

# Modeling of Diabetic Patients Associated with Age: Polynomial Model Approach

Md. Rafiqul Islam

Department of Population Science and Human Resource Development, Rajshahi University, Bangladesh

**Abstract** Diabetes mellitus is a set of diseases that involves troubles with the hormone insulin. It is characterized by chronic elevation of blood glucose level exceeding normal value. In this paper, an effort has been made to fit mathematical model to diabetic patients as well as its cumulative distribution for both sexes associated with age of Rajshahi City in Bangladesh. For this purpose, the data have been taken from Noor (2008). In this study, an attempt has been given attention to show that the polynomial model is tried to fit to the distribution of diabetic patients associated with age as well as its cumulative distribution. It is found that the distribution of diabetic patients for both sexes associated with age follows bi-quadratic polynomial model. Moreover, it is investigated that cumulative distribution of diabetic patients follow cubic polynomial model. Cross validity prediction power is employed to the fitted model to verify the stability of the model in this manuscript.

**Keywords** Distribution Of Diabetic Patients, Polynomial Model, Variance Explained ( $R^2$ ), Cross Validity Prediction Power (Cvpp), F-Test

## 1. Introduction

Diabetes is a number of diseases that create problems with hormone insulin. Generally, the pancreas releases insulin to assist our body store and use the sugar and fat from the food that we eat. Diabetes, also familiar as Diabetes Mellitus, is a collection of metabolic disorders characterized by chronic hyperglycemia owing to relative insulin deficiency or resistance or both. Diabetes affects the body's ability as well as capacity to use blood sugar for energy. Diabetes affects people of all kinds of socio-economic conditions all over the present technological world. For this, in the era of globalization, an attention is given to concentrate to build mathematical model for the diabetic patients. In fact, mathematical model is a very realistic and sophisticated mechanism to express data in mathematical formulation. Mathematical models are actually of great useful to all discipline of applied knowledge in distinguishing important and unimportant variables. At last but not least, model is important for estimation, prediction as well as forecasting purposes. In this study, deterministic models are used to build up some mathematical relation between diabetic patients related to age of Rajshahi City in Bangladesh. Conventionally, one can draw graphs for any event but it is very important to know what types of models are apt for that event. For this purpose,

a polynomial model is chosen to apply for age association of diabetic patients. A brief discussion about polynomial model is addressed in the following that will be employed in various branches of health sciences:

A general expression of the structure

$$y = f(x) = a_0 + a_1x + a_2x^2 + a_3x^3 + \dots + a_nx^n \quad (a_n \neq 0)$$

(Waerden, 1948)

where  $a_0$  indicates the constant term;  $a_i$  represents the coefficient of  $x^i$  ( $i = 1, 2, 3, \dots, n$ ) but  $a_1, a_2, \dots, a_n$  are also constants. In addition to these constants belong to a field indicating a nonempty set in which group for addition, group for multiplication and left as well as right distributive law hold and  $n$  is the positive integer,

is called a polynomial of degree  $n$  and the symbol  $x$  is called an indeterminate.

An effort has been made here to find out what types of models are appropriate to the distribution of diabetic patients associated with age. Thus, the fundamental objectives of this study are briefly mentioned below:

i) to build up mathematical models to the distribution of diabetic patients associated with age of Rajshahi City in Bangladesh,

ii) to fit a mathematical model to the cumulative distribution of diabetic patients related to age of Rajshahi City in Bangladesh, and

iii) to apply cross-validity prediction power (CVPP),  $\rho_{cv}^2$ , to the model to verify the stability of the model.

This paper is mainly organized into five sections. First section is introduction and sources of data are included in

\* Corresponding author:

rafique\_pops@yahoo.com (Md. Rafiqul Islam)

Published online at <http://journal.sapub.org/statistics>

Copyright © 2011 Scientific & Academic Publishing. All Rights Reserved

second section. Section three contains materials and methods in which data smoothing, model fitting, model validation, shrinkage coefficient of the fitted model, F-test as well as velocity and acceleration curve of a function are briefly described here. Results of this study are narrated in section four. Lastly, section five concludes the discussion of this paper.

## 2. Sources of Data

The data on the distribution of diabetic patients for both sexes (jointly male and female) associated with age of Rajshahi City in Bangladesh have been taken from Noor (2008).

## 3. Materials and Methods

### 3.1. Data Smoothing

Before going to fit mathematical model to the data set, first the data is plotted in the graph paper, then we see that there is some sort of distortions in the data set. It is unexpected. For this reason, to remove any anomalies in the data, the data is smoothed by Free Hand Curve Method. The smoothed data is demonstrated in Table 1 and thereafter, it is used to fit mathematical model to fulfill the objectives of this manuscript.

### 3.2. Model Fitting

Using the scattered plot of the distribution of diabetic patients associated with age, it is observed that the distribution of diabetic patients can be fitted by polynomial model with respect to age. Therefore, an  $n$ th degree polynomial model is considered and the form of the model is

$$y = a_0 + \sum_{i=1}^n a_i x^i + u \quad (\text{Gupta and Kapoor, 1997})$$

where,  $x$  indicates mid value of age group;  $y$  represents distribution of diabetic patients associated with age;  $a_0$  is the constant;  $a_i$  is the coefficient of  $x^i$  ( $i = 1, 2, 3, \dots, n$ ) and  $u$  is the stochastic error term of the model. Here a suitable  $n$  has been selected for which the error sum of square is minimum.

The software STATISTICA was used to fit the mathematical model to this data aggregate.

### 3.3. Model Validation

To check the stability of the model, the cross validity prediction power

(CVPP),  $\rho_{cv}^2$ , is applied. The formula for CVPP is given in the following

$$\rho_{cv}^2 = 1 - \frac{(n-1)(n-2)(n+1)}{n(n-k-1)(n-k-2)}(1-R^2)$$

where,  $n$  is the number of cases,  $k$  = the number of regressors in the model and the cross-validated  $R$  is the correlation

relation between observed and predicted values of the dependent variable (Stevens, 1996). Islam (2003 and 2005) and Islam et al. (2003 and 2005) had employed CVPP as model validation or accuracy test of the fitted model.

### 3.4. Shrinkage Coefficient of the Model

The shrinkage coefficient of the model is the absolute value of the difference of  $\rho_{cv}^2$  and  $R^2$ . The stability of  $R^2$  of this model is equal to 1- shrinkage. The smaller quantity of the shrinkage value indicates the better fit of the model. Islam and Ali (2004) employed shrinkage coefficient as the better fit of the model.

### 3.5. F-test

To verify the measure of the overall significance of the fitted model as well as the significance of  $R^2$ , the F-test is employed to this model. The formula for F-test in mathematics is as follows:

$$F = \frac{R^2 / (k-1)}{(1-R^2) / (n-k)}$$

where  $k$  is the number of parameters to be estimated,  $n$  is the number of cases and  $R^2$  is the coefficient of determination in the model (Gujarati, 1998).

### 3.6. Velocity and Acceleration Curve of the Diabetic Patients

Let any curve can be represented by  $y = f(x)$  in which  $x$  is argument and  $y$  is entry. Then, the velocity curve of  $y$  with respect to  $x$  is symbolized by  $\frac{dy}{dx} (= v)$  and is defined by

$$\frac{dy}{dx} = \lim_{\partial x \rightarrow 0} \frac{f(x + \partial x) - f(x)}{\partial x}.$$

That is, the velocity is the rate of change of  $y$  with respect to  $x$  (Loney, 1990). Moreover, acceleration is defined by the rate of change of velocity due to age  $x$ . In this paper, the velocity and acceleration curve only for the fitted model of diabetic patients in terms of age are enumerated and depicted in Figure 3 and Figure 4 respectively in result and discussion section.

## 4. Results of This Study

The polynomial model is assumed for the distribution of diabetic patients associated with age in Rajshahi City, Bangladesh and the fitted model is

$$y = 864.3376 - 114.386x + 4.960512x^2 - 0.07856x^3 + 0.000411x^4 \quad (1)$$

t-stats	7.1771	-8.067	8.920692	-8.89111	8.399420
p-value	0.0189	0.015	0.012334	0.01241	0.013880

providing  $R^2 = 0.9997$  and  $\rho_{cv}^2 = 0.994857$ . This is the polynomial of degree four, i.e. bi-quadratic polynomial model.

The polynomial model is assumed for the cumulative distribution of diabetic patients associated with age of Rajshahi City in Bangladesh and the fitted model is

$$y = 412.8274 - 49.5236x + 1.74262x^2 - 0.0131x^3 \quad (2)$$

t-stats 4.9038 -7.0197 10.12432 -10.3118  
p-value 0.0162 0.0059 0.00205 0.0019

providing  $R^2 = 0.99998$  and  $\rho_{cv}^2 = 0.999983$  and shrinkage coefficient = 0.00001414. This is the polynomial of degree three, i.e. cubic polynomial model.

From this statistics we see that the fitted models are highly cross-validated and their shrinkages are 0.004842857 and 0.00001414 for the models (1) and (2) respectively. These imply that both the fitted model is more than 99% stable. Moreover, all the parameters of the fitted model are highly statistically significant with exceeding 99% of variance explained. Moreover, the stability of  $R^2$  of these models is also more than 99%.

In this study, the calculated values of F-statistics are 1666.2 and 9999 for the models (1) and (2) respectively, that is, which mean that both the fitted model is overall significant at 1% level. Therefore, from these statistics it is seen that the fitted model and corresponding  $R^2$  are highly statistically significant. As a result, both the models are better fit. Thereafter, the prediction is better done and the predicted values of the model are also demonstrated in the Table1.

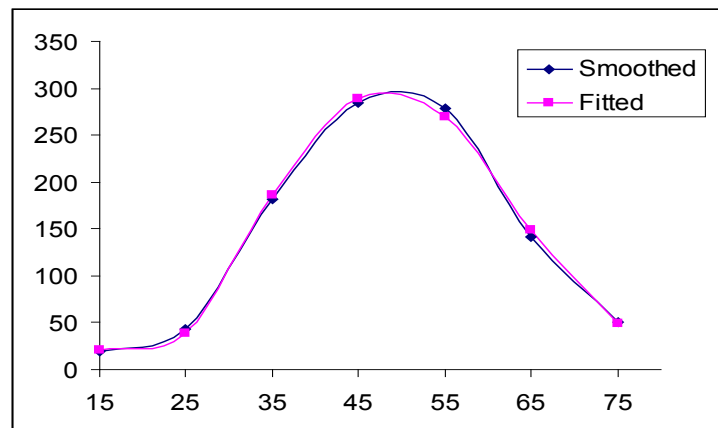
After the fitted model, velocity and acceleration curve are estimated only for the distribution of diabetic patients in terms of age and shown in Figure 3 and Figure 4 correspondingly.

**Table 1.** Distribution of diabetic patients associated with age in Rajshahi City, Bangladesh

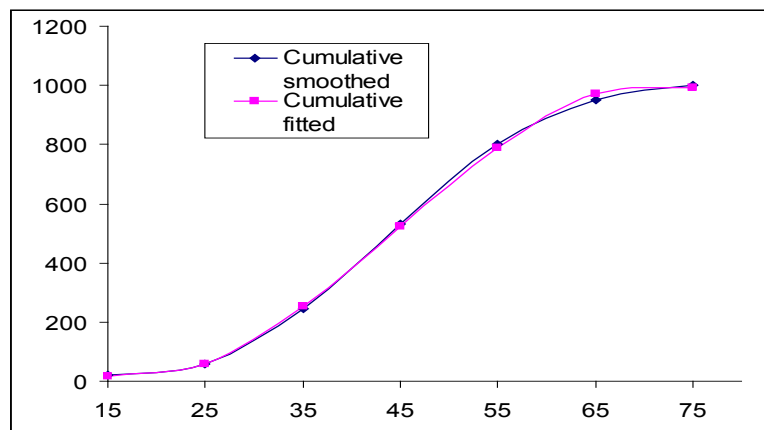
Age group (in years)	Observed number of diabetic patients	Smoothed number of diabetic patients	Fitted number of diabetic patients	Smoothed cumulative number of diabetic patients	Fitted cumulative number of diabetic patients
11-20*	19	19	20	19	18
21-30	47	42	38	62	60
31-40	172	181	186	243	254
41-50	278	286	288	528	523
51-60	300	277	269	807	788
61-70	130	144	149	949	971
70-80**	54	51	49	1000	993

\*Less than 10 years age group is assumed as 11-20 years age group in this study.

\*\*Above 80 years is included in the age group 70-80.



**Figure 1.** Smoothed and fitted diabetic patients associated with age in Rajshahi City, Bangladesh



**Figure 2.** Smoothed and fitted cumulative distribution of diabetic patients associated with age in Rajshahi City, Bangladesh

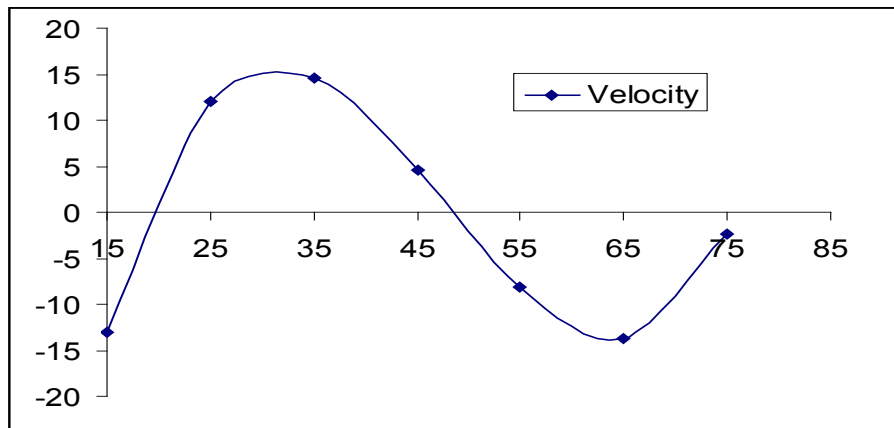


Figure 3. The velocity curve for the fitted model of diabetic patients associated with age in Rajshahi City, Bangladesh

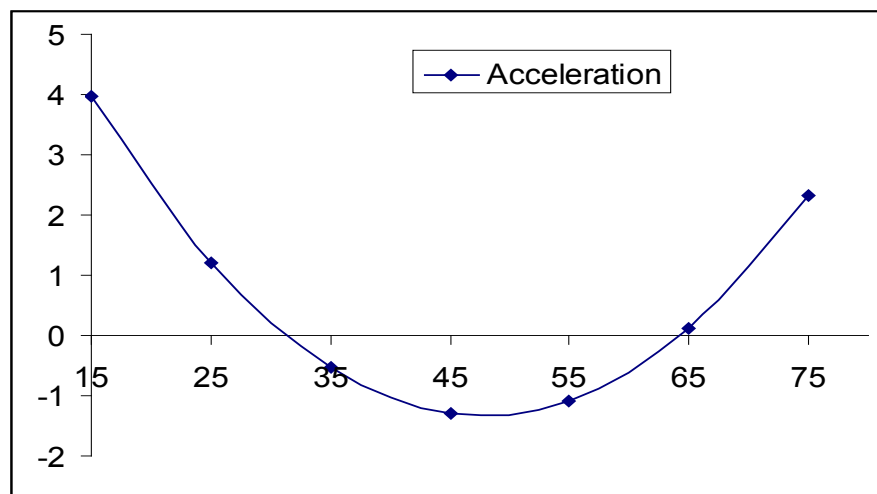


Figure 4. The acceleration curve for the fitted model of diabetic patients associated with age in Rajshahi City, Bangladesh

## 5. Conclusions

In this paper, it is found that fourth degree polynomial model containing more than 99% variation is fitted to the distribution of diabetic patients associated with age of Rajshahi City in Bangladesh. On the other hand, cumulative distribution of diabetic patients in accordance with age follow cubic polynomial model with explaining more than 99% variation. It should be noted that the usual models such as exponential model, Gompertz model, Makeham model and log-linear model were also applied for the distribution of diabetic patients and cumulative distribution of diabetic patients but seems to be worse fitted with respect to their coefficient of determination as well as shrinkage coefficients. Therefore, the results of these models were not shown here.

## ACKNOWLEDGEMENTS

I heartily thanks to Behzad Noor for submitting the M. Sc. thesis to the Dept. of Population Science and Human Resource Development, University of Rajshahi, Bangladesh.

## REFERENCES

- [1] Gujarati D N. Basic Econometrics, Third Edition, McGraw Hill, Inc., New York 1998.
- [2] Gupta SC and Kapoor VK. Fundamentals of Mathematical Statistics, Ninth Extensively Revised Edition, Sultan Chand & Sons, Educational Publishers, New Delhi 1997.
- [3] Islam Md Rafiqul. Modeling of Demographic Parameters of Bangladesh-An Empirical forecasting, Unpublished Ph.D. Thesis, Rajshahi University 2003.
- [4] Islam Md Rafiqul, Islam Md Nurul, Ali Md Ayub and Mostofa Md Golam Construction of Male Life Table from Female Widowed Information of Bangladesh, International J. of Statistical Sciences, Vol. 2, Dept. of Statistics, University of Rajshahi, Bangladesh 2003: 69-82.
- [5] Islam Md Rafiqul and Ali M Korban. Mathematical Modeling of Age Specific Fertility Rates and Study the Reproductivity in the Rural Area of Bangladesh During 1980-1998, Pakistan Journal of Statistics, Pakistan 2004: 20(3):381-394.
- [6] Islam Md Rafiqul. Construction of Female Life Table from Male Widowed ,Information of Bangladesh, Pakistan J. of Statistics 2005: 21(3): 275-284.

- [7] Islam Md Rafiqul, Islam Md. Nurul, Ali M Korban and Mondal Md Nazrul Islam. Indirect Estimation and Mathematical Modeling of Some Demographic Parameters of Bangladesh, *The Oriental Anthropologist* 2005; 5(2): 163 - 171.
- [8] Loney S L. *An Elementary Treatise on the Dynamics of a Particle and of Rigid Bodies*, Cambridge University Press, London 1990.
- [9] Noor Behzad. *Health Status and Consciousness of the Diabetic Patients: A Study on the Diabetic Patients of Rajshahi City, Bangladesh*, M.Sc Thesis, Submitted to the Dept. of Population Science and Human Resource Development, University of Rajshahi, Bangladesh 2008.
- [10] Stevens J. *Applied Multivariate Statistics for the Social Sciences*, Third Edition, Lawrence Erlbaum Associates, Inc., Publishers, New Jersey 1996.
- [11] Waerden BL Van Der. *Modern Algebra*, Vol. 1, ICK Ungar Publishing Co. New York 1948.