Uso da análise de séries temporais em estudos epidemiológicos

POR: CLEVIA BENTO DE OLIVEIRA

SUMÁRIO

INTRODUÇÃO

Esta apresentação tem como objetivo descrever aspectos conceituais e a análise de medidas de interesse de séries temporais para estudos epidemiológicos.

Medidas de interesse para análises epidemiológicas

Tendência

é definida como um padrão de crescimento/descrecimento da variável em um certo período de tempo.

Associação

a interpretação das tendências em séries temporais com outras informações sobre o fenômeno em questão

Sazonalidade

quando os fenômenos que ocorrem durante o tempo se repetem a cada período idêntico de tempo

MATERIAL E MÉTODOS

Tendência

Quando estudamos séries temporais em estudos epidemiológicos, um primeiro elemento da análise focaliza a tendência da medida.
Dependendo da análise ela pode apresentar trechos com diferentes tendências. Podemos ver esse comportamento na figura seguinte que mostra uma série temporal da mortalidade infantil na cidade de São Paulo-SP.

Podemos ver que em alguns intervalos de anos o coeficiente de mortalidade infantil possui tendência de crescimento e em outros tendência de queda

Fonte: Antunes[1]

Figura 3 – Coeficiente de mortalidade infantil na cidade de São Paulo, estado de São Paulo. Brasil,1900-1994

Associação

Associando a tendência de uma série com informações adicionais podemos ampliar a interpretação dos resultados e entender melhor o comportamento do estudo. Essas informações adicionais podem ser qualitativas, auxiliando a interpretar motivos para o aumento, diminuição ou persistência dos valores de uma medida de interesse para a saúde, ou podem ser quantitativas, dando ensejo à aplicação de técnicas estatísticas para estimar sua associação com a série temporal que se tenta explicar.

Nesta figura vemos o coeficiente de mortalidade em São Paulo entre os anos de 1900 a 1990. Visualmente podemos notar que aproximadamente em 1918 houve um grande pico de mortalidade. Sabendo que em 1918 foi o ano em que houve um grave surto de gripe espanhola, podemos associar que esse aumento significativo de mortes teve ligação com o surto de gripe

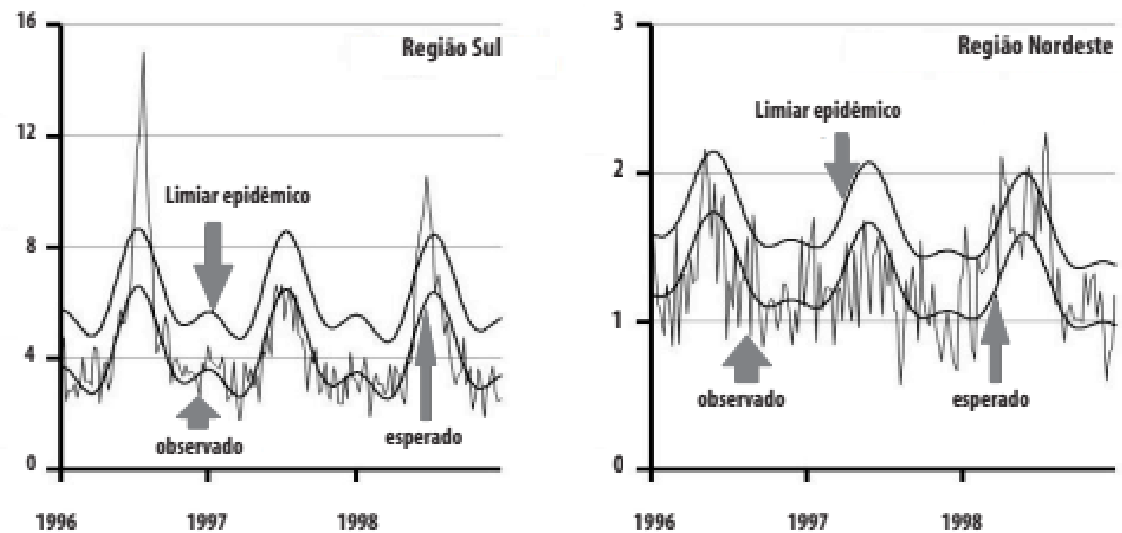
Fonte: Antunes[3]

Figura 2 – Coeficiente de mortalidade geral na cidade de São Paulo, estado de São Paulo. Brasil, 1900 a 1994

Sazonalidade

A percepção de que os fenômenos de interesse para a saúde também podem apresentar repetições organizadas no tempo, ou seja, de que há ritmo a ser reconhecido na análise de séries temporais, é muito importante para a epidemiologia.
Um exemplo são as variações sazonais e cíclicas que afetam a medida de muitas doenças.

Podemos ver que ambas as séries apresentam tendência estacionária com variação sazonal. Ao se observar a diferença de escala no eixo vertical, percebe-se que a mortalidade é mais elevada na região Sul que no Nordeste do país.



Fonte: Oliveira e colaboradores [5]

Figura 4 – Mortalidade semanal de idosos (65 anos ou mais) por influenza e pneumonia nas regiões Sul e Nordeste do país. Brasil, 1996 a 1998

VARIAÇÃO ALEATÓRIA

é o elemento de variação das medidas epidemiológicas organizadas no tempo que propiciaria perturbação na percepção de tendências, associações e variações sazonal

essa perturbação é causada pela variação aleatória da medida, a qual se manifesta visualmente na forma de rugosidade nas linhas dos gráficos de séries temporais.

A variação aleatória em séries temporais é definida como flutuações irregulares e erráticas – que não são importantes em si mesmas –, causadas por fatores do acaso, impossíveis de serem antecipados, detectados, identificados ou eliminados.

ASPECTOS METODOLÓGICOS DA ANÁLISE DE SÉRIES TEMPORAIS

ESTIMAR TENDÊNCIAS

Séries temporais podem apresentar tendência crescente, decrescente ou estacionária, e até tendências diferentes em trechos sequenciais. Para estimar a tendência, funções matemáticas são ajustadas aos pontos observados, seja para a série temporal como um todo, seja para o segmento em foco. A quantificação da tendência visa permitir a comparação entre diferentes séries temporais.

UTILIZANDO O MÉTODO ORIGINALMENTE PROPOSTO POR ANTUNES E WALDMAN

Sendo Y a escala dos valores da série temporal e X a escala de tempo, a reta de melhor ajuste entre os pontos da série temporal, ou um trecho para o qual se pretende estimar a tendência, é definida pela seguinte equação:

Fórmula 1: equação de regressão linear com componente de tendência

$$Y = b_0 + b_1 X$$

Na Fórmula 1, o valor b0 corresponde à interseção entre a reta e o eixo vertical; o valor b1 corresponde à inclinação da reta. Para cada mudança de uma unidade na escala de X, o valor de Y é acrescido de b1 unidades. Porém, o valor bruto da variação é expresso em unidades, o que dificulta sua comparação com fatores medidos em escalas diferentes.

Fórmula 1: equação de regressão linear com componente de tendência

$$Y = b_0 + b_1 X$$

Nesse sentido,
é preferível estimar a taxa percentual de variação.
Para mensurar a taxa de variação da reta que ajusta os pontos da série temporal, aplica-se a transformação logarítmica dos valores de Y, o que propicia vantagens adicionais para a análise de regressão linear, como a redução da heterogeneidade de variância dos resíduos da análise de regressão.

$$\boxed{\begin{gathered}\log Y_i = b_0 + b_1 X_i \\ \text{e} \\ \log Y_{i+1} = b_0 + b_1 X_{i+1}\end{gathered}}$$

Diferenciando os termos destas duas equações:

$$\log Y_{i+1} - \log Y_i = b_0 + b_1 X_{i+1} - b_0 - b_1 X_i = b_1 (X_{i+1} - X_i)$$

Como Xi+1 e Xi são períodos (dias, meses, anos) subsequentes, sua diferença é sempre igual a um. E por propriedades da álgebra de logaritmos:

$$\log Y_{i+1} - \log Y_i = \log (Y_{i+1} / Y_i) = b_1$$

ou

$$Y_{i+1} / Y_i = 10^{b1}$$

Subtraindo 1 de ambos os lados da equação:

$$Y_{i+1}/Y_i - 1 = -1 + 10^{b1}$$

ou

$$(Y_{i+1} - Y_i)/Y_i = -1 + 10^{b1}$$

Contudo, (Yi+1–Yi)/Yi é justamente a taxa de mudança, pois foi dimensionada para um período genérico ‘i’. Basta, então, estimar o valor de b1 para inferir a taxa de mudança anual (mensal ou diária) da medida de interesse.

Como b1 é estimado por regressão linear, deve-se aplicar o intervalo de confiança desse coeficiente na Fórmula 2, para se calcular o intervalo de confiança da medida. Com isso, teremos uma expressão sintética para a estimação quantitativa da tendência.

Fórmula 2: Tendência ou mudança percentual

$$APC=[-1+10^{b1}]*100\%$$

$$IC_{95\%}=[-1+10^{b1min.}]*100\%;[-1+10^{b1máx.}]*100\%$$

MODELAR SAZONALIDADE

As variações sazonais podem ser aferidas por medidas diárias, semanais ou mensais. É preciso notar que há alguma irregularidade na forma de registro do tempo, nas avaliações de sazonalidade. Se as medidas forem diárias, há irregularidade nos anos bissextos. Para medidas mensais, os meses do calendário podem ser enumerados sequencialmente, mas não têm o mesmo número de dias. Para dados semanais, utiliza-se a definição de semanas epidemiológicas para contornar a divisão não exata do ano em semanas.

Para identificar se há variação sazonal, é preciso decompor a série temporal, isolando o componente e verificando se atende à hipótese de significância estatística. Essa decomposição usa a equação de regressão linear com dois componentes, um para indicar tendência e outro para sazonalidade.

Fórmula 3: equação de regressão linear com componente sazonal

$$Y_i = b_0 + b_1 * X_i + b_2 * sen(2\pi X_i / L) + b_3 * cos(2\pi X_i / L)$$

Na Fórmula 3, Yi é a medida da série temporal para cada momento genérico 'i' e Xi é a numeração sequencial dos momentos de tomada da medida (dia, semana, mês), π é a conhecida constante 3,141592654... e L é uma constante relativa à forma da medida: 12 para medidas mensais, 52 para semanais e 365 para diárias. O coeficiente b0 é o intercepto da equação de regressão, b1 é o estimador da tendência, e b2 e b3 são os coeficientes que modelam a sazonalidade.

Quando a série temporal usar dados diários ou semanais, o maior número de pontos permite maior poder estatístico para a análise de regressão linear. Pode-se, então, incluir um segundo termo harmônico para sazonalidade. A Fórmula 4 apresenta a nova equação, sendo b4 e b5 os coeficientes que definem o segundo harmônico.

Fórmula 4: equação de regressão linear com componente sazonal segundos harmônicos

$$Y_i = b_0 + b_1 * X_i + b_2 * sen(2\pi X_i/L) + b_3 * cos(2\pi X_i/L) + b_4 * sen(4\pi X_i/L) + b_5 * cos(4\pi X_i/L)$$

O limiar epidêmico indicado na Figura 4 pode ser calculado como função (Fórmula 5) de Yi. O limiar epidêmico (Zi) serve para compor o diagrama de controle da série temporal, e para reconhecer a emergência de surtos sempre que os valores observados superarem o limiar epidêmico por duas ou mais semanas consecutivas.

Fórmula 5: limiar epidêmico do diagrama de controle da série temporal

$$Z_i = Y_i + 1{,}645 * DP(Y_i)$$

SÉRIES TEMPORAIS INTERROMPIDAS E ANÁLISE DE REGRESSÃO SEGMENTADA

A análise de séries temporais interrompidas foi considerada o mais efetivo recurso não experimental para avaliar o efeito longitudinal de intervenções.

Entretanto, sua aplicação não se restringe a isso, servindo também para testar hipóteses sobre fatores que

modificam o comportamento no tempo das medidas de interesse para a saúde.

Essa análise pressupõe segmentar, algébrica e graficamente, a série temporal.

Tomando a Fórmula 1 como referência, a Fórmula 6 sintetiza a análise de regressão segmentada com dois e três segmentos. É possível incluir mais segmentos, se houver número suficiente de observações (ao menos oito pontos para cada segmento). Também é possível incluir os termos da avaliação de sazonalidade (fórmulas 3 e 4).

Fórmula 6: análise de regressão segmentada

dois segmentos: $Y_i = b_0 + b_1{*}\text{tempo} + b_2{*}\text{degrau} + b_3{*}\text{rampa}$

três segmentos: $Y_i = b_0 + b_1{*}\text{tempo} + b_2{*}\text{degrau}_1 + b_3{*}\text{rampa}_1 + b_4{*}\text{degrau}_2 + b_5{*}\text{rampa}_2$

A variável “degrau” é construída de modo dicotômico, com 0 (zero) nos pontos anteriores à intervenção e 1 na vigência da intervenção, isto é, após o início do segmento. A variável ‘rampa’ mede o tempo após a intervenção, sendo construída com 0 (zero) nos pontos que antecedem a intervenção e valores sequenciais – 1,2,3… – após o início do segmento.

CONSIDERAÇÕES FINAIS

Antever o futuro é uma primeira e óbvia aplicação da análise de séries temporais.

Para variáveis quantitativas, é sempre possível conhecer os valores passados, não os futuros.

Uma segunda aplicação do instrumental de análise das séries temporais refere-se à previsão do passado, o que, embora pareça estranho e desnecessário, tem várias aplicações epidemiológicas.

Há, ainda, uma terceira forma de antever o futuro, a partir da análise de séries temporais. Em vez de focalizar os valores que as variáveis quantitativas assumirão no futuro, essa terceira forma refere-se ao reconhecimento dos padrões de variação da medida.