

**A LINEAR MODEL AS AN APPROXIMATION TO GOMPERTZ MODEL**

**Dr. R. B. Tiwari\***

Department of Statistics, D.A-V. College, CSJM University, Kanpur, U.P., India.

\*Corresponding Author: **Dr. R. B. Tiwari**

Department of Statistics, D.A-V. College, CSJM University, Kanpur, U.P., India.

Article Received on 16/07/2022

Article Revised on 06/08/2022

Article Accepted on 26/08/2022

**ABSTRACT**

In this study a linear model has been suggested as an approximation to the existing non-linear model viz. Gompertz curve. It has been proved by fitting the suggested model to different data sets obtained from various authentic sources. The proposed model has been applied to data set obtained from Kutner et al (2004) in table 1.1 and N Okendro Singh et al (2009) in table 1.2. The results in terms of relevant parameter values have been displayed through tables and the graphical representation has been done by using a Statistical Software 'Statgraphics'. Just by having a careful look over the results one can conclude easily that the proposed linear model is a good approximation to the existing non-linear model.

**KEYWORDS:** *Linear model, non-linear model, Gompertz curve, linear approximation to non-linear models.*

**INTRODUCTION**

Based on the relationship between monomolecular model and Gompertz model, a model having linearly appearing parameters has also been developed as an adequate approximation of Gompertz model. Ratkowsky (1983) has given the following functional form for Gompertz model.

$$Y = \exp\{\alpha + \beta \cdot \rho^x\} \quad \dots\dots (1)$$

On taking Natural Log Transformation of (1) we have

$$\log(Y) = \alpha + \beta \cdot \rho^x \quad \dots\dots (2)$$

Equation (2) is another form of writing the model (1).

If a graph is plotted between Y and X, it will show an 'S'- shape Curve. From equation (2) we learn that the graph between log(Y) and X will show an asymptotic shape. This fact has been observed for the suggested data sets considered in the present investigation.

In early studies we have considered an approximation of the Right Hand side of equation (2), on the basis of suggestions of Misra (1992) as

$$\alpha + \beta \cdot \rho^x = a + b \cdot X + \frac{c}{X} \quad \dots\dots (3)$$

Combining equations (2) and (3) we can write an approximation of the model (1) as

$$\log(Y) = a + b \cdot X + \frac{c}{X} \quad \dots\dots (4)$$

The appropriateness of model (4) as an approximation to Gompertz model can be verified by fitting it to various data sets obtained from various sources and carrying out

its analysis and it's residual studies as described in bellow. Equation (4) is the **proposed model**.

**Estimation of Parameters and Analysis of Residuals**

The proposed model (4) can be written as

$$Z = \log(Y) = a + b \cdot X + \frac{c}{X} \quad \dots\dots (5)$$

The constants a, b, c, α, β and ρ are all unknown parameters. Applying the Standard Procedures the least square estimates of parameters a, b, and c will be given by

$$\hat{B} = (X' X)^{-1} X' Z \quad \dots\dots (6)$$

where

$$\hat{B} = [a \ b \ c]'$$

$$X = \begin{bmatrix} 1 & X_1 & \frac{1}{X_1} \\ 1 & X_2 & \frac{1}{X_2} \\ \dots\dots\dots & \dots\dots\dots & \dots\dots\dots \\ 1 & X_n & \frac{1}{X_n} \end{bmatrix}_{n \times 3}$$

$$Z = [Z_1 \ Z_2 \ \dots\dots \ Z_n]'$$

The variance-covariance matrix of  $\hat{B}$  denoted by  $\text{Var}(\hat{B})$  will be

$$\text{Var}(\hat{B}) = (X'X)^{-1}\sigma^2 \quad \dots\dots (7)$$

The values of coefficient of determination  $R^2$ , Residual Mean Square  $s^2$ , Mean Absolute Error (MAE) and Standardized residuals are computed in a manner identical to methods discussed in earlier study as in Tiwari (2010). Similarly the analysis of residuals i.e. residual plots and Normal Probability Plot for the residuals have also been carried out as discussed in Tiwari (2010). The analysis table displaying values of  $\log(Y)$ ,  $\log(\hat{Y})$ ,  $e_i = \log(Y) - \log(\hat{Y})$  etc. are also given for the data sets from 1.1 to 1.2.

It is very important to note here that Draper and Smith (1998, Pg. 278) have specifically stated that the residuals to examine, for model like (5) in which log transformation has been used on dependent variable Y, should be of the form  $\log(Y) - \log(\hat{Y})$ . They have further stated that all tests and confidence statements must be made in the Transformed Space only.

Draper and Smith (1998, III Edition, Pg. 278) have specifically commented that: "Suppose we have fitted a model to  $\ln(Y)$ , and we make a prediction in  $\ln \hat{Y}$  at a certain set of X's. We can, if we wish, evaluate  $\hat{Y} = \exp\{\ln \hat{Y}\}$  and predict in the original space.

Also, confidence statement on  $E(\ln Y)$  with interval (a, b) can be translated into a confidence statement with interval  $(e^a, e^b)$  in the Y-space. It will not be symmetric about the predicted value  $\hat{Y}$ , of course. We can also evaluate residuals  $Y_i - \hat{Y}_i$  at the data points, if we wish. These residuals are not, however, checked; these are not the residuals that should satisfy the residuals checks for normality, and so on".

Box and Cox (1964) have given a family of transformations on dependent variable Y as

$$(i) (Y^\lambda - 1) / \lambda, \text{ for } \lambda \neq 0$$

$$(ii) \log(Y), \text{ for } \lambda = 0$$

The model (4) also admits a transformation on dependent variable Y as in Box and Cox (1964) family of transformations if  $\lambda = 0$ .

It has also been remarked that there is no relationship between the parameters of the transformed model and those of the original model, therefore any attempt to find such a relationship is not fruitful. The analysis of the

residuals has also been carried out in the same lines as in Misra (1992).

#### Fitting of Model (4) to Suggested Data Sets

A variety of the data sets from 1.1 to 1.2 has been obtained from different sources to which the model (4) has been applied.

The data set 1.1 has been obtained from Kutner et al (2004) which relates to a study on effectiveness of coupons offering a price reduction on a given product. The predictor variable X represents the amount of price reduction offered to coupons and the response variable Y represents the number of coupons redeemed.

The data set 1.2 are obtained from N Okendro et al (2009), they relate to growth performance of *Tor Putitora* (a fish species) under the monoculture and polyculture systems with the stocking densities of 4000 and 1600 per hectare respectively in earthen ponds of 0.1 hectare each. Here Y is the Average weight gain in gm and X is Time in months.

The least square estimates of coefficients a, b and c of the model (3.4) fitted to the above data sets are given in their respective tables. The values of coefficient of determination  $R^2$ , Residual Mean Square  $s^2$  and Mean Absolute Error (MAE) are also provided. The tables also provide the values of least square estimates of  $\hat{Z} = \log(\hat{Y})$  and residuals  $Z - \hat{Z}$  or  $\log(Y) - \log(\hat{Y})$  (as suggested by Draper and Smith 1998, pg. 278).

The values of Standardized residuals have been calculated for all data sets. In all cases, the values of standardized residuals lie in between -3 to +3. The values of standardized residuals confirm that residuals do not reflect model inadequacy, (Montgomery et al, 2003).

The graphs describing observed and fitted values, residual plots and Normal Probability Plot for the residuals are also drawn for the data sets. The graph of observed versus fitted values confirms an excellent fit in case of all data sets. It is very interesting to note that plot of  $\log(Y)$  versus X and also plot of  $\log(\hat{Y})$  versus X are both asymptotic in shape. This confirms type of relationship between asymptotic model and Gompertz model.

It has been observed that the values of  $R^2$  are very high and values of Residual Mean Square  $s^2$  are very small for all data sets. The Mean Absolute Error is also considerably small.

The residual plots of  $e_i$  versus  $X_i$  show a horizontal band for all data sets. It confirms that residuals fulfil the model assumptions. The points of Normal Probability Plot for the residuals are approximately in a straight line which means that errors follow normal distribution. All these observations lead to the conclusion that the model (1.4)

can be considered as a good approximation of Gompertz model represented by equation (1.2). Moreover without loss of generality we can consider  $X=1$  instead of  $X=0$ , wherever necessary.

It is to be emphasized that the analysis of residuals on all data sets has been carried out as per suggestions of Draper and Smith (1998, III Edition, Pg. 278), Montgomery et al (2003) as well as Box and Cox (1964). The analysis along with the necessary graphs for all data sets is as follows.

**Data Set: 1.1 Source of the Data: Kutner et al (2004). Ch.-13. Pg. 569.**

**Analysis of the data**

**Table-1 A**

S. No.	X	Y	Log(Y)	Log( $\hat{Y}$ )	Residual (Log(Y)- Log( $\hat{Y}$ ))	Standardized Residuals
1	5	30	3.401200	3.399770	0.001430	0.022835
2	10	55	4.007330	3.998320	0.009010	0.143879
3	15	70	4.248500	4.305870	-0.057370	-0.916131
4	20	100	4.605170	4.540680	0.064490	1.029829
5	30	137	4.919980	4.937530	-0.017550	-0.280253

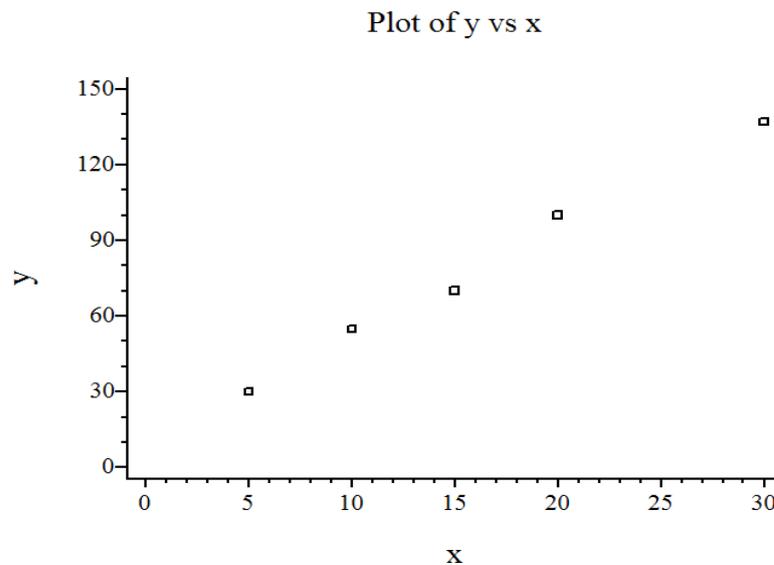
**Parameter Estimates**

**Table-1 B**

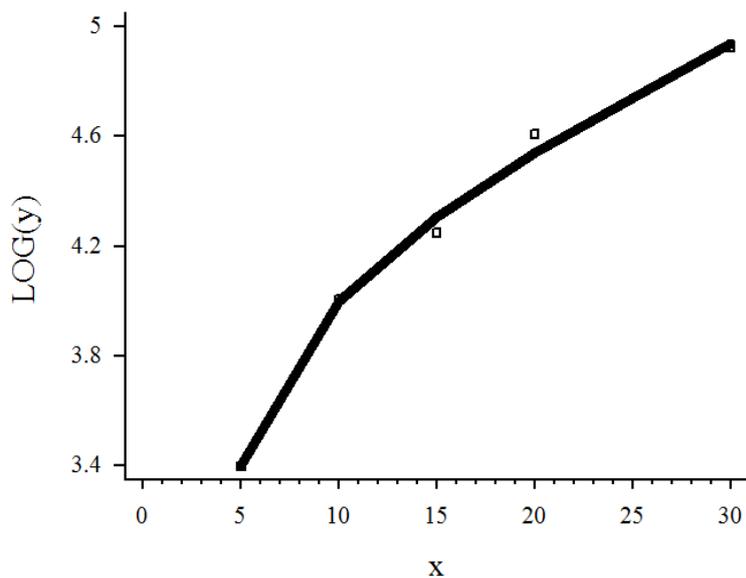
S. No.	Results	
1	$\hat{a}$	4.11072
2	$\hat{b}$	0.0324102
3	$\hat{c}$	-4.36499

Coefficient of determination ( $R^2$ ) = 99.4205 percent  
 Residual Mean Square ( $s^2$ ) = 0.00392152  
 Mean Absolute Error = 0.0299705

**GRAPH FOR THE DATA SET 1.1**

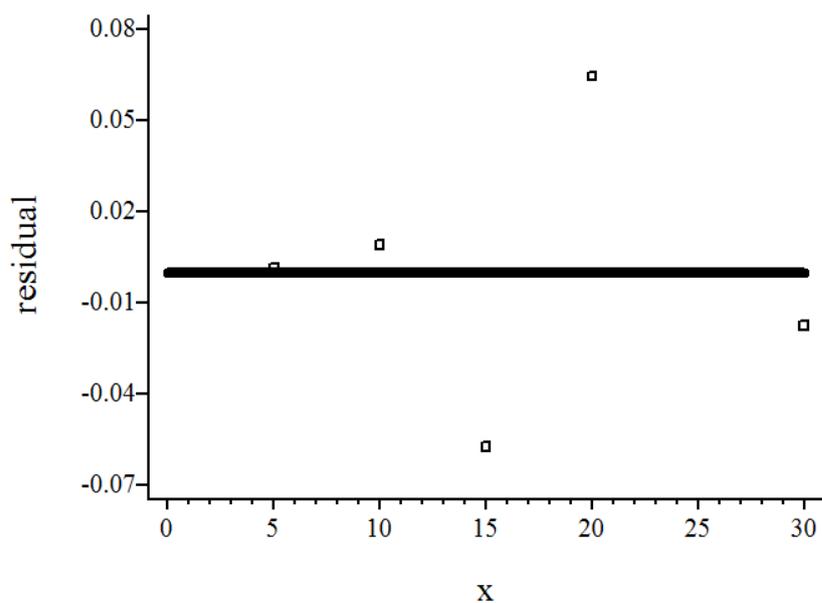


Plot of LOG(y) with Predicted Values

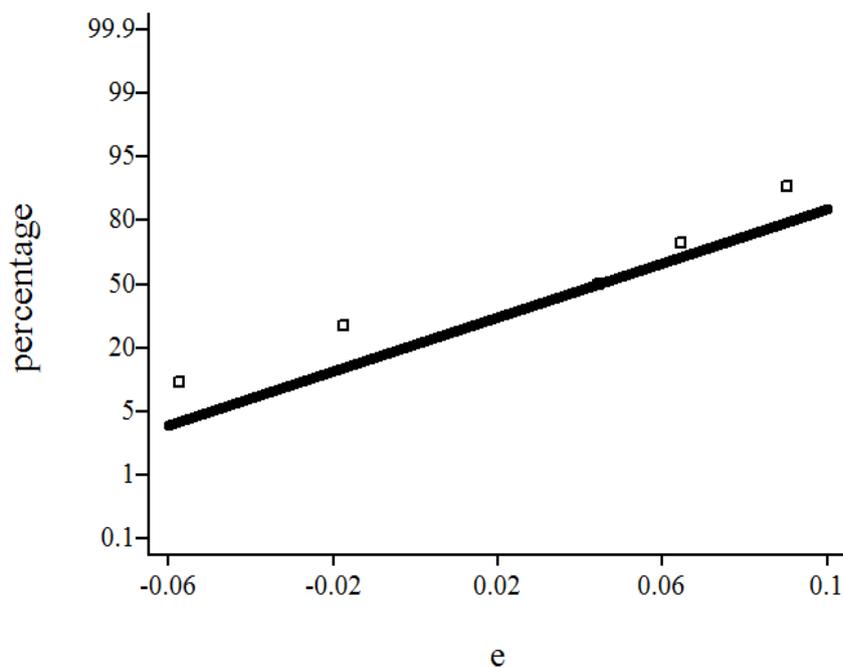


GRAPH FOR THE DATA SET 1.1

Residual Plot



Normal Probability Plot for e



Data Set: 3.3.4

Data Set: 1.2

Source of the Data: N Okendro Singh et al (2009). (Monoculture)

Analysis of the data

Table-2 A:

S. No.	X	Y	Log(Y)	Log( $\hat{Y}$ )	Residual (Log(Y)- Log( $\hat{Y}$ ))	Standardized Residuals
1	1	320	5.768320	5.761010	0.007310	0.242572
2	3	400	5.991460	6.014960	-0.023500	-0.779813
3	6	475	6.163310	6.179190	-0.015880	-0.526954
4	9	550	6.309920	6.314520	-0.004600	-0.152644
5	12	650	6.476970	6.442630	0.034340	1.139522
6	15	725	6.586170	6.567850	0.018320	0.607922
7	18	825	6.715380	6.691620	0.023760	0.788440
8	23	950	6.856460	6.896240	-0.039780	-1.320040

Parameter Estimates

Table-2 B

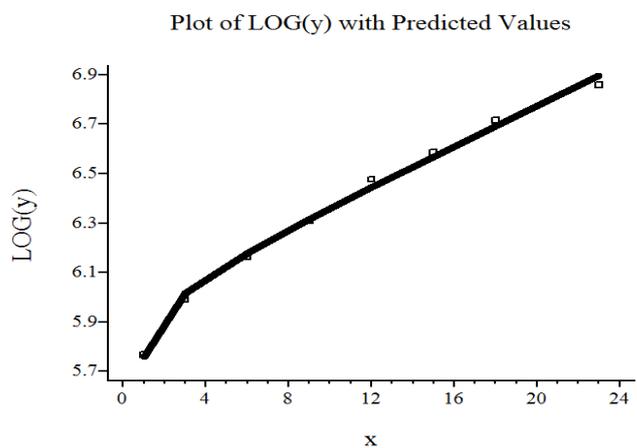
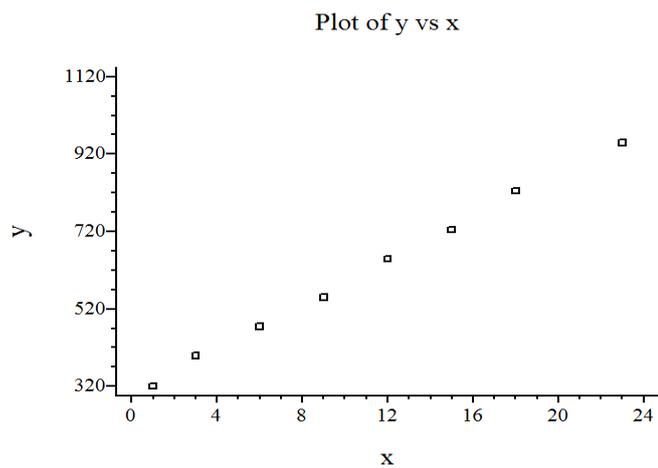
S. No.	Results	
1	$\hat{a}$	5.98075
2	$\hat{b}$	0.0402953
3	$\hat{c}$	-0.260041

Coefficient of determination ( $R^2$ ) = 99.5293 percent

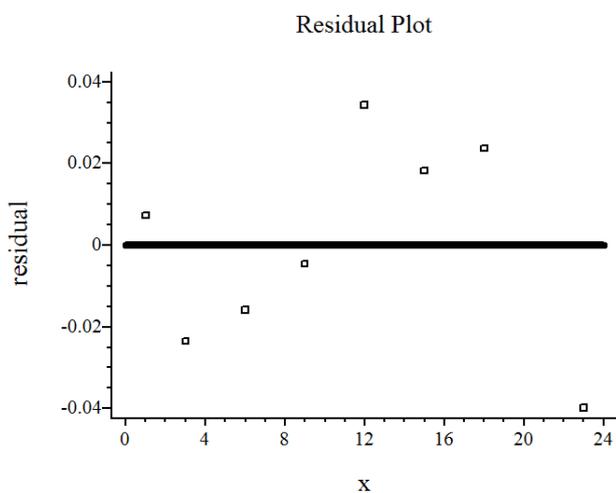
Residual Mean Square ( $s^2$ ) = 0.000908145

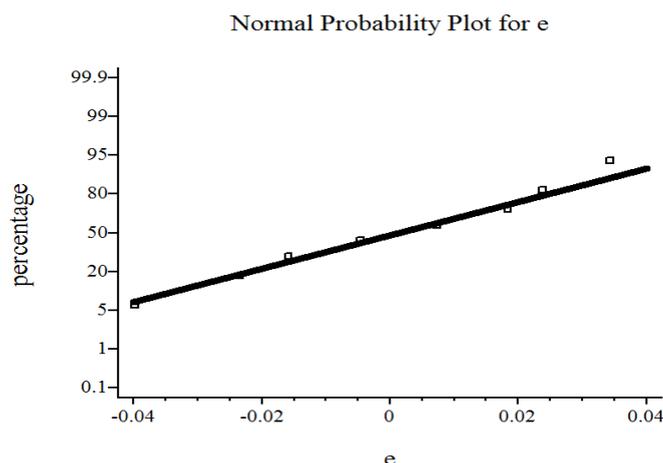
Mean Absolute Error = 0.0209357

GRAPH FOR THE DATA SET 1.2



GRAPH FOR THE DATA SET 1.2





### CONCLUSION

The data sets arising from different fields have been considered for the application of proposed model (4), which has satisfactorily described the data sets. The graph between X and Y resembles a ‘S’- shape where as the graph between  $\log(Y)$  and X as well as super imposed graph between  $\log(\hat{Y})$  and X are both asymptotic in nature. It confirms the theoretical relationship between Gompertz and asymptotic model. The graphs between  $\log(\hat{Y})$  and X show an excellent fit for all data sets.

A very high values of  $R^2$  in percentage have been observed for all data sets, The magnitude of residual sum of squares  $s^2$  is also very small for all data sets. The values of mean absolute errors are also very small in all cases. The values of Standardized residuals are also between -3 to +3 which confirm the absence of any outliers in the data.

The residual plots are showing horizontal band for all data sets and can be considered as satisfactory and are not showing any unusual pattern. The Normal Probability Plot for the residuals obtained for all data sets also show a straight line. All these discussions lead to the conclusion that the model (4) is a good approximation of Gompertz model represented by equation (2).

### REFERENCES

- Alba Vilanova 1,2, Bo-Young Kim 1, Chang Ki Kim 1 and Hyun-Goo Kim 1, Linear-Gompertz Model-Based Regression of Photovoltaic Power Generation by Satellite Imagery-Based Solar Irradiance, *Energies*, 2020; 13: 781.
- Aggrey S E on Logistic nonlinear mixed effect model for estimating growth parameters. *Journal of Poultry Science*, 2009; 88: 276-280.
- Akbar Atif, Pusha G.R., Aslam, Muhammad Warsi S.K.A. “Statistical Modeling for citrus yield in Pakistan”. *European journal of Scientific Research* ISSN 1450-216X, 2009; 31(1): 52-58.
- Cochran W G Sampling Techniques. John Wiley & Sons, 1977.
- Jessen R J Statistical Survey Techniques. Wiley, New York, 1978.
- Mahir A, Rozita W D W, Khairiah J and Ismail B S Fitting a Nonlinear Regression Model to Gauge Heavy Metal Uptake in Spinach (*Amaranthus hybridus* L.). *American-Eurasian J. Agric. & Environ. Sci.*, 2009; 5(2): 236-243.
- Misra G C “Some Estimation Precedures in Non-Linear Statistical Models”. The Unpublished Thesis of Ph. D., Kanpur University, Kanpur, 1992.
- Moh. Saat N Z, Jemain A A, Al-Mashoor S H A A Comparison of Weibull and Gamma Distribution in Application of Sleep Apnea. *Asian Journal of Mathematics and Statistics*, 2008; 1(3): 132-138.
- Mukhopadhyay, P. Theory and Methods of Survey Sampling. Prentice Hall of India Pvt. Ltd. New Delhi, India, 1998.
- Nygaard R W, Katharina V E and Ronald A S Models of reading performance in older adults with normal age-related vision. *Journal of rehabilitation Research & Development*, 2008; 45(6): 901-910.
- Singh N O, Alam M W, Paul A K and Kumar S Length-Weight Relationship and Growth Pattern of Tor Putitora (Hamilton) under Monoculture and Polyculture Systems: A Case Study. *J. Ind. Soc. Statist*, 2009; 63(1): 85-89.
- Shukla et.al. New Linear Model for Optimal Cluster Size in Cluster Sampling. *Statistics in Transition new Series*, 2020; 21(2): 189-200.
- Tiwari R.B. Application of Three Parameter Model For Determination of optimum Cluster Size, *International Journal of Research and Analytical Reviews ijar*, 2022; 09(03): 134-143.