International Journal of Applied Research 2016; 2(9): 215-219

ISSN Print: 2394-7500
ISSN Online: 2394-5869
Impact Factor: 5.2
IJAR 2016; 2(9): 215-219
www.allresearchjournal.com
Received: 02-07-2016
Accepted: 03-08-2016

Sunanda Tanaji Wagh
Vice Principal MVP Samaj's Arts Science & Commerce College, Ozar (Mig) Dist Nashik, Maharashtra, India

Correspondence
Sunanda Tanaji Wagh
Vice Principal MVP Samaj's Arts Science & Commerce College, Ozar (Mig) Dist Nashik, Maharashtra, India

Cost analysis of alternative modes of delivery by lognormal regression model

Sunanda Tanaji Wagh

Abstract
By investigating relationship between response variable cost of delivery & explanatory variables mode of delivery, type of institution, area of residence, this study compares the estimates of response variable cost of delivery by linear multivariate regression analysis & log-linear multiple regression model. When the linear multivariate analysis is carried out the estimated value for explanatory variable is negative & significant which seems to be unrealistic. Even though by removing outliers, the results do not change. When log-linear multivariate analysis is carried out the estimates are positive & significant hence reliable. By comparing coefficient of determination R^2, it is concluded that log transformed data is more suitable than the conventional multiple regression as coefficient of determination is higher for log transformed data.

Keywords: Cost of delivery, skewness, log normal regression, least square method, estimation of cost

1. Introduction
The use of statistical linear regression is bounded by several key assumptions such as linear relationship, multivariate normality, no (or very low) multicollinearity, no auto-correlation & homoscedasticity. This further extended as error terms must be mutually independent and identically distributed (normally distributed) [1]. To explain the statistical relationship between explanatory and response variables Log Linear & linear multiple regression analysis has an extensive application in various fields like business & economics, engineering, agriculture and health related data. In real life situations & in decision-making processes the data may not satisfy the required assumptions.
Generally use of a logarithmic transformation variable is to pull outlying data. From a positively skewed distribution closer to the bulk of the data in a quest to have the variable be normally distributed. In regression analysis the logs of variables are usually taken, not necessarily to accomplish with normal distribution of the dependent and / or the explanatory variables but for interpretability. The coefficients in a regression analysis are interpreted as unit change in the independent variable results in the relevant regression coefficient change in the expected value of the dependent variable while all the predictors are held constant. Interpreting a log transformed variable can be done in such a manner; however, such coefficients are routinely interpreted in terms of percent change. In this case dependent variable has large values in thousands which are very much higher as compare to others having 0 or 1 value. If we take logarithms the dependent variable the variables can be put on same level [2].
The focus of this multiple regression analysis is model interpretation. We noticed that possibility of Caesarean Section (CS) is highly related to socioeconomic status of mother. Therefore attempt is made to estimate amount of money paid by the woman in caesarean section and normal delivery in government & private hospitals and in rural and urban setup. In view of these costs the reason behind increased caesarean section rate is assessed. Objectives of this analysis is to estimate the cost factors involved in caesarean section and normal delivery in various government and private hospitals in rural and urban areas of Nasik division. To analyse the increased caesarean section rate a cross-sectional study is carried out in Nasik division.

2. Materials and Methods

A regression Model that involves more than one explanatory or independent variables, called as multiple linear regression model. This model generalizes the simple linear regression in two ways. It allows the mean function E(Y) to depend on more than one explanatory variable. Let Y denotes the cost incurred for delivery procedure in study area. This variable is linearly related to 3 explanatory variables. The response variable Cost of Delivery is continuous variable may be the function of independent variables Mode of delivery (X_1), region of residence (X_2) & Institution type (X_3), has a Normal distribution, its mean could be linked to a set of explanatory variables using simple linear regression model

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$

Error ε follows normal distribution. In this equation β_0 is Y intercept, which is expected value of Y when all Xi's are zero. The partial regression coefficients are β_1, β_2, β_3. Where β_j is interpreted as expected change in Y for unit change in X_j assuming all other X_i's are held constant. These coefficients are estimated by ordinary Least Square Method. The above discussed models are applicable only when Y is continuous.

About 440 cases who had delivered either by normal (153) or Caesarean Section (287) were interviewed by pretested questionnaire [3, 4]. More CS cases concentrated in the private sector. Information regarding expenditure before and after delivery which includes cost of medicine, hospital stay cost, etc. were collected from the cases. Separate costs for normal vaginal delivery and delivery by CS were calculated for government hospital delivery and delivery at nursing homes. The average cost of delivery is estimated as Rs. 20586 in rural areas and that in urban Rs. 29636, regardless mode of delivery either normal or CS & institute type private or government. The difference is significant as value of t = 30.082 (p = 0.003). Table 1 shows average cost of delivery in rural & urban areas carried out in private & government institution either normal or by caesarean section.

Table 1: Cost involved in Normal & CS type of delivery in Government and Private Hospitals

Mode of delivery	Place of Delivery Institution	Type Of Locality	Cost of Delivery in Rs (Mean)	n	Std. Deviation
Normal	Government Hospital	Rural	784.00	25	1047.887
		Urban	450.00	4	208.167
		Total	737.93	29	979.582
	Private Hospital	Rural	10230.68	88	7901.365
		Urban	10930.56	36	6389.761
		Total	10433.87	124	7475.204
	Total	Rural	8140.71	113	8015.227
		Urban	9882.50	40	6839.886
		Total	8596.08	153	7741.514
Caesarean Section	Government Hospital	Rural	3083.33	12	3299.541
		Urban	1250.00	2	1060.602
		Total	2821.14	14	1436.830
	Private Hospital	Rural	30442.68	164	10521.588
		Urban	31660.55	109	15910.857
		Total	30928.94	273	12931.184
	Total	Rural	28577.27	176	12313.628
		Urban	31315.32	111	16004.063
		Total	29636.24	287	13895.094

The histogram is displayed in figure 1 shows non-normal data type.

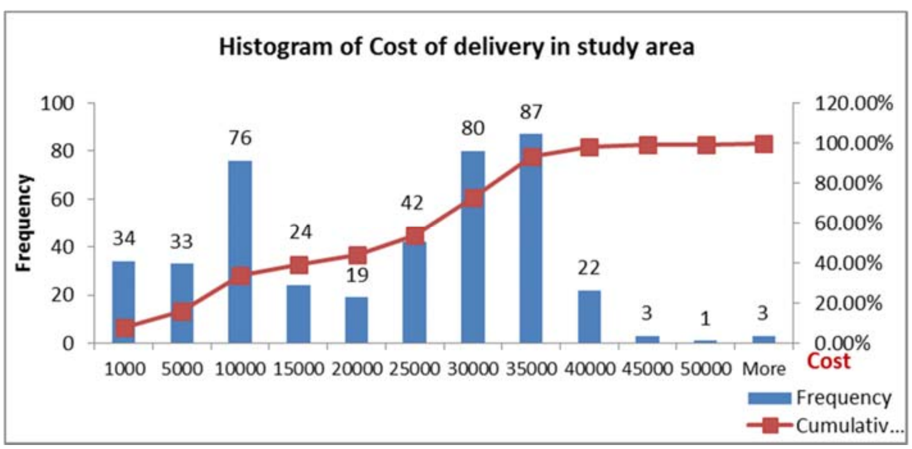

Fig 1: Frequency Distribution of Cost of delivery in Study Area.

3.1 Least Squares Method

Multivariate regression analysis is carried out to analyse cost of delivery with three categorical variables namely Mode of delivery (X_1), region of residence (X_2) & Institution type (X_3). As sample size is large enough to carry out regression analysis, the table 2 shows the regression coefficients & its significance. The regression coefficient of locality (whether rural or urban) is not significant

Table 2: Multivariate regression Analysis of Y on X1, X2 & X3

Variable	β	Std. Error	t	Sig
(Constant)	-4330.097	1757.804	-2.463	.014
Mode of delivery (X1)	18689.307	1159.570	16.117	.000
Type Of Locality(X2)	1323.979	1144.611	1.157	0.248
Place Of Delivery Institutions(X3)	15522.143	1863.657	8.329	0.000

& equation is
Cost = -4330.097+ 18689.307* Mode of delivery+ 1323.979 * Locality + 15522.143 *Institution)
With Multiple Correlation Coefficient R= 0.702.
$R^2 = 0.493$ (Adjusted R Square = 0 .490)

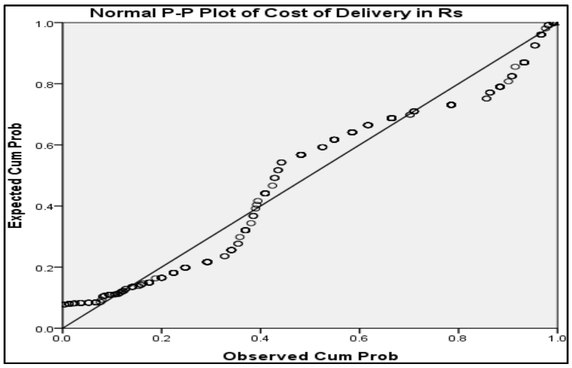

Fig 2: Normal P-P Plot of Cost of Delivery based on 440 observations

Shapiro-Wilk Test of Normality is not significant (test statistic is 0.862 with p value < 0.000). The is non-normal. The data is significantly deviate from normal, which can be judged through histogram & P-P plot (Fig 1 & Fig 2). Histogram is negatively skewed & not matching a normal distribution. There are too many high values and too few low values. Since sample size is large, according to central limit theorem, the regression estimated cost of Normal delivery in government institute in rural area is -4330/- which seems to be unrealistic. Though cost is continuous here it is not linear as shown in figure 2 therefore we carry out log linear transformation for cost. Logarithmic transformations are convenient means of transforming a highly skewed variable into one that is more approximately normal [5].

3.2 Estimation of parameters of Log Linear Multiple regression

The data is obtained by sampling. The Cost variable is skewed to the right and its distribution does not follow normal distribution. Here we transform data by log transformation as part of a regression analysis to achieve linearity, to achieve homogeneity of variance, that is, constant variance about the regression equation & to achieve normality or, at least, symmetry about the regression equation [6].

When logarithmic transformations are applied to both variables, the distributions of the individual variables are less skewed

Let $\text{Log } Y = b_0 + b_1X_1 + b_2X_2 + b_3 X_3$

Which is equivalent to

$Y = 10^{b0 + b1X1 + b2X2 + b3 X3}$

The histogram of transformed cost variable is shown in figure 3. From SPSS the Estimated Distribution is Normal Distribution with Parameters Location 4.14 & Scale 0.63.

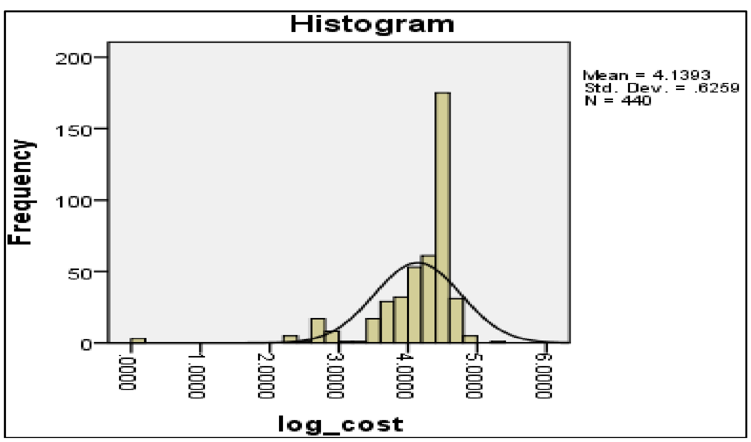

Fig 3: Histogram of Log transformed data of Cost

Y= Log (Cost) &
Cost = Antilog ($b_0 + b_1$*Mode of delivery+ b_2* Locality+ b_3 *Institution)
Carrying out linear regression of log Y on X_1, X_2, & X_3, the output is displayed in table 3

Table 3: Multivariate regression analysis of log Y on X_1, X_2, X_3

	Beta	S E	t Stat	P-value	Lower 95%	Upper 95%
Intercept	2.5752	0.0558	46.1822	0.0000	2.4656	2.6848
X_1 Mode	0.4923	0.0368	13.3841	0.0000	0.4200	0.5646
X_2 Locality	0.0371	0.0363	1.0209	0.3079	-0.0343	0.1084
X_3 Institution	1.3635	0.0591	23.0631	0.0000	1.2473	1.4797

Log Cost = 2.575 + 0.4923*Mode of delivery+0.0370* Locality+1.3634 *Institution
Cost = Antilog (2.575 + 0.4923 * Mode of delivery+ 0.0370 * Locality + 1.3634 *Institution)
With R = 0.8235 R^2 = 0.6782 (Adjusted R Square = 0.6760)
(b_0, b_1, b_3 significant are Significant while b_2 is insignificant.)

When Outliers are removed, n = 425, R = 0.8269, R2 = 0.68378 (b_0, b_1, b_3 significant are Significant while b_2 is insignificant.)
Y = 2.5764 + 0.4874 *Mode of delivery+0.0396 * Locality+1.349 *Institution
Cost of Delivery =
$10^{(2.5764 + 0.4874 * \text{Mode of delivery} + 0.0396 * \text{Locality} + 1.349 * \text{Institution})}$

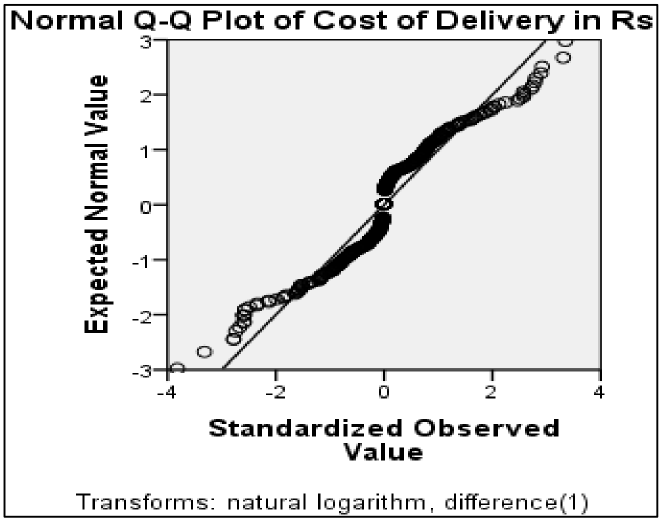

Fig 4: Normal Q-Q Plot of Log transformed data of Cost

The Q-Q plot of transformed data satisfies test of normality after removing extreme observations.

The estimated costs by log transformed regression model are

Mode of delivery (Normal=0, LSCS=1) Type Of Locality (Rural = 0, Urban=1) Place Of Delivery Institutions (Gov=0,Private=1)	Observed from Data	Estimated By multivariate regression
Normal Rural Government (0,0,0)	377.0509	737.93
Normal Rural Private NRP(0,0,1)	8421.705	9230.7
Normal Urban Private NUP(0,1,1)	9225.714	10130.6
Normal Urban Government NUG(0,1,0)	413.0475	450
LSCS Rural Government LRG(1,0,0)	1158.244	2083.33
LSCS Rural Private (1,0,1)	25870.21	30442.7
LSCS Urban Private LUP(1,1,1)	28340.01	31660.6
LSCS Urban Government (1,1,0)	1268.82	1250

3.3 Coefficient of determination (R^2)

R is the multiple correlation coefficient between Y and joint effect of X1, X2, X3 on Y. R^2 is called as coefficient of determination. It describes the degree of association between original values & estimated values as given by regression model. It is a measure of goodness of fit of the model. R^2 is the fraction by which the variance of the errors is less than the variance of the dependent variable. Adjusted R-square as defined below is always smaller than R^2 [7, 8].

Adjusted $R^2 = 1 - \frac{(n-1)(1-R^2)}{(n-p-1)}$

Where R^2 is sample MCC square. p= number of predictors = 3, n= total sample size=440
Adjusted R^2 can be interpreted as the proportion of the variation in the dependent variable accounted by the explanatory variables.

Model	MCC -R	Coefficient of Determination R^2	Adjusted R^2
Linear	0.702	0.493	0.490
Log linear	0.8235	0.6782	0.6760

By comparing R^2 for both models, log linear model is efficient as it explains 67.60 % variation in dependent variable.

4. Conclusions

The results discussed in paper clearly point out the usefulness of log linear model when dependent variable is not normal & having large values as compare to other variables. In this paper least square regression analysis and Log linear multiple regression models are compared using coefficient of determination for goodness of fit. The problem under study shows that the log linear multiple regression model is better than the linear model. Hence produce better prediction as compared to linear model.
By interpreting the coefficients it is concluded that socioeconomic status of mother that is area of residence,

institution type and mode of delivery are contributing to variation in cost of delivery. From the analysis it is revealed that additional costs is needed for both types of delivery in private set up as compare to government hospitals. The rising caesarean section rate may be the consequence of inequalities in economic costs of alternative modes of delivery. From the above discussion it is clear that type of maternity care is influenced by social class of woman. The mode of the delivery definitely affects the cost of maternity care.

5. References

1. Rao CR, Toutenburg H, Shalabh Heumann C. Linear Models and Generalizations Least Squares and Alternatives.
2. Srivastava Shalabh AK. Predictions in Linear Regression Models with Measurement Errors, Indian Journal of Applied Economics, 1995; 4(2):1-14
3. Shepard D. Cost Effectiveness of Ambulatory Surgery in Columbia Health Policy and Planning. A Procedure for Projecting Hospital Recurring Cost, 8(2); 136-142.
4. Barrum Kutzin H. Public Hospitals in Developing Countries: Resource, Use, Cost, Financing. John Hopkins University Press, 1993.
5. Shalabh. Prediction of Values of Variables in Linear Measurement Error Model, Journal of Applied Statistics, 1995; 27(4):475-482.
6. Introductory Econometrics: A Modern Approach by Woolridge for discussion and derivation.
7. Christensen R. Log-Linear Models and Logistic Regression (2nd ed.) Springer, 1997. http://asio.jde.aca.mmu.ac.uk/new_gis/analysis/loglin.htm - A brief tutorial on log-linear models