US\$ de 1995			Tabela de US\$ de 1996		
Jan	R\$	0,84	Jan	R\$	0,98
Fev	R\$	0,85	Fev	R\$	0,99
Mar	R\$	0,90	Mar	R\$	0,99
Abr	R\$	0,91	Abr	R\$	0,99
Mai	R\$	0,90	Mai	R\$	0,99
Jun	R\$	0,92	Jun	R\$	1,00
Jul	R\$	0,94	Jul	R\$	1,01
Ago	R\$	0,95	Ago	R\$	1,02
Set	R\$	0,96	Set	R\$	1,02
Out	R\$	0,96	Out	R\$	1,03
Nov	R\$	0,96	Nov	R\$	1,03
Dez	R\$	0,97	Dez	R\$	1,04

Tabela de US\$ de 1997

Tabela de US\$ de 1997		
Jan	R\$	1,04
Fev	R\$	1,05
Mar	R\$	
Abr	R\$	

Mai R\$
 Jun R\$
 Jul R\$
 Ago R\$
 Set R\$
 Out R\$
 Nov R\$
 Dez R\$

11) Os campos CEP_PACIEN e DATA_INT estão zerados nos arquivos de 92.

12) O tempo de permanência é calculado segundo o seguinte algoritmo:

$$\begin{aligned} \text{PERMANEN} &= \text{n\# de dias de data_alta se IDENT} = 5 \\ &= \text{data_alta} - \text{data_int} \quad \text{se IDENT} <> 5 \end{aligned}$$

13) O campo VALOR_RN foi incluído a partir da competência Dez/94.

III) Relat.,o dos códigos IBGE e SINPAS para cada UF.

SINPAS	IBGE	SIGLA	NOME
02	27	AL	Alagoas
03	13	AM	Amazonas
04	29	BA	Bahia
05	23	CE	Cear
06	50	MS	Mato Grosso do Sul
07	32	ES	Espírito Santo
08	52	GO	Goiás
09	21	MA	Maranhão
10	51	MT	Mato Grosso
11	31	MG	Minas Gerais
12	15	PA	Pará
13	25	PB	Paraíba
14	41	PR	Paraná
15	26	PE	Pernambuco
16	22	PI	Piauí
17	33	RJ	Rio de Janeiro
18	24	RN	Rio Grande do Norte
19	43	RS	Rio Grande do Sul
20	42	SC	Santa Catarina
21	35	SP	São Paulo
22	28	SE	Sergipe
23	53	DF	Distrito Federal
24	12	AC	Acre
25	16	AP	Amapá
26	11	RO	Rondônia
27	14	RR	Roraima
28	17	TO	Tocantins

B - Estrutura do arquivo de dados do SIM

Cada óbito dá entrada no sistema com as seguintes variáveis:

1. Tipo de óbito
2. Mês do óbito
3. Ano do óbito
4. Estado civil
5. Sexo
6. Idade
7. Local de ocorrência
8. Município de ocorrência
9. Município de residência
10. Ocupação habitual
11. Naturalidade (unidade federada ou país)
12. Grau de instrução
13. Assistência médica durante a doença que ocasionou a morte
14. Médico que assina o atestado que atendeu ao falecido
15. O diagnóstico foi confirmado por:
 - exame complementar
 - cirurgia
 - necrópsia
16. Causa Básica do Óbito segundo a 9ª Revisão da Classificação Internacional de Doenças (8)
17. Para causas externas
 - tipo de violência (homicídio, suicídio, acidente)
 - acidente de trabalho
 - local do acidente

Para óbitos fetais e de menores de um ano, são acrescentadas as seguintes variáveis:

- a) Ocupação habitual do pai
- b) Grau de instrução do pai
- c) Ocupação habitual da mãe

- d) Idade da mãe
- e) Grau de instrução da mãe
- f) Número de filhos tidos pela mãe (incluindo aquele cujo óbito está sendo declarado):
 - nascidos vivos
 - nascidos mortos
- g) Duração da gestação em semanas
- h) Tipo de gravidez
- i) Tipo de parto
- j) Peso ao nascer (óbitos fetais e em menores de 28 dias)

Para efeito de crítica e correção dos dados, as variáveis de cada óbito são classificadas em:

Indispensáveis - ano do óbito e tipo de óbito. Não são computados os óbitos que não tenham a definição dessas duas variáveis;

Essenciais - sexo, idade, município de ocorrência, município de residência, causa básica e tipo de violência. Essas variáveis, que são as mais frequentemente usadas em estudos de mortalidade, têm prioridade para crítica e correção de dados;

Secundárias - todas as demais. Embora todas tenham grande importância, essas, pelo menor uso e maior dificuldade na obtenção de níveis satisfatórios de qualidade, não têm caráter prioritário para crítica e correção. Este desdobramento é feito especialmente nas áreas de implantação parcial ou mais recente do Sistema, que correspondem àquelas de menor grau de desenvolvimento social, onde o contingente de “ignorados” é significativamente importante.