ISSN: 2319-6505

Available Online at http://journalijcar.org

International Journal of Current Advanced Research
Vol 4, Issue 7, pp 168-170, July 2015

International Journal of Current Advanced Research

ISSN: 2319 - 6475

RESEARCH ARTICLE

LINEAR REGRESSION TECHNIQUE FOR SOFTWARE RISKS

Subrahmanyam DVSS

Department of CSE, Malla Reddy Engineering College (Autonomous), Dhulapally, Secunderabad – 500 100, Telangana State

A R T I C L E I N F O

Article History:

Received 18th, June, 2015
Received in revised form 20th, June, 2015
Accepted 15th, July, 2015
Published online 28th, July, 2015

Key words:

cost risks, quality risks, Bayes' Theorem, conditional probability, linear Regression, scatter diagram.

A B S T R A C T

Many risks are involved in software engineering process. Major risks in software engineering which are primarily based on software quality risks. If the relation between these two risks are known, then it will be more useful for software industries to predict software costs risks in advance as per the available data on software quality risks. A simple linear regression method can be adopted for these predictions, which is purely a statistical method .

INTRODUCTION

Risks have an increasing phenomena everywhere in every domain, especially they undergo through various functions in software industries. Attention is generally paid on major possible risks among all the risks which can affect the projects very badly. In order to reduce major possible risks, it is inevitable to adopt some technical tools such as graphical calculators, statistical software etc. which are available in the market to predict the project outcomes. A basic statistical method can be adopted to predict project estimations in advance without depending on any other software tools. Regression Analysis provides us various tools for predictions on the basis of the available statistical data.

Regression Analysis provides some techniques by the name of curve fittings. One of the method is Linear Regression Method which is simple and very easy to understand and very effective tool to make predictions . Thus technique helps software industries to take decisions in advance in order to reduce their project risks, if any. Regression studies are designed to describe the relationship between the mean of a random variable (Y) and one or more other variables(X).

The primary purpose of Regression Analysis is prediction. Simple algebraic equations are developed to study Regression Analysis. One of the techniques, which is described in Regression Analysis, is single Linear Regression (SLR) method or simply Linear Regression method (LR).

This linear relationship is a straight line which is described by two variables X and Y.

If X is depended on Y, than Y is an independent variable in this relationship.

If Y is depended on X, than X is called an independent variable.

Independent variable is also known as Regressor variable or Predictor variable. Thus the curve based on X and Y is called Regression curve of X on Y. Regression curve provides the graph of the mean value of the dependent variable over various values of the Regressor variable. Thus this technique is the simpler one in providing a very good approximations to predict the output over the range of the available statistical data without depending on other software for predictions.

Here equation for the straight line is to be estimated from the available data. Many lines can be drawn from the available data in order to make points closer to the lines. Among the available many straight lines, only one straight line is to chosen by the name of Best Fit Curve. Here Least Squares Method is used to find the Best Fit Curve. The procedure of finding the equation of the line which best fits a given data, is called the Method of Least Squares.

$$Y = a + bX \qquad (1)$$

is the straight line to be fitted into the curve, which is the technical tool of Linear Regression, where a and b are parameters.

Then normal equations of the above equation (1) are

$$\sum Y = an + b\sum X \qquad and \qquad \sum XY = a\sum X + b\sum X2$$

Solving the above normal equations ,

$$A = \frac{(\sum Y - b\ \sum X)}{n} \qquad (2)$$

$$b = \frac{(n\sum XY - \sum X\ \sum Y)}{n\sum X2 - (\sum X)2} \qquad (3)$$

substituting the values of a and b from equations (2) and (3) in equation (1), then the equation of the straight line becomes

$$Y = \frac{(\sum Y - b \sum X)}{n} + \frac{(n \sum XY - \sum X \sum Y)}{n \sum X2 - (\sum X)2} X \quad (4)$$

Here software quality risks can be considered as an independent variable or Regressor variable (X) and Software cost risks can be regarded as a dependent variable(Y) in the Linear Regression equation.

Then Linear Regression equation can be written in terms of software risks as,

Software Cost Risks = a + b (Software Quality Risks) (5)

i.e., $Y = a + bX$

Use of Linear Regression technique is always suggested in case the available data is closely related with each other.

Linear Regression considers all the data available from previous cases and on the basis of these statistical data, it makes predictions approximately i.e., near to accurate output , which are on par with the available Statistical Software. Let us consider a medium level software company which has been in Software business more than a decade. Considering all projects which have been handled by that company and collecting the required Statistical data of the cost of projects, then software Quality risks and their affects on the cost of the entire projects can be analyzed in an easy way. This enables a company to make predictions on its current and future projects.

All the projects handled by the company from the year 2004 to 2014 are considered and the cost of the projects, Quality risks and the affects of Quality risks on the cost of the project are taken into consideration to make current and future predictions regarding their business. All these details have been collected in a tabular form. The following table contains year, project cost, Quality risks and its affect on the cost of the project.

Table 1 software risks year wise

year	Cost of project in lakhs of Rupees	Quality risks in percentages	Affects of Quality risks on the cost of the project
2004	45	8	18.9
2005	60	7.2	18.5
2006	72	7	17
2007	70	6.3	18
2008	110	6.1	17.1
2009	130	6	17.4
2010	155	4.7	15
2011	200	4.5	13.8
2012	188	4.2	13.1
2013	195	3.8	12.8
2014	216	3.1	10

Here Quality risks are taken in percentages. The percentage indicates the project failed to meet the required standards of Quality. For example in 2005 the Quality risk was 7.2% i.e., the project failed to meet the required Quality standards by 7.2%. Because of not meeting these standards the cost of the project was affected by 18.5% in that year. i.e., an additional amount of 18.5% of the total cost of the project was spent in 2005 (Table.1). It is the additional burden on the total cost of the project.

Now considering Quality risks and its impact on the cost of the project in a separate table in terms of Regressor variable and dependent variable. It can be shown in Table.2.. Here the number of observations considered is 11 i.e., n = 11. Here all percentages are converted into real numbers. A separate Table.4 is used to include values of $\sum X$, $\sum Y$, $\sum XY$ and $\sum X2$.

These values are needed for finding the values of a and b.

Table 3 values of X, Y, XY and X^2

X	Y	XY	X^2
0.08	0.189	0.01512	0.0064
0.072	0.185	0.01332	0.005184
0.07	0.17	0.0119	0.0049
0.063	0.18	0.01134	0.003969
0.061	0.171	0.010431	0.003721
0.06	0.174	0.01044	0.0036
0.047	0.15	0.00705	0.002209
0.045	0.138	0.00621	0.002025
0.042	0.131	0.005502	0.001764
0.038	0.128	0.004864	0.001444
0.031	0.10	0.0031	0.000961

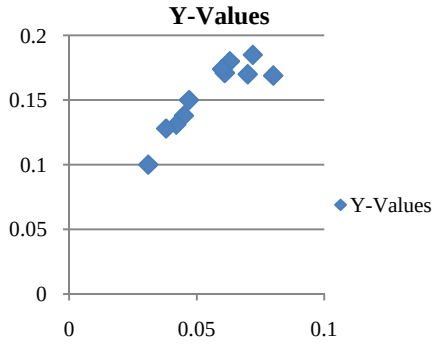

Figure 1 Scatter diagram

The above Table.3 indicates Scatter diagram of X- variables (Regressor variable) and Y- variables (Dependent variable). A regression line is to be drawn from the available data which covers most of the points in the XY-domain.

Table 2 values of X and Y

Quality risks in (X)	Affects of Quality risks on the cost of the project (Y)
0.08	0.189
0.072	0.185
0.07	0.17
0.063	0.18
0.061	0.171
0.06	0.174
0.047	0.15
0.045	0.138
0.042	0.131
0.038	0.128
0.031	0.10

$$\sum X = \sum Y = \sum XY = \sum X2 = 0.609 \quad 1.716 \quad 0.099277 \quad 0.036177$$

Now substituting the above values in equations (3) and (2) , then they become as

$$b = \frac{(n \sum XY - \sum X \sum Y)}{n \sum X2 - (\sum X)2}$$

$$= \frac{11(0.099277) - (0.609)(1.716)}{11(0.036177) - (0.609)^2}$$

i.e., b = 0.047003 / 0.027066 = 1.7366

and $a = \frac{(\Sigma Y - b \Sigma X)}{n}$

i.e., a = (1.716 – 1.7366 * 0.609) / 11

i.e., a = 0.05986

Now substituting the values of a and b in equation (1) , it becomes as

$Y = a + bX = 0.05986 + 1.7366 X$

i.e., Y = 0.05986 + 1.7366 X (6)

Now substituting the value of (6) in (5), it becomes as

Software Cost Risks = a + b (Software Quality Risks)

Software Cost Risks =0.05986 + 1.7386 (Quality Risks) (7)

which is the required Linear Regression equation for future predictions.

Now predictions can be made on the basis of the developed Regression line (7). If the company wants to develop a software with not more than 2% of Quality risks,

Software Cost Risks = 0.05986 + 1.7386 (0.02) = 0.094632

i.e., nearly the company has to spend Rs. 9.4 lakhs (approximately) in addition to the existing budget.

Similarly, if a software product is developed by the company and the company spent some extra amount in lakhs, then the incurred Quality risks can be calculated as

Quality risks = (Cost risks – 0.05986) / 1.7386.

Thus Regression Analysis provides a very easy technique to calculate software Cost risks and Quality risks.

CONCLUSION

Statistics provides a plenty of tools to apply to our day-to-day related applications. These have been most widely used in business organizations, industries and many. Most of the methods are cost-effective and provide a wide range of applications. Big Data Analytics is a very good example of these outcomes. They have a vast scope and applications especially in software industry.

References

1. Bayesian Statistics and Marketing by Peter, Rossi and Allenby (Wiley Series)
2. Software Estimation by Steve McConnell
3. Professional Software Development by Steve McConnell

****