International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org ||Volume 7 Issue 4 Ver. V|| April 2018 || PP 61-65

Model selection for forecasting growth rate of Hepatitis B patients

B.Sarojamma[1], B.Hari mallikarjuna Reddy*2, S.V.Subramanyam[3]
[1]*Assistant Professor, Department of Statistics, S.V. University, Tirupathi.A.P.*
[2,3]*Research Scholar, Department of Statistics, S.V. University, Tirupathi.A.P.*
**Corresponding auther: B.Sarojamma*

***Abstract**: Hepatitis B is a viral infection that affecting liver function throughout the world around two billion people are infected by hepatitis B virus. There are many types of hepatitis's they are hepatitis B, hepatitis A, hepatitis C, hepatitis D, hepatitis E, autoimmune hepatitis, chronic hepatitis, etc.*
In this paper we proposed three time series models for time series data of hepatitis B for forecasting growth rate of Hepatitis patients. The models are tested for its accuracy by using R^2 criteria and Root Mean Square Root Error (RMSE).
***Keywords:** Hepatitis-B, time series Model-1, time series Model -2, time series Model-3, R^2 criteria, Root Mean Square Error (RMSE).*

Date of Submission: 10-04-2018 Date of acceptance: 26-04-2018

I. Introduction

Hepatitis B is a viral infection that attracts on liver function. Hepatitis is a chronic disease. All over the word two billion people are infected by hepatitis B virus. These virus will spread through blood to blood contact as like HIV. It will spread through sexual activity and injecting drug users. Hepatitis B is transmitted during young adult hood through sex. Hepatitis B virus is more infection than HIV. It is 50 to 100 times more infection than HIV. There are so many types of Hepatitis virus they are Alcoholic Hepatitis, Toxic and drug-induced Hepatitis, ischemaic hepatitis, giant cell hepatitis, viral hepatitis, hepatitis A, hepatitis B, hepatitis C, Hepatitis D, hepatitis E, autoimmune hepatitis chronic hepatitis, etc.

The entire paper divided in to four sections, in first section we discussed brief introduction about the study and the rest of sections includes detailed proposed models, empirical investigation and conclusions respectively.

II. Models for forecasting growth rate.

The model-I

Let $y_t = \dfrac{\beta_0}{\beta_1^t}$

Where y_t = number of Hepatitis B patients and t = time (no of years), and where β_0 and β_1 areknown as the constants of the model .

Taking log on both sides of the above equation then we get modified equation of form

$\therefore$ Y_i=A-B.t_i (2.1.1)

By solving the above equation using MLE method we produce below fitted equation

$\therefore$ The fitted regression model is $\hat{y}_t = \frac{\hat{\beta}_0}{\hat{\beta}_1^t}$

Model: II

Another proposed model is as below

$y_t = \beta_0 + \beta_1^{t_i}$

Where y_t = number of Hepatitis B patients and t = time (no of years), and where β_0 and β_1 known as the constants of the model are respectively the intercept and slope coefficients.

Taking log on both sides then the above equation we get

$$\text{Log}(y_t) = \log(\beta_0 + \beta_1^{t_i})$$

$$\log y_t = \log \beta_0 + \log \beta_1 . t_i$$

Let Log(y_t) = Y , log(β_0) = A, $\log\beta_1$ = B

$$\therefore \quad Y_i = A+B.t_i \qquad (2.2.1)$$

By solving the above equation using MLE method we produce below fitted equation

$\therefore$ the fitted regression model is $\hat{y}_t = \hat{\beta}_0 + \hat{\beta}_1^t$

Model: III

$$y_t = \frac{\beta_0}{\beta_1^t + \beta_2^{t^2}}$$

Where y_t = number of Hepatitis B patients t = time (no of years), β_0 and β_1 β_2 andareknown as the constants of the model are respectively the intercept and slope coefficients.

Taking log on both sides of the above equation we get

$$\text{Log}(y_t) = \log(\beta_0)\text{-}\log(\beta_1^t + \beta_2^{t^2})$$

$$\log y_t = \log\beta_0 \ -\log\beta_1.t - \log\beta_2.t^2$$

Let Log(y_t) = Y , log(β_0) = A, $\log\beta_1$ = B, log(β_2) = C

$$\therefore \quad Y=A\text{-}B.t\text{-}C.t^2 \qquad (2.3.1)$$

By solving the above equation using MLE method we produce below fitted equation

$\therefore$ The fitted regression model is $\hat{y}_t = \frac{\hat{\beta}_0}{\hat{\beta}_1^t + \hat{\beta}_2^{t^2}}$

Generalized model:

$$y_t = \frac{\beta_0}{\beta_1^t + \beta_2^{t^2} + \ldots\ldots + \beta_2^{t^n}}$$

Where y_t = number of Hepatitis B patients t = time (no of years), and where β_0, and β_1 β_2..... and areknown as the constants of the model, are, respectively, the intercept and slope coefficients.

Taking log on both sides then we get

$$\text{Log}(y_t)=\log(\beta_0)\text{-}\log(\beta_1^t + \beta_2^{t^2} +\ldots)$$

$$\log y_t = \log\beta_0 \ -\log\beta_1.t - \log\beta_2.t^2$$

Let Log(y_t) =Y , log(β_0) = A_0, $\log\beta_1$ = A_1, log(β_2) = A_2 , ...log(β_n) = A_n

$$\therefore \quad Y=A_0\text{-}A_1.t\text{-}A_2.t^2\text{-}\ldots\text{-}A_n t^n \qquad (2.4.1)$$

By solving the above equation using MLE method we produce below fitted equation

Therefore the fitted regression model is $\hat{y}_t = \frac{\hat{\beta}_0}{\hat{\beta}_1^t + \hat{\beta}_2^{t^2} + \cdots\ldots\ldots\ldots + \hat{\beta}_2^{t^n}}$

III. Empirical Investigations:

By utilizing proposed models in section 2, for the data we observed the following results.

Model I: the first model for forecasting growth rate is $y_t = \frac{\beta_0}{\beta_1^t}$

Where y_t = number of Hepatitis B patients and t = time (no of years), and where β_0 and β_1 are called the parameters of the model and generally termed as Intercept and slope terms.

Table 3.1.1.

Year	**ti=(year-2008.5)**	**Observed patients**	**yi=log (observed deaths)**	**yi*ti**	**ti^2**
2005	-3.5	3860	3.5865	-12.5530	12.25
2006	-2.5	4400	3.6434	-9.1086	6.25
2007	-1.5	4668	3.6691	-5.5036	2.25
2008	-0.5	3826	3.5827	-1.7914	0.25
2009	0.5	3203	3.5055	1.7528	0.25
2010	1.5	3460	3.5390	5.3086	2.25
2011	2.5	3995	3.6015	9.0038	6.25
2012	3.5	3424	3.5345	12.3708	12.25
Total	**0**	**30836**	**28.6626**	**0.5207**	**42**

∴ The fitted model is β_0& β_1 are obtained as follows

$$Y_t = \frac{3826}{(1.02896)^t}$$

Table 3.1.2 : The observed expexted number of patients using model 1

Year	Observed patients	expected value
2005	3860	4248.05
2006	4400	4129.05
2007	4668	4013.40
2008	3826	3901.01
2009	3203	3793.77
2010	3460	3685.62
2011	3995	3582.45
2012	3424	3482.18
R^2		0.99924026
adj R^2		0.999113637

Model II:

The prosed model is $y_t = \beta_0 + \beta_1^{t_i}$

Table 3.2.1

Year	ti= (year-2008.5)	Observed patients	yi=log (observed deaths)	yi*ti	ti^2
2005	-3.5	3860	3.5866	-12.5531	12.25
2006	-2.5	4400	3.6435	-9.1086	6.25
2007	-1.5	4668	3.6691	-5.5037	2.25
2008	-0.5	3826	3.5827	-1.7914	0.25
2009	0.5	3203	3.5056	1.7528	0.25
2010	1.5	3460	3.5391	5.3086	2.25
2011	2.5	3995	3.6015	9.0038	6.25
2012	3.5	3424	3.5345	12.3709	12.25
Total	**0**	**30836**	**28.66259937**	**-0.5207**	**42**

∴ The fitted model is $Y_t = 3826 + (1.02896)^t$

Table 3.2.2: The Observed and expexted number of patients using model 2

Year	Observed patients	expected value
2005	3860	3854.40
2006	4400	3854.43
2007	4668	3854.46
2008	3826	3854.49
2009	3203	3854.51
2010	3460	3854.54
2011	3995	3854.57
2012	3424	3854.61
R^2		0.999186013
adj R^2		0.99905035

Model III: The prosed model for forecasting number hepatitis B patients is

$$y_t = \frac{\beta_0}{\beta_1^t + \beta_2^{t^2}}$$

Table 3.3.1

Year	ti= (year-2008.5)	Observed patients	yi=log (observed deaths)	yi*ti	yi* ti^2	yi*ti^3	ti^2	ti^3	ti^4
2005	-3.5	3860	3.5866	-12.5531	43.94	-153.8	12.25	-42.875	150.0625
2006	-2.5	4400	3.6434	-9.10863	22.78	-56.9	6.25	-15.625	39.0625
2007	-1.5	4668	3.6691	-5.5037	8.25	-12.4	2.25	-3.375	5.0625
2008	-0.5	3826	3.5827	-1.79137	0.89	-0.4	0.25	-0.125	0.0625
2009	0.5	3203	3.5056	1.7527	0.88	0.4	0.25	0.125	0.0625
2010	1.5	3460	3.5391	5.3086	7.96	11.9	2.25	3.375	5.0625
2011	2.5	3995	3.6015	9.0037	22.51	56.3	6.25	15.625	39.0625
2012	3.5	3424	3.5345	12.3708	43.29	151.5	12.25	42.875	150.0625
Total	**0**	**30836**	**28.6626**	**-0.5207**	**150.50**	**-3.3**	**42**	**0**	**388.5**

The fitted Model is

$$Y_t = \frac{3818.5633}{(1.02896)^t + (0.9996)^{t^2}}$$

Table 3.2.2: The Observed and expexted number of patients using model 3

Year	Observed patients	expected value
2005	3860	4247.59
2006	4400	4128.82
2007	4668	4013.40
2008	3826	3901.22
2009	3203	3792.20
2010	3460	3686.25
2011	3995	3583.28
2012	3424	3483.21
r^2		0.9991863
adj R^2		0.99905068

IV. Conclusions:

In order to choose the best model among proposed models one can use R2 criteria and Root Mean Square Error (RMSE). The fitted models are presented below:

Model-I: $Y_t = \dfrac{3826}{(1.02896)^t}$

Model-II : $Y_t = 3826 + (1.02896)^t$

Model-III: $Y_t = \dfrac{3818.5633}{(1.02896)^t + (0.9996)^{t^2}}$

Forecasted values

Year	Observed patients	expected value 1	expected value 2	expected value 3
2005	3860	4248.05	3854.40	4247.59
2006	4400	4129.05	3854.43	4128.82
2007	4668	4013.40	3854.46	4013.40
2008	3826	3901.01	3854.49	3901.22
2009	3203	3793.77	3854.51	3792.20
2010	3460	3685.62	3854.54	3686.25
2011	3995	3582.45	3854.57	3583.28
2012	3424	3482.18	3854.61	3483.21

Model	Equation	R^2	$R^2_{adjusted}$	MSE	Root Mean Square Error
I	$Y_t = \frac{3826}{(1.02896)^t}$	0.9992	0.9991	153952.5	392.3679
II	$Y_t = 3826 + (1.02896)^t$	0.9992	0.9991	218201.2	467.1201
III	$Y_t = \frac{3818.5633}{(1.02896)^t + (0.9996)^{t^2}}$	0.9992	0.9991	153661.1	391.9963

Last four columns explain R^2, adjusted R^2, Mean Square Errorand mean square root error for three models listed above. The best model among three is model3, Though R^2 value resembles in case of model 2and model3 but on observation mean square root error is minimum for model 3.

The present study help to forecast the disease before onset and then mankind can be saved from dreadful disease.

References

[1]. Akaike, H (1973), “Information theory and an Extension of the Maximum Likelihood Principle, in proceedings of the 2nd International Symposium on Information theory”, ed.by N.Petrov and F.Csad Kt, Budapeset: Akademiaikiado, 267-281.
[2]. Akaike, H (1981) “Likelihood of an Model and Information Criteria”, Journal of Econometrics, 16, 3-14.
[3]. Amemiya, T (1980), “Selection of Regressors”, International Econometric Review, 21, 331-354.
[4]. Ames Lopez Bernal, Steven Cumins, Antonio Gasparini, (2016) “Interrupted time series regression for the evaluation of public health interventions: a tutorial” International Jounral of Epidemiology. 1-8.
[5]. Ashok Chandra, K (2007), “Criterion for Selection of Regressors in Econometrics”, unpublished Ph.D., Thesis, S.V. University, Tirupati, Andhra Pradesh State, India.
[6]. Baltagi, B.H (1999), “Econometrics”, 2nd Edition, Springer-verlong, New York.
[7]. Benerjee, A.N. and Magnus, J.R. (2000), “on the sensitivity of the usual t and F-tests to Covariance Misspecification”, Journal of Econometrics, 95, 157-176.

B.Sarojamma "Model selection for forecasting growth rate of Hepatitis B patients "International Journal of Engineering Science Invention (IJESI), vol. 07, no. 04, 2018, pp 61-65