APPLICATION OF REGRESSION ANALYSIS IN NUMEROUSTIMES

Prof. Jadhav Pratibha V.,Prof. Shendkar Archana M.

***Abstract:*Regression analysis is a statistical method for the investigation of relationshipsbetween variables. It is mathematical measure expressing an average of relationship between two or more variables in terms of the original units of the data. Regression analysis is a statistical procedure for estimating or predicting the relationships among variables. The variable is predicted on the basis of other variable is called 'dependent' or the 'explained' variable and other is the 'independent' or 'predicting' variable. The paper gives idea about basics about Regression, Simple Regression, Multiple Regression, its assumption, types, and application in various areaslike social, behavioral, health sciences and many more also techniques used in regression such as marginal or frailty modeling techniques, marginal Cox model, and Cox regression model.**

***Index Terms-* Statistical methods;Inference; Models; Multivariate current status data; Marginal or frailty modeling techniques; Cox regression model; Gamma-frailty PH model.**

I. INTRODUCTION

Regression analysis is a statistical methodfor the investigation of relationships betweenvariables. It is a mathematical measure expressing an average of relationship between two or more variables in terms of the original units of the data. Regression analysis is a statistical procedure for estimating or predicting the relationships among variables. The variable is predicted on the basis of other variable is called 'dependent 'or the 'explained' variable and other is the 'independent' or 'predicting' variable. It includes many techniques for modeling and analyzing and it gives relationship between adependent variable and one or more independent variables.Forexample the effect ofa price increases upon demand, or the effect of changesin the money supply upon the inflation rate. This relationship between two variables can be considered between, say, rainfall and agricultural production, price of an input and the overall cost of production, consumer expenditure and disposable income.

These relations can be measured by means of elasticity in economics but the statistical technique of regression analysis helps in finding out an average of relationship between a series of related observations in the form of a regression equation. For example shoe size andfoot size where shoe size as an independent variable and foot size as a dependent variable which shows a very high regression coefficient and highly significance of parameter estimator, but we should not conclude that higher size of shoe causes higher size of foot . Because of that all mathematician can tell us is whether or not they are correlated, and if so, by how much. It is important to recognize that regression analysis is fundamentally different from ascertaining the correlations among different variables. Correlation determines the strength of the association between variables, while regression challenges to describe that relationship between these variables in more detail. It is important to recognize that regression analysis is fundamentally different from ascertaining the correlations among different variables. Correlation determines the strength of the relationship between variables, while regressionanalysis tries to describe that relationship between these variables in more detail. Regression techniques have long been central to the field of economic statistics ("econometrics"). Now days they have become also important to lawyers and legal policy makers. Regression has been offered as evidence of liability under Title VII of the Civil Rights Act of 1964.

For analyzing multivariate current status data are usually centered on the estimation of the survival functions, assessing the significance of covariate effects, and estimating the correlation between failure times. Most existing approaches for regression analysis of correlated current status data can be classified into two categories: marginal or frailty modeling techniques. Various regression models have been proposed and investigated for studying current status and interval-censored data using this modeling technique. For example, Goggin's, W.B., Finkelstein, D.M and Kim, M.Y., Xue, X.N Studies correlated interval-censored data using the marginal proportional hazards (PH) Model[13][14]. Wang, L., Sun, L., Sun, J .proposed a goodness-of-fit test for the marginal Cox model for multivariate interval-censored data [15]. Wen and Chen proposed a computational method under the Gamma-frailty PH model, which involves iteratively updating the regression and frailty parameters through a Newton–Raphson technique [16].When there is no possibility of direct compression test, geotechnical engineers and practitioners may utilize regression models (i.e., equations) to estimate the uniaxial compressive strength, UCS, of rock from point, since there are established relationships between the two properties. This study developed an approach for selection of the most appropriate site-specific regression model and probabilistic characterization of uniaxial compressive strength UCS, using only a limited number [17]. Cox regression model is a semi parametric regression models that are examining the relationship of independent variables with failure time (survival time) also estimated hazard ratio (HR) of two individuals with different covariates and data concerning survival time or survival time with censored data. Cox

ISSN: 2278 – 7798

regression model gives a good estimate from the regression coefficient β, and the hazard ratio (HR), used to determine the risk factors of the two individuals.[18].It provide an overview of the most basic techniques of regression analysis and its application in various areas.

A. Simple linear regression model

Regression is mainly used for prediction and causal inference. A scatter plot of Y against X can give a very general picture of the relationship, Statisticians can establish the direction, size and significance of a potential association between an independent and dependent variable using a simple linear regression model.Simple because one explanatory variable is actuality tested, rather than several.Linear because it is evaluating whether there is a straight-line relationship between X and Modelbecause the process is a simplified yet useful concept of the complex processes that determine values of Y.The general notation for a simple regression model is in the form of an equation of a straight line: $Yi = \alpha + \beta\, Xi + \varepsilon$.Where Y_i is the dependent variable, X_i the independent variable, and the status of the dependent variable when the independent variable is absent is given by the intercept parameter α.(i.e. the value of Y when β equals zero).The magnitude and direction of that relation are given by the slope parameter β,It is the regression coefficient (the slope of the line that shows the increase (or decrease) in Y given a one-unit increase in X_i.).An error term ε captures the amount of variation not predicted by the slope and intercepts terms. These parameters are estimated for reducing ε and produces a line of best fit. ThusRegression shows us how variation in one variable related with variation in another [1][2][3].

1.1 Assumptions of the Regression Model

- The relation between x and y is given by $Yi = \alpha + \beta\, Xi + \varepsilon$, ε is a random variable, which may have both positive and negative values, so
- ε is normally distributed.
- $E(\varepsilon) = 0$.
- The standard deviation of ε, σ_{yx} is constant beyond the whole range of deviationof x . This property is called "homoscedasticity."
- Since $\varepsilon = 0$, we're supposing that $E(Yi) = \alpha + \beta\, Xi + E(\varepsilon) = \alpha + \beta\, Xi$ [4].

B. Multiple Regression

Multiple regression is a very advanced statistical tool and extremely powerful. Multiple regression is a flexible method of data analysis that may be appropriate whenever a quantitative variable (the dependent or criterion variable) is to be examined in relationship to any other factors it may be termed as independent or predictor variables. Relationship isnonlinearand independent variables are quantitative or qualitativewhich observes the effects of a single variable or multiple variables with or without the properties of other variables. Multiple regressionsis a technique which allows additional factors to enter the analysis separatelyso that the influence of each can be estimated and it is appreciated for measuring the influence of various simultaneous influences upon a single dependent variable. Additional, as of misplaced variables preference with simple regression and multiple regression is regularly essential even when the investigator is only interested in the effects of one of the independent variables. For example We can predict or Explain a person's current compensation based on a number of employee characteristics; We can predict or Explain patient survival life after surgery based on a number of personal characteristics; For example there is a positive association between per capital income and the number of robberies that occurredbetweencountries in the U.S.in 1995.Multiple regression analysis is having capability of dealing with an arbitrarily large number of explanatory variables. However people lack the capacity to imagine in more than three dimensions, including mathematicians. For n,explanatory variables, multiple regression analysis is estimate the equation of a "hyper plane" in n-space such that the sum of squared errors has been minimized and its intercept indicates the constant term and slope in each dimension indicates one of the regression coefficients. By way of in the case of simple regression, the SSE criterion is quite appropriate computationally. Formulae for the parameters α, β, γ can be derived readily and evaluated simply by using computer, again using only the observed values of the dependent and independentvariables. Theanalysis of the coefficient estimates in a multiple regression is given by following model.

$$Y_i = \alpha + \beta\, X_{1i} + \gamma\, X_{2i} + \varepsilon_i$$

Where

- α is the mean reaction at $X_1 = 0$ and $X_2 = 0$. Otherwise there is no useful interpretation.
- β is the change in the mean reaction per unit increase in X_1 when X_2 is held constant.
- γ is the change in the mean reaction per unit increase in X_2 when X_1 is held constant.
- β, γ are called Partial Regression Coefficients.

If X_1 and X_2 are independent then they are called seasoning effects. [1][2][5][6]

There are three types of multiple regressions:

1. Standard multiple regression
2. Hierarchical, or sequential regression
3. Stepwise or statistical regression

1. Standard multiple regressions:

ISSN: 2278 – 7798

Standard multiple regression is used to evaluate the relationships between a set of independent variables and a dependent variable Where all of the independent variables are entered into the regression equation at the identical time. Multiple R and R^2 measure the power of the relationship between the set of independent and the dependent variable. The F test is used to determine if the relationship can be generalized to the population represented by the sample.A t-test is used to evaluate the individual relationship between each independent variable and the dependent variable.

2. Hierarchical multiple regressions

Hierarchical or sequential regression is used to study the relationships between a set of independent variables and a dependent variable, after adjusting for the effects of some other independent variables on the dependent variable where the independent variables are entered in two stages.In the first stage, the independent variables that we want to control for entered into the regression. In the next stage, the independent variables whose relationship is to examine after the controls are entered.Statistical test of the change in R^2 from the first stage is used to evaluate the importance of the variables entered in the second stage.

3. Stepwise multiple regressions

Stepwise or statistical regression is used to recognize the subset of independent variables that has the strongest relationship to a dependent variable and is designed to find the most economical set of predictors that are most effective in predicting the dependent variable. Variables are directly added to the regression equation and using the statistical criterion of maximizing the R^2 for the included variables.When none of the possible addition can make a statistically significant improvement in R^2, the analysis stops [10].

1.1 Assumptions

- Multiple regression technique does not test whether data are direct. On the opposing, itcontinues by supposing that the association between the Y and each of X_i 's is linear.Hence as a rule, it is sensible to always look at the scatter plots of (Y, X) i=1,2,.....k. ,. If any plotsuggests non linearity one may use suitabletransformation to attain linearity.
- The important assumption is nonexistence of multicollinearity means the independent variables are not related among themselves which can be tested by computing thecorrelation coefficient between each pair of independent variables.
- It includes homoscedasticity and normality.[7] [8]

II. APPLICATIONS OF REGRESSION ANALYSIS

- Multiple regression analysis is used when one is interested in predicting a continuousDependent variable from a number of independent variables.[7][8]
- These models are used in various fields in Agriculture, Education,Medicine, and Industry. "Least SquareMethod" is used for estimate the parameters for that we use to fit a Polynomial Model for certain order data by using this model [11].
- Regression analysis is generally used for prediction and forecasting and it also used on largeconnection with the field of machine learning. In regulatedconditions, regression analysis can be used to infer causal relationships.
- Linear regression is used in the formation of trend lines, which uses previous data to predict performance in future. Trend lines are used in business to display the movement of financial or product qualities above time. Stock prices, oil prices, or product specifications can also be analyzed using trend lines.
- Linear regression is used in the field of business to predict events, to manage product quality and examine a variation of data types for decision-making.

III. CONCLUSION

The linear regression model offers a powerful tool for analyzing the association between one or more explanatory variables and a single outcome variable. Some beginner researchers wish to move quickly beyond this model and learn to use more sophisticated models because they get discouraged about its limitations and believe that other regression models are more appropriate for their needs of analysis. Beyond these there are many situations where alternative regression models are needed the linear model should not be overlooked or underutilized. Linear regression will frequently provide accurate results. There are many practical uses and its application in the social, behavioral, and health sciences for linear regression analysis. The key to using linear regression analysis appropriately is to understand its strengths and weaknesses. It works quite nicely with dummy and continuous explanatory variables. As long as the outcome variable can be transformed to something close to normality, the coefficients and standard errors are within a reasonable distance to the true population parameters have a good adjustment with the standard errors that minimize the negative effects ofheteroscedasticity and influential observations.Linear regression will fail to give good results when any of the assumptions are not valid.

REFERENCES

[1] Dan CampbellandSherlock Campbell by *"Stat labWorkshop Introduction to RegressionandData Analysis*"October 28, 2008.

ISSN: 2278 – 7798

[2] "TheInaugural Coase Lecture: An Introduction to Regression Analysis" Alan O. Sykes Chicago Working Paper in Law & Economics.

[3] David Knott & Paul Bolton Social by,"Statistical literacy guide : A basic outline of regression analysis , Last updated:" March 2009

[4] "Stats II", Lecture Notes.

[5] Jim Higgins by, "Excerpted from The Radical Statistician" by, Ed.D. Copyright 2005.

[6] "Linear Regression Analysis: Applications and Assumptions" Second Edition

[7] URL: https://explorable.com/multiple-regression-analysis https://explorable.com/linear-elationship

[8] URL**:**https://explorable.com/multiple-regression-analysishttps://explorable.com/non-linear-relationship

[9] Jamie DeCoster Sterdam Van der Boechorststraat by "Notes on Applied Linear Regression"April 10.

[10] "Data Analysis & Computers II" SW388R7.

[11] PrajneshuIndian Agricultural Statistics Research Institute by "NON-LINEAR REGRESSION MODELS AND THEIR APPLICATIONS" New Delhi 1100.

[12] NaichenWanga, LianmingWanga, Christopher S. Mc Mahanbby, "Regression analysis of bivariate current status data under the Gamma-frailty proportional hazards model using the EM algorithm"

[13] Goggins, W.B., Finkelstein, D.M. by," A proportional hazards model for multivariate interval-censored failure time data Biometrics 56" 940–943.

[14] Kim, M.Y., Xue, X.N., 2002. By" The analysis of multivariate interval-censored survival data. Stat. Med. 21" 3715– 3726.

[15] Wang, L., Sun, L., Sun, J., 2006 by" A goodness-of-fit test for the marginal Cox model for correlated interval-censored failure time data. Biom . J. 5" 1–9.

[16] Wen, C.-C., Chen, Y.-H., 2011. "Nonparametric maximum likelihood analysis of clustered current status data the Gamma-frailty Cox model. Comput. Statist. Data Anal." 55, 1053–1060.

[17] YuWangn, deyemiEmmanAladejareby"Selection of site-specific regression model for characterization of uniaxial compressive strength of rock".

[18] AzimmatulIhwahby"The Use of Cox Regression Model to Analyze the Factors that Influence Consumer Purchase Decision on a Product"

[19] John P. by "Linear Regression Analysis : Applications and Assumptions Second Edition"

[20] Paul E. Roundy William M. FrankBy "Applications of a Multiple Linear Regression Model to the Analysis of Relationships between Eastward- and Westward-Moving Intraseasonal Modes.

Prof. Jadhav Pratibha V.

Dept of First year Engg., JSPM's Imperial college of Engineering and Research Wagholi, Pune, Maharashtra, pin- 412207

Prof. Shendkar Archana M.

Dept of First year Engg., JSPM's Imperial college of Engineering and Research Wagholi, Pune, Maharashtra, pin- 412207

ISSN: 2278 – 7798