ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 3, Issue 12 , December 2016

Curvilinear Regression For Software Quality Analysis

DVSS.SUBRAHMANYAM

Dept. of CSE , Swami Vivekananda Institute of Technology (SVIT), Secunderabad-500003 , Telangana

ABSTRACT: The main function and purpose of Regression analysis is predictions in any domain irrespective of its nature. Regression analysis aims at designing the relationship between the mean of a random variable and one or more other variables. Even though Regression analysis provides many Regression tools for predictions but much importance is attached to curvilinear regression as most of the real life data are non-linear in nature. Thus most of the regression curves, for predictions, are categorized under curvilinear regressions. It has a prominent role in the domain of Software Quality Analysis to give predictions accurately and effectively with respect to their quality, durability, cost of the project , evaluating internal dependence of components etc.,. By this methodology any non-linear data can be converted into linear data by adopting suitable transformations.

KEYWORDS: Regression tools, curvilinear Regression, Software Quality Analysis

I.INTRODUCTION

The fundamental Regression technique for predictions is Linear Regression[6][4][5][7]. This technique provides good and reliable predictions if the scatter diagram of the entire given data is gathered around a straight line like path. This accumulated data seems to be a straight line in its scatter diagram[2][3]. Then this data is transformed into a linear equation to get the required predictions, which is called linear regression technique[2][3][6]. But real practical applications have , in most of the cases, data which are of the form of non-linear in nature. In this case linear regression technique is not applicable and worked out. Then other non-linear regression techniques are to be taken into consideration. One of them is Curvilinear regression technique which helps to convert a given data of non-linear nature into a data of linear nature[6][5]. Curvilinear regression technique ,Mathematically, converts a non-linear function into a linear function to give a linear shape to the data in order to make predictions in a particular way[1][2][3][6]. Regression techniques have many applications in Software Engineering. They can be used in any type of domains of Software Engineering for making accurate predictions[4][5][7]. Here the main focus is on Software Quality with respect to used software in Software companies. Software companies invest millions of rupees to improve quality of software they used for their customers. Software quality has been the yardstick for assessing Quality Analysis in Software companies[5][7]. It is very difficult for companies to invest millions of rupees every year in order to make new software for their customers instead they always makes efforts to reuse the existing software for any future requirements. For this reason, the quality of the software that the companies used is to be analyzed. Curvilinear regression technique helps to assess the quality of the software that companies are using for their customers[4][6].

II. CURVILINEAR REGRESSION

It is more useful for non-linear nature of data. Even though there are many non-linear regression models under curvilinear regression methodology, but here the main interest of discussion is on exponential regression technique[2][3][6], which is the more frequently used data by software companies. This exponential regression technique is of the form of $y = a . b^{x}$, which is non-linear, where x, y are variables and a, b are constants. Then $y = a . b^{x}$ ---------- (1) is transformed into a linear equation by taking logarithms both sides. Least Squares Method is used for equation (1)[1][3][6]. A software company collected data of used software that it supplied to its customers. It wanted to know how much software was unused by its customers based upon their nature of operations[7]. The company wants to update the existing software without going for a new software for their customers in order to support their day-to-day operations for their business activities[6][7]. The company collected data from their customers regarding the percentage of software that is unused

ISSN: 2350-0328
International Journal of Advanced Research in Science, Engineering and Technology
Vol. 3, Issue 12 , December 2016

by them. The company wants to know how much actual percentage is being utilized by their customers and the company wants to predict the same for their future operations. The company collected data from their customers as follows :

Table 1 : Software used for number of months and percentage of software used by customer companies

Software used for no. of months (x)	10	19	26	35	39	46	49	58	69	73	79	84	80	95	100
Percentage of software used (y)	34	50	38	45	39	44	31	51	41	32	35	30	29	50	42

The data can be put into a scatter diagram as shown in the above Figure 1. The entire data above is not linear so it is to be adjusted into another technique called exponential regression technique of curvilinear regression, which is a good fit for the given data. All the above data of Software used for no. of months is considered on X- axis and Percentage of Software used is considered on Y-axis as shown below in Figure 1. A scatter diagram (Figure 1) is generated on the basis of the data on X-axis and Y-axis. All data are gathered not in a proper order, that is why linear regression technique can't be used in this case. Because of the irregular gatherings of data, a non-linear regression technique is considered in this case. Here almost all data entries equally influence the output predictions of this regression technique. All data are equally considered irrespective of their position in the scatter diagram. A limited number of data are considered here for predictions because of some limitations.

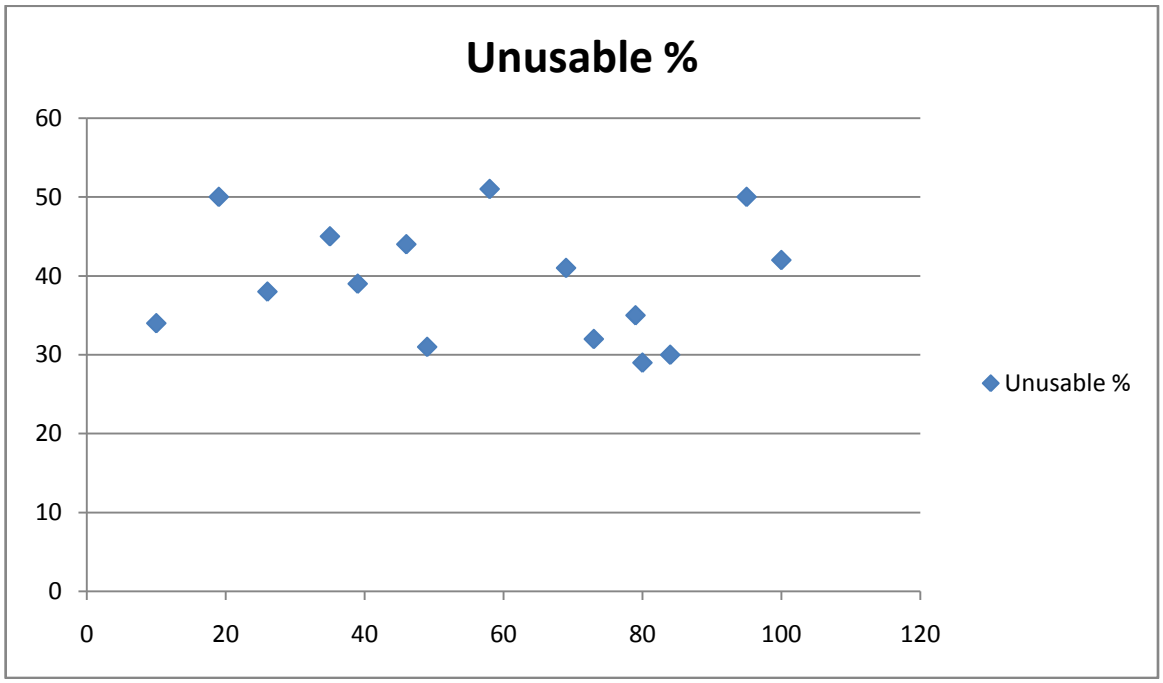

Figure 1 : Scatter diagram of unusable percentage of software not used by customer companies

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 3, Issue 12 , December 2016

III. METHODOLOGY

Exponential regression technique is [3][6] $y = a \cdot b^{x}$ ---------- (1)
taking logarithms both sides, the above equation becomes as $\log y = \log a + x \cdot \log b$ ---------- (2).
Assuming $Y = \log y$, $A = \log a$ and $B = \log b$ equation (2) becomes as $Y = A + B x$ ---------- (3),
which is a linear regression equation[2][3][6].
Normal equations are to be framed for the equation (3).
Normal equations are $\sum Y = A n + B \sum x$ ----------(4)
and $\sum x \cdot Y = A \sum x + B \sum x^{2}$ ---------- (5)
where A and B are predictable constants and x and Y are variables and n is the number of data items used .
Solving equations (4) and (5) , values of A and B can be obtained as follows :
$A = \{ \sum Y \cdot \sum x^{2} - \sum x \sum x Y \} / \{ n \cdot \sum x^{2} - (\sum x)^{2} \}$ ---------- (6) ,
where n is the number of data items used , here it is n =15 ,
$B = \{ n \sum x \cdot Y - \sum x \cdot \sum Y \} / \{ n \sum x^{2} - (\sum x)^{2} \}$ ---------- (7)
Here predictable constants A and B are expressed in terms of variables x , Y and its combinations. Even though these equations seem to be a complicated one for calculations but all these are necessary and sufficient equations for obtaining the actual values of A and B. Then these are to be converted in terms of a and b for actual fit.

Table 2 : Calculations of data items of equations (6) and (7)

x	y	Y = log y	x . Y	x^{2}
10	34	1.5315	15.315	100
19	50	1.6990	32.281	361
26	38	1.5798	41.0748	676
35	45	1.6532	57.862	1225
39	39	1.5911	62.049	1521
46	44	1.6435	75.601	2116
49	31	1.4914	73.07861	2401
58	51	1.7076	99.0408	3364
69	41	1.6128	111.2832	4761
73	32	1.5051	109.8723	5329
79	35	1.5441	121.9839	6241
84	30	1.4771	124.0764	7056
90	29	1.4624	131.616	8100
95	50	1.6990	161.405	9025
100	42	1.6232	162.32	10000
$\sum x = 872$		$\sum Y = 23.8208$	$\sum x \cdot Y = 1378.859$	$\sum x^{2} = 62276$

Substituting the above values in equations (6) and (7) , the following outputs are observed :
$A = \{ 23.8208 * 62276 - 872 * 1378.859 \} / \{ 15 * 62276 - (872)^{2} \} = 1.61778$ ---------- (8)
$B = \{ 15 * 1378.859 - 872 * 23.8208 \} / \{ 15 * 62276 - (872)^{2} \} = \{ - 88.8526 \} / \{ 173756 \}$
i.e., **B = - 0.0936** ---------- (9)
Now substituting the value of (9) in (8) ,
$A = 1.588 - 58.133 * (- 0.0936) = 1.588 + 5.441 = 7.029$ i.e., **A = 1.61778** ---------- (10)
Here equations (9) and (10) give values of A and B respectively. It is clear from equation (2) that
$Y = \log y$, $A = \log a$ and $B = \log b$
Here to find the actual values of a and b , Anti-logarithms are to be calculated for A and B respectively
. i.e., $\log a = A = 1.61778 \rightarrow a =$ Anti-log of $1.61778 = 41.47$ i.e ., **a = 41 . 47** ---------- (11)
Again $\log b = - 0.0936 = 0.091$ i.e ., **b = 0. 091** ---------- (12)
Now the above equation (2) will get the actual form of the equation (1)
i.e., $\mathbf{y = 41.47 * (0.091)^{x}}$ ---------- (13)
which is the required equation of the exponential regression technique for the given data[3][6][1]. This relation gives required predictions for any type of data. If the Software company desires to know the unused percentage software used by one of its customers if a given software is used by the customer for 10 months on an average , then it can

ISSN: 2350-0328

International Journal of Advanced Research in Science, Engineering and Technology

Vol. 3, Issue 12 , December 2016

calculated by $y = 41.47 * (0.091)^{x}$ Thus the quality of the software can be adjudged on the basis of curvilinear regression technique. It can be extended to any type of domains irrespective of their nature. Here predictions always give an approximate values but 100 % accuracy output can never be expected Internal dependent characteristics and strength of components of Software can also be calculated by using the same technique[4][5][7]. This is purely manual technique that is based on calculations. The same calculation oriented software is available in the market . No major differences could be found between these two but in calculation time. This technique gives nearly an accurate and effective predictions comparing with any of the existing software in the market. This tool has an unlimited applications to apply any domain[7][5].

IV. CONCLUSION

Here in this paper a limited number of data items are collected because of the limitations of the paper. If the data items are increased and its domain is expanded then more accurate predictions are possible by the same regression technique. The degree of predictions is increasing as data domain increases proportionately. As the main purpose of using Regression technique is predictions , it is more closely and accurately applicable to Software Quality Analysis (SQA)[4][5][7]. As the domain and range of SQA is vas t, so many regression techniques can be used as per the requirement of the situations. In whatever way Regression techniques are used in Software Quality Assurance , they give predictions very accurately, effectively and efficiently[4][5][7].

V. REFERENCES

[1] Alan Clements, Principles of computer hardware. Second Edition, Oxford Science Publications, page no. 10 -15, March 2006
[2] Janet Crawshaw, Probability and Statistics 1 for Cambridge International A Level, page no. 75-80 , Nov 2001
[3] James Nicholson, Mathematics for Cambridge International AS & A Level: Oxford Probability & Statistics 2 for Cambridge International AS & A Level, pages no.100-110, April 2015
[4] Yudistira Ansar , Volha Bryl and Paolo Giorgini. Using Risk Analysis to Evaluate Design Alternatives. Page no.66, March 2006
[5]Y.Ansar and P.Giorgini. Modelling Risk and Identifying Countermeasures in Organisations , page no. 35-40 , July 2006
[6].Bayesian Statistics and Marketing by Peter, Rossi and Allenby (Wiley Series) page no. 40 – 85 August 2004
[7] Professional Software Development by Steve McConnell , Wiley India Private Limited ,page no. 120 – 145, 2006

AUTHOR'S BIOGRAPHY

Dr.D.V.S.S.Subrahmanyam is Professor at Dept. of Computer Science & Engineering at Swami Vivekananda Institute of Technology (SVIT), Secunderabad. He did his graduation AMIETE, M.Tech and Ph.D in Computer Science & Engineering, Also did M.Sc in Industrial Mathematics and another M.Sc in Mathematics and M.Phil in Mathematics. Areas of interest include Software Engineering, Software Quality Analysis, Cyber Security and Big Data Analysis..