

COMPARAÇÃO DE MODELOS DE PREVISÃO APLICADOS A UMA SÉRIE DE DEMANDA DA VACINA CONTRA HEPATITE B

G. P. N. Duque*, M. L. O. Novaes*, R. M. V. R. Almeida*

*Programa de Engenharia Biomédica/ Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brasil
graciela.duque@yahoo.com.br

Resumo: A vacina contra Hepatite B confere imunidade aos grupos de riscos expostos à doença, sendo amplamente utilizada no país. O objetivo desta pesquisa foi comparar dois modelos de séries temporais (Autorregressivo Integrado de Médias Móveis, ARIMA e Suavização exponencial, SE) em horizontes de previsão distintos (um, três e 12 meses) em uma série de doses aplicadas dessa vacina (155 observações). Para o estudo, foram construídas dez subséries de 60 observações cada, as quais foram analisadas relativamente a seus erros médios (MAPE) e coeficientes de determinação (R^2). Os resultados indicaram pouca variação nos valores médios do MAPE e R^2 em relação aos horizontes de previsão utilizados, com um comportamento melhor do modelo ARIMA em relação ao SE. Conclui-se que o modelo ARIMA pode ser empregado na previsão de séries temporais de dados de vacinação da Hepatite B nos horizontes pesquisados.

Palavras-chave: Vacina, hepatite B, ARIMA, Suavização Exponencial.

Abstract: *The Hepatitis B vaccine intends to provide immunity to risk groups exposed to the disease, and is widely used in Brazil. The purpose of this study was to compare two models of time series analysis (Auto-Regressive Integrated Moving Averages, ARIMA and Exponential Smoothing, ES) in different forecast horizons (one, three and twelve months) in a series of doses of the vaccine (155 observations). For the study, ten sub-series were built, each with 60 observations. These sub-series were analyzed relatively to their mean average errors (MAPE) and their coefficient of determination (R^2). Results indicated little variation on the MAPE and R^2 average values according to the forecast horizons, with an overall better behavior of ARIMA relatively to ES. In conclusion, ARIMA can be used for predicting Hepatitis B vaccine series in the analyzed forecast horizons.*

Keywords: *Vaccines, Hepatitis B, ARIMA, Exponential smoothing*

Introdução

Séries temporais permitem a aplicação e análise de modelos a muitos tipos de variáveis na área de saúde [1, 2]. Um exemplo é a análise de dados vacinais, que são disponibilizados no Brasil através do Ministério da Saúde e do Programa Nacional de Imunizações (PNI),

pela Internet e pelas Secretarias de Saúde municipais [3], e cuja modelagem é extremamente importante para a definição de estratégias vacinais adequadas e para a avaliação do impacto de campanhas de vacinação. Esses modelos podem também auxiliar na estimativa do número de vacinas a serem utilizadas, reduzindo o desperdício desses imunobiológicos [1].

No entanto, apesar de existirem na literatura métodos de previsão em séries temporais consagrados, suas aplicações ainda suscitam discussões. Por exemplo, existem dúvidas sobre qual o horizonte de previsão mais adequado para um determinado modelo; qual o método, entre os disponíveis, capaz de fornecer resultados com menor erro para certo horizonte; que magnitude de erros pode-se considerar em uma aplicação e qual o número de observações mínimo a ser utilizado em um dado método [1, 2].

É de supor-se que intervalos de previsão menores (por exemplo, um mês) forneçam resultados mais confiáveis [1]. No entanto, o desenvolvimento de parâmetros melhor definidos é fundamental para que estudos epidemiológicos possam ser realizados com maior confiabilidade e robustez [4]. Nesse contexto, a presente pesquisa objetivou comparar dois modelos de séries temporais (Autorregressivo Integrado de Médias Móveis (ARIMA) e Suavização exponencial (SE)) relativamente a seus erros em horizontes de previsão distintos (um, três e 12 meses). Uma série temporal de doses aplicadas da vacina contra Hepatite B foi utilizada para a análise.

Materiais e métodos

Coleta de dados - Para a realização deste estudo foram utilizadas observações mensais de doses aplicadas da vacina contra Hepatite B (HepB) em Juiz de Fora - MG, entre os anos 1999 e 2011, totalizando 155 observações. Valores referentes à cobertura dessa vacina (quantidade de doses aplicadas sobre a população-alvo) foram também obtidos, para fins descritivos.

Nesse município, a HepB é aplicada como parte do Programa Nacional de Imunizações (PNI) em 46 salas de vacinação, sendo seus dados disponibilizados pela Secretaria Municipal de Saúde de Juiz de Fora em registros mensais contendo os números de doses recebidas, utilizadas, perdidas, remanejadas e existentes no estoque (saldo).

Modelos ARIMA e de Suavização Exponencial -

Para a verificação da melhor estratégia de previsão, utilizaram-se horizontes de previsão (intervalo entre a última observação real e a observação que será estimada [5]) de um, três e 12 meses. Dois métodos foram aplicados: o modelo ARIMA (*Autoregressive Integrated Moving Average*) clássico e o modelo de Suavização Exponencial (SE).

O ARIMA, modelo paramétrico do domínio do tempo, consiste na combinação de três “filtros”: o autorregressivo (AR), o “filtro” de integração (I), e o de médias móveis (MA) [6]. O modelo de médias móveis baseia-se na média de observações passadas, dada pela equação [6]:

$$Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (1)$$

Onde “ μ ” representa a média, “ ε ” são os erros aleatórios não explicados pelo modelo e “ θ ” corresponde ao parâmetro do modelo MA de ordem q . O valor de q significa o número de termos que atuam no período de cálculo da média móvel [7].

Já o modelo Autorregressivo (AR), descrito pela equação (2), especifica que a variável de saída depende linearmente de seus valores anteriores [8].

$$Z_t = \varphi_1 Z_{t-1} + \varphi_2 Z_{t-2} + \dots + \varphi_p Z_{t-p} + a_t \quad (2)$$

Onde Z_t corresponde à observação da série temporal no tempo t , φ representa o parâmetro do modelo AR de ordem p , e a_t representa o erro de eventos aleatórios que não podem ser explicados pelo modelo [9]. O valor de φ é determinado por meio da resolução da equação característica (3).

$$1 - \varphi_1 x - \varphi_2 x^2 - \dots - \varphi_p x^p = 0 \quad (3)$$

Quando uma série é formada pelas partes autorregressivas (AR) e médias móveis (MA), tem-se o modelo mais geral de séries temporais, o ARMA [10]. Z_t será um modelo “autorregressivo de médias móveis” de ordem “ p ” e “ q ” e parâmetros φ e θ , obtendo-se a equação (4).

$$Z_t = \varphi_1 Z_{t-1} + \varphi_2 Z_{t-2} + \dots + \varphi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (4)$$

Adicionalmente, uma série temporal que necessita ser diferenciada para tornar-se estacionária é chamada de versão integrada desta série [11]. Assim, uma série que passou pelo processo de diferenciação para tornar-se estacionária ($W_t = \Delta dZ_t$) é representada por um modelo ARMA, chamado de ARIMA (termo “I” referente à integração) e definido pela equação (5).

$$\varphi(X)W_t = \theta(X)a_t \quad (5)$$

A equação (5) é um modelo ARIMA de ordem “ p ”, “ d ” e “ q ”, que relaciona a ordem “ p ” ao modelo autorregressivo, a ordem “ d ” à diferenciação da série e a ordem “ q ” ao modelo médias móveis.

Já modelos de suavização exponencial são frequentemente utilizados na previsão de dados na área de economia. Basicamente, sua ideia é prever os valores em uma série pela atribuição de pesos aos seus valores passados, sendo que os valores mais recentes recebem pesos maiores que os mais distantes [11]. Assim, pode ser considerado como um modelo média móvel, no qual são dados pesos maiores às observações mais recentes [12]. Este esquema é exemplificado pela equação (6), na qual são realizadas previsões a partir de “demandas reais” e “demandas prévias”.

$$NP = \alpha(\text{demanda real}) + (1-\alpha)(\text{demanda prévia}) \quad (6)$$

Em (6), NP é a nova previsão e α é o fator de ponderação, com valores $0 < \alpha < 1$. Logo, quanto maior o valor de α , mais elevado é o peso colocado nos níveis mais recentes de demanda. Esse tipo de modelo responde mais rapidamente às mudanças na série temporal [12].

Avaliação dos modelos - Duas estatísticas foram utilizadas para a avaliação dos modelos: O MAPE e o valor de R^2 . O MAPE (*Mean Absolute Percentage Error*) é definido como a média da porcentagem do erro entre a série real e a série prevista [13], sendo representado pela equação (7).

$$M = [\sum |Z_t - F_t| / Z_t] / n \quad (7)$$

Onde Z_t são os valores observados em uma série em função do tempo t , e F_t são os valores previstos pelo modelo de séries temporais utilizado, também como função do tempo.

A estatística R^2 é muito utilizada para a avaliação do ajuste de modelos lineares, sendo também chamada de “coeficiente de determinação” ou “percentual da variância explicada”. Essa estatística é definida como “o percentual da variação de x que ocorre concomitantemente à variação de y ”, estando o valor de R^2 situado entre zero e um, sendo que, quanto mais próximo o valor de um, melhor o modelo [14].

Para a comparação das previsões foram extraídas da série original dez subséries, com 60 observações cada. Para isso, dez números inteiros no intervalo 1-155 (equivalentes aos pontos de observação da série original) foram sorteados e colocados em ordem crescente; e a partir desses pontos, foram separados os grupos de 60 observações. Esse número foi escolhido com o auxílio de indicações da literatura [15], sendo também definido um *hold out* (5% da série, ou seja, as três últimas observações) para a estimativa *out-of-sample* das estatísticas de comparação para cada subsérie. Foram excluídos os números correspondentes às observações de 96 a 155, por não permitirem amostras do tamanho desejado.

A partir das subséries definidas acima, os modelos ARIMA e SE foram utilizados para a investigação de três horizontes de previsão: um, três e 12 meses (ou seja, intervalos entre a última observação real e a observação estimada de um, três e 12 meses). Os

resultados são apresentados para cada série e também na forma de média das modelagens realizadas, e o programa utilizado foi o *Forecast Pro for Windows Version 3.50 Extended Edition*, um software de previsão para séries temporais [11].

Resultados

A Tabela 1 apresenta os valores de doses aplicadas e a cobertura de HepB no município estudado em cada ano analisado. A Tabela 2 apresenta os valores do MAPE e do R² para cada *hold-out* das dez subséries.

Em termos gerais, observaram-se resultados melhores para o modelo ARIMA em relação ao R² e MAPE nos horizontes de previsão de um e três meses (valores médios de 0,21 e 0,5 respectivamente), relativamente aos valores SE (valores médios aproximados de 0,24 e 0,32 respectivamente). Notou-se também pouca variação nos valores do R² e MAPE ao longo dos horizontes.

Tabela 1: Doses aplicadas e cobertura percentual da vacina HepB, por ano, município de Juiz de Fora- MG.

Ano	Doses Aplicadas	Cobertura (%)
1999	45.279	-
2000	36.389	-
2001	74.119	-
2002	88.594	-
2003	71.919	-
2004	43.224	89,64
2005	39.645	96,06
2006	43.876	94,20
2007	40.296	88,94
2008	39.560	89,97
2009	36.912	90,53
2010	42.587	82,95
2011	21.623	87,72

Discussão

A HepB é aplicada em um esquema de três doses em toda a população, priorizando-se o grupo de risco de hemofílicos, usuários de hemodiálise/transfusões de sangue, profissionais de saúde, prostitutas, reclusos e homossexuais masculinos [16]. Observa-se na Tabela 1 que esta vacina tem uma grande cobertura na população priorizada, com uma alta taxa de doses aplicadas, o que fornece uma série temporal robusta. No presente estudo, a série temporal de HepB foi utilizada para verificar a capacidade preditora dos modelos ARIMA e SE em subséries construídas a partir dessa série original.

Tabela 2: Média e desvio-padrão para MAPE e R² (*hold-outs*) em modelos ARIMA e SE, horizontes de previsão de um, três e 12 meses, subséries construídas a partir de uma série de HepB, Juiz de Fora- MG.

Modelo	Observações	Horizonte de Previsão		
		1 mês	3 meses	12 meses
		MAPE/ R ²	MAPE/ R ²	MAPE/R ²
ARIMA	Fev/99- Jan/04	0,27/0,51	0,26/0,53	0,26/0,55
	Nov/99-Out/04	0,26/0,50	0,27/0,49	0,27/0,51
	Jan/00- Dez/04	0,25/0,51	0,25/0,51	0,28/0,53
	Set/00-Ago/05	0,22/0,46	0,23/0,45	0,26/0,47
	Jul/01-Jun/06	0,20/0,66	0,20/0,67	0,21/0,59
	Out/01-Set/06	0,14/0,77	0,20/0,65	0,20/0,62
	Fev/02-Jan/07	0,21/0,57	0,20/0,59	0,01/0,87
	Jun/03-Mai/08	0,17/0,19	0,17/0,19	0,16/0,25
	Dez/03-Nov/08	0,14/0,12	0,14/0,12	0,16/0,02
	Jan/07-Dez/11	0,23/0,27	0,24/0,22	0,18/0,02
	<i>Média</i>	<i>0,21/0,48</i>	<i>0,21/0,49</i>	<i>0,19/0,52</i>
	<i>Desvio Padrão</i>	<i>0,04/0,02</i>	<i>0,04/0,19</i>	<i>0,05/0,26</i>
	SE	Fev/99-Jan/04	0,27/0,31	0,26/0,31
Nov/99-Out/04		0,26/0,31	0,27/0,30	0,27/0,28
Jan/00-Dez/04		0,25/0,32	0,26/0,32	0,28/0,28
Set/00-Ago/05		0,42/0,40	0,25/0,31	0,27/0,26
Jul/01-Jun/06		0,25/0,68	0,24/0,70	0,25/0,66
Out/01-Set/06		0,19/0,66	0,17/0,69	0,18/0,67
Fev/02-Jan/07		0,19/0,62	0,19/0,61	0,17/0,68
Jun/03-Mai/08		0,17/0,37	0,16/0,37	0,16/0,37
Dez/03-Nov/08		0,16/0,13	0,15/0,17	0,15/0,22
Jan/07-Dez/11		0,25/0,22	0,25/0,18	0,17/0,08
<i>Média</i>		<i>0,25/0,34</i>	<i>0,24/0,31</i>	<i>0,21/0,32</i>
<i>Desvio Padrão</i>		<i>0,07/0,18</i>	<i>0,04/0,19</i>	<i>0,05/0,21</i>

Em termos de comparação inter-modelos, pôde-se constatar que o modelo ARIMA apresentou um melhor comportamento em relação ao SE nas estatísticas e horizontes de previsão investigados. Assim, em relação aos horizontes “curtos” (um e três meses), observaram-se resultados melhores para o modelo ARIMA, tanto para o valor de R² quanto para o MAPE, e, no horizonte mais distante (12 meses), essa superioridade ainda pôde ser observada relativamente ao R². Por exemplo, os valores do coeficiente de determinação para o modelo ARIMA foram tipicamente R²=0,5, contra valores na faixa 0,21-0,25 para o SE. Já no que diz respeito aos horizontes de previsão, não foi observada grande variabilidade entre os horizontes pesquisados, com valores ARIMA para MAPE e R² tipicamente próximos a 0,2/0,5.

Na literatura internacional, outro estudo avaliou horizontes de previsão, modelando o número de casos de dengue na cidade do Rio de Janeiro, RJ entre os anos de 1997 e 2004. Esse estudo verificou, por meio de função de autocorrelação da série (ACF), que a melhor previsão ocorreu no horizonte de um mês (relativamente a um horizonte de 12 meses) [2]. No entanto, como mencionado, no presente trabalho não foram identificadas diferenças relevantes entre os horizontes de previsão, talvez devido ao menor erro (sub-registro) presente na série analisada aqui. Essa pouca variabilidade longitudinal é conceitualmente importante em termos de modelagem de séries epidemiológicas, dado que, caso a validade de uma previsão só pudesse

ser considerada como confiável para um intervalo pequeno, ações de planejamento/intervenção com base nesses modelos seriam de pouca utilidade prática [1].

Ainda outro trabalho, com dados populacionais dos Estados Unidos entre 1900 e 1980, observou que os erros de previsão aumentam com o aumento do horizonte de previsão a partir da aplicação de modelos de extrapolação (linear e exponencial) [6]. Porém, como mencionado neste estudo, não foi identificado aumento dos erros de previsão em decorrência do aumento dos horizontes de previsão (por exemplo, no modelo SE, observou-se MAPE médio de 0,25 para um mês e de 0,21 para 12 meses).

Conclusão

Os resultados do presente estudo sugerem que o modelo ARIMA apresentou um melhor ajuste para a previsão de séries temporais com dados de vacinação em Hepatite B, quando comparado com o modelo de SE, e também que não é necessária a utilização de um horizonte de previsão curto (um mês), como, por vezes, sugerido na literatura. Estudos de maior escopo devem ser realizados para a definição de outros parâmetros, como o comportamento de modelos não-lineares, horizontes de previsão maiores e uma análise mais cuidadosa das características inter-modelos.

Agradecimentos

Os autores agradecem ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à CAPES pelo auxílio financeiro ao presente trabalho. Agradecemos também ao Departamento de Vigilância Epidemiológica da Secretaria Municipal de Saúde de Juiz de Fora - MG pela disponibilização dos dados referentes às doses aplicadas de Hepatite B.

Referências

- [1] Novaes MLO, Almeida RMVR, Bastos RR, Figueiredo BB, Centellas CDR, Rangel JMC, Silva MHS, Flores PA. Caracterização das perdas da vacina contra Rotavírus e de seus custos associados. In: XXIII Congresso Brasileiro em Engenharia Biomédica- XXIII CBEB; 2012. Out 02-05; Recife, Brasil. 2012.
- [2] Luz PM, Mendes BVM, Codeço CT, Struchiner CJ, Galvani AP. Time series analysis of dengue incidence in Rio de Janeiro, Brazil. *The American Society of Tropical Medicine and Hygiene*. 2008; 79(6):933-939. pni.datasus.gov.br/index.asp
- [3] Ministério da Saúde. Sistema de informações do PNI. DATASUS [internet]. 2014 Set; Disponível: <http://goo.gl/xf2Kko>
- [4] Steelwagen EA, Goodrich RL. Forecast pro-
statistical reference manual. 5ª ed. Business Forecast Systems: Belmont; 2008.
- [5] Chatfield C. Time series forecasting. Chapman and Hall: London; 2000.
- [6] Smith SK, Sincich T. An empirical analysis of the effect of length of forecast horizon on population forecast errors. *Demography*. 1991; 28 (2):261-74.
- [7] Mattos RS. Modelos ARIMA- Metodologia de Box & Jenkins. Faculdade de Economia/ Universidade Federal de Juiz de Fora: Juiz de Fora; 2012.
- [8] Fava VL. Análise de séries de tempo. Em: Atlas, editor. Manual de Econometria. São Paulo: MAS Vasconcelos e D Alves; 2000.
- [9] Mueller A. Uma aplicação de redes neurais artificiais na previsão do mercado acionário [dissertação]. Florianópolis: Universidade Federal de Santa Catarina; 1996.
- [10] Bezerra MIS. Análise de séries temporais. Universidade Estadual Paulista: São Paulo; 2006.
- [11] Nau RF. Introduction to ARIMA. Decision 411-
Statistical Forecasting [internet]. 2013 Nov [citado novembro 2013]; Disponível: tinyurl.com/l5523j4
- [12] Ballou RH. Gerenciamento da cadeia de suprimentos. 4ª ed. Bookman: Porto Alegre; 2001.
- [13] Steelwagen EA, Goodrich RL. Forecast pro-
statistical reference manual. 7ª ed. Business Forecast Systems: Belmont; 2011.
- [14] Moore DS, Notz W, Fligner M. The basic practice of statistics, 12ª ed. Macmillan, New York, 2012.
- [15] Oracle. Crystal ball. Statistical Guide [internet]. 2013 Ago [citado agosto 2013]; Disponível: docs.oracle.com/cd/E12825
- [16] Ministério da Saúde: Fundação Nacional da Saúde. Manual de normas de vacinação. 3ª ed. Brasília; 2001.