ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 4, October 2013

A Binary Logistic Regression with Threshold Values

DBS Suresh Kumar, M. Chitti Babu, K Bhavani Shankar

Abstract:-In this paper it was proposed to examine the suitability of binary regression for assessing the threshold values. A threshold effect indicates an association between a risk factor and a defined outcome above the threshold value but none below it. An important field of application is occupational medicine where, for a lot of chemical compounds and other agents which are non-corcinogenic health hazards, so called threshold limit values are specified. In this paper a model with threshold values within the frame work of logistic regression model was developed. A method for estimating and testing a threshold value is discussed. The method for assessing a threshold consists of an estimation procedure using the maximum likelihood technique and a test procedure based on the likelihood ratio statistic following under the hypothesis of no threshold , a quasi one sided chi-square distribution with one degree of freedom. A random sample of diabetic patients collected from King George Hospital, Visakhapatnam, were interviewed and the information on characteristics such as age, fasting blood glucose level, income, disease history, whether the patient has Blood Pressure (B.P), family history of the disease, type of disease, type of medicine were recorded for each patient. The relationship between B.P. and the explanatory variables was discussed with the help of the above model.

Index Terms—**about four key words or phrases in alphabetical order, separated by commas.**

I. INTRODUCTION

A threshold effect indicates an association between a risk factor and a defined outcome above the threshold value but none below it. An important field of application is occupational medicine where, for a lot of chemical compounds and other agents which are non-carcinogenic health hazards, so called threshold limit values are specified. In this paper a model with threshold values within the framework of logistic regression model is presented. A method for estimating and testing a threshold value is described. In most available statistical packages the concept of a threshold is disregarded. The method for assessing a threshold consists of an estimation procedure using the maximum likelihood technique and a test procedure based on the likelihood ratio statistic which follows under the H_0 the null hypothesis (no threshold) a quasi one –sided Chi-square distribution with one degree of freedom.

II. BINARY REGRESSION

Suppose that for each individual or experimental unit the response Y_i can take only one of two possible values, denoted for convenience each by 0 and 1 respectively. Observations of this nature arise for example in medical trials where at the end of the trial period the patient has either recovered (Y=1) or has not (Y=0) similarly, in this study a patient has blood pressure
(Y=1) or has not (Y=0). We may write

$$P(Y_i=0) = 1-\theta_i \quad , \quad P(Y_i=1) = \theta_i$$

In most investigations, we have associated with each individual or experimental unit a vector of covariates or explanatory variables ($x_1, x_2 \cdots x_k$). The covariates may be qualitative, quantitative or a combinational of both. The principal objective is to investigate the relationship between the response probability θ and the explanatory variables $x_1, x_2 \cdots x_k$. Often, otherwise, a subset of the x 's is of primary importance but due allowance must also be made for the effect of the remaining covariates.

The simplest empirical regression is to suppose that θ_i , the value of θ for the i[th] individual, is linearly related to the explanatory variables, i.e., for the i[th] individual

$$\theta_i = \beta_0 + \sum_{j=1}^{k} x_{ij} \beta_j \quad \ldots\ldots\ldots\ldots\ldots..(2.1)$$

Where β_0 is an unknown intercept and β_j 's are the unknown regression coefficients; x_{ij} is the value of the j[th] explanatory variable on the i[th] individual, with the condition

$$0 \le \theta_i \le 1 \quad \ldots\ldots\ldots\ldots\ldots..(2.2)$$

A model in which (2.2) is automatically satisfied is the logistic distribution (Cox.D.R et al) with location $\beta_0 + X\beta$ and unit scale

This has the cumulative distribution function

$$\frac{exp(\mu - \beta_0 - X\beta)}{\{1 + exp(\mu - \beta_0 - X\beta)\}} \quad ,so$$

that

$$F(0; x) = \frac{1}{\{1 + exp(\beta_0 + x\beta)\}} \quad ,$$

From which it follows that

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 4, October 2013

$$\theta = p(Y=1;x) = \frac{exp(\beta_0 + x\beta)}{\{1+exp(\beta_0 + x\beta)\}},$$

$$1-\theta = p(Y=0;x) = \frac{1}{\{1+exp(\beta_0 + x\beta)\}}.$$

Where X is the vector of explanatory variables and β is the vector of unknown regression coefficients. This relation is linearized by the transformation

$$\ln\left(\frac{\theta}{1-\theta}\right) = \beta_0 + x\beta$$

……………………………(2.3)
Which is called the logit function?

III. LIKELIHOOD FUNCTION FOR BINARY DATA

Let the responses $y_1, y_2, \ldots\ldots y_N$ correspond to independent random variables $Y_1, Y_2, \ldots\ldots\ldots Y_N$. Where Y_i is assumed to be binomially distributed with index n_i and parameter θ_i. Ungrouped data are considered with n_i=1 (i=1,2…N). The log likelihood considered as a function of the vector θ =($\theta_1, \ldots\ldots \theta_N$) is

$$l(\theta ;Y) = \sum_{i=1}^{N}\left[y_i ln\left(\frac{\theta_i}{1-\theta_i}\right) + n_i ln(1-\theta_i)\right]$$

……….(3.1)
In particular, if

$$ln\left(\frac{\theta}{1-\theta}\right) = \beta_0 + x\beta$$

We have

$$l(\beta;Y) = \sum_i \sum_j y_i x_{ij}\beta_j - \left[\sum_i n_i \ln\left(1+\exp\sum_j x_{ij}\beta_j\right)\right] \quad \ldots(3.2)$$

The β vector of regression coefficients can be estimated by maximizing (3.2). Many statistical packages have logistic regression as a module and β can be estimated directly.

IV. BINARY LOGISTIC REGRESSION WITH THRESHOLD VALUES

Model:

The dependent variable Y has two possible outcomes 0 for individuals without disease and 1 for those with disease and the explanatory variables are denoted by $x_1, x_2 \cdots x_k$. Consider a model with two explanatory variables. The logistic model is given by

$$ln\left(\frac{\theta}{1-\theta}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

……………….. (4.1)

Where β_0, β_1 and β_2 are constants.

If the explanatory variables x_1 and x_2 have thresholds denoted by t_1, and t_2 respectively model (4.1) will be modified to

for $x_1 \le t_1$, $x_1 \le t_1$
for $x_1 > t_1, x_2 > t_2$.

or equivalently

$$ln\left(\frac{\theta}{1-\theta}\right) = \beta_0 + \beta_1(x_1-t_1)I(x_1-t_1) + \beta_2(x_2-t_2)I(x_2-t_2) \text{ with}$$

.. (4.2)
I(x) = 0 for x≤0
= 1 for x >0

Usually the parameters β_0, β_1, β_2, t_1, and t_2 are unknown and have to be estimated . But by contrast to β the parameters t_1 and t_2 are constrained by $min\, x_1 \le t_1 \le max\, x_1$ and $min\, x_2 \le t_2 \le max\, x_2$

A. A THRESHOLD MOEL:

In respect of the present study (4.2) can be extended to three explanatory variables with two explanatory variables having threshold values. The model for three variables where only two variables x_1 and x_2 have threshold values t_1, and t_2 is given by

$$\ln\left(\frac{\theta}{1-\theta}\right) = \beta_0 + \beta_1(x_1-t_1)I(x_1-t_1) + \beta_2(x_2-t_2)I(x_2-t_2) + \beta_3 x_3$$

..(4.1.1)

Where $ln\left(\frac{\theta}{1-\theta}\right)$ is called the logit

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 4, October 2013

B. ESTIMATION PROCEDURE:

To estimate the parameters $\beta_0, \beta_1, \beta_2, \beta_3$ and t_1 and t_2 of model (3.1.1), the likelihood function L or the log likelihood L=l ($\beta_0, \beta_1, \beta_2, \beta_3, t_1, t_2$; y) is maximized: l

$$l(\beta_0,\beta_1,\beta_2,\beta_3,t_1,t_2;y)=\sum_{i=1}^{N}\left[y_i \ln\theta(x_i)+(1-y_i)\ln(1-\theta(x_i))\right]1$$

.. (4.2.1)

where $ln\left(\frac{\theta(x_i)}{1-\theta(x_i)}\right)$

is the logit for the i^th individual ,and y_i is the binary response for the i^th individual.

In particular (4.2.1) becomes,

$$l(\beta,t;y)=\beta_0\sum_{i=1}^{N}y_i+\beta_1\sum_{i=1}^{N}(x_{1i}-t_1)I(x_{1i}-t_1)+\beta_2\sum_{i=1}^{N}(x_{2i}-t_2)I(x_{2i}-t_2)+\beta_3\sum_{i=1}^{N}y_i x_{3i}$$

…… (4.2.2)

where x_{ij} is the value of the explanatory variable x_j for subject i and N is the sample size, for j=1,2,3 and β , t , are the vectors of parameters.

$$E(Y_i)=n_i\,\theta_i \quad \text{i=1,2,…..N}$$

and the $cov(y_1,y_2,....y_N)=diag[n_i\theta_i(1-\theta_i)]$. Where y_i is assumed to be binomial with index n_i and parameter θ_i . Ungrouped data can be treated with n_i=1 for i=1,2…N .

Following an approximation by Hawkins.D.M (1977) one can calculate the log likelihood values $l_i=l_i(\beta,t^{(i)};y)$ assuming threshold, $t^{(1)}$, has value t_{1i} and, $t^{(2)}$ has value t_{2i} :

$$t^{(j)}=t_{ji} \qquad \text{i=1,2…N} \quad \text{j=1,2}$$

where

$$t_{1i}=min\,x_{1i}+\frac{max.x_{1i}-min.x_{1i}}{N-1}(i-1)$$

……………..(3.2.3)

$$t_{2i}=min.x_{2i}+\frac{max.x_{2i}-min.x_{2i}}{N-1}(i-1)$$

………………..(4.2.4)
for i=1,2,……N .

If l_i is the maximum of all these values then the estimates $\hat{t}_1$ and $\hat{t}_2$ will be between $t^{(j-1)}$ and $t^{(j+1)}$ for j=1,2.

ESTIMATION OF β 'S :

β 's in (4.3.2) were estimated using Newton-Raphson iterative ,procedure (Rajaraman.V.2001) .In general the k+1^th approximation for a β_j is obtained from k^th iteration as

$$\beta_j^{k+1}=\beta_j^k-\frac{l(\beta^k,t;y)}{\frac{\partial l}{\partial\beta_j}(\beta^{(k)},t;y)} \qquad \text{for j=0,1,2,3.}$$

…(4.2.5)

The partial differential coefficients of the log likelihood function given by (4.3.2) with respect to β_0 , β_1 , β_2 ,and β_3 are obtained as follows :

$$\frac{\partial L}{\partial\beta_0}=\sum_{i=1}^{N}y_i-\sum_{i=1}^{N}\frac{\exp[\beta_0+\beta_1(x_{1i}-t_1)+\beta_2(x_{2i}-t_2)+\beta_3\sum_{i=1}^{N}y_i x_{3i}]}{1+\exp[\beta_0+\beta_1(x_{1i}-t_1)+\beta_2(x_{2i}-t_2)+\beta_3\sum_{i=1}^{N}y_i x_{3i}]}$$

……..(4.2.6)

$$\frac{\partial L}{\partial\beta_1}=\sum_{i=1}^{N}y_i(x_{1i}-t_1)-\sum_{i=1}^{N}\frac{\exp[\beta_0+\beta_1(x_{1i}-t_1)+\beta_2(x_{2i}-t_2)+\beta_3\sum_{i=1}^{N}y_i x_{3i}](x_{1i}-t_1)}{1+\exp[\beta_0+\beta_1(x_{1i}-t_1)+\beta_2(x_{2i}-t_2)+\beta_3\sum_{i=1}^{N}y_i x_{3i}]}$$

……..(4.2.7)

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 4, October 2013

$$\frac{\partial L}{\partial \beta_2} = \sum_{i=1}^{N} y_i (x_{2i} - t_2) - \sum_{i=1}^{N} \frac{\exp[\beta_0 + \beta_1 (x_{1i} - t_1) + \beta_2 (x_{2i} - t_2) + \beta_3 \sum_{i=1}^{N} y_i x_{3i}](x_{2i} - t_2)}{1 + \exp[\beta_0 + \beta_1 (x_{1i} - t_1) + \beta_2 (x_{2i} - t_2) + \beta_3 \sum_{i=1}^{N} y_i x_{3i}]}$$

… …
(4.2.8)

$$\frac{\partial L}{\partial \beta_3} = \sum_{i=1}^{N} y_i x_{3i} - \sum_{i=1}^{N} \frac{\exp[\beta_0 + \beta_1 (x_{1i} - t_1) + \beta_2 (x_{2i} - t_2) + \beta_3 \sum_{i=1}^{N} y_i x_{3i}]\sum_{i=1}^{N} y_i x_{3i}}{1 + \exp[\beta_0 + \beta_1 (x_{1i} - t_1) + \beta_2 (x_{2i} - t_2) + \beta_3 \sum_{i=1}^{N} y_i x_{3i}]}$$

...(4.2.9)

V. DIABETES AND HYPERTENSION

Diabetes is not a disease of the blood sugar alone. Diabetes affects the entire body from our mind to all organ systems. Therefore, controlling not only blood sugar but the factors controlling blood sugar as well, is important. High blood pressure is a worldwide condition of almost epidemic proportions. Yet because there are no tell-tale symptoms, it is often overlooked or treated casually. Although hypertension is more common among diabetics than among non-diabetics does not mean he or she will have hypertension. In some cases diabetics, high blood pressure, kidney diseases are linked closely. Diabetics who have kidney disease often develop high blood pressure as a result of the impaired function of their kidneys. Because the problems of diabetics, hypertension and kidney diseases are so closely interrelated preventing or controlling are helps reduce the complications of others. In this paper the dependence of the presence or absence of the response variable BP on the covariates AGE, DH, and TD was considered and discussed. The diabetic patients who were diagnosed for BP at a later date were coded as 1 and others as 0 for the purpose of analysis and treated as a binary response.

A. FITTING OF A LOGISTIC REGRESSION:

The data on the diabetic patients was used to estimate the relationship between the response probability θ (where the dependent variable Y is whether a diabetic patient had B.P or not) and the relevant covariates. AGE,DH and TD. The calculations were performed using the statistical package STATISTIKA. By regressing B.P on the covariates AGE, DH, and TD obtained the following results.
Model; Logistic Regression
-2*log (likelihood) = 235.4902.
Chi-square=5.7303, d.f. =3, p=. 135517
Ml estimates of the regression coefficients;

$\hat{\beta}_0$ =-1.7466, $\hat{\beta}_1$=0.0121, $\hat{\beta}_2$=0.0614, $\hat{\beta}_3$= -0.2314
Thus the estimated logit function of (3.2.3) is given by

$$\log\left(\frac{\theta}{1-\theta}\right) = -1.7466+0.012 \text{ AGE}+0.0614 \text{ DH}-0.2314\text{TD} \quad \text{……….(5.1.1)}$$

Using (5.1.1) the chance that a diabetic patient will be diagnosed for B.P at a later date can be obtained. For instance the patient (s.no. =1) whose AGkE = 30, DH = 1 and TD = 0 has the chance of 0.21 being diagnosed for BP at a later date. It can be noticed from (4.4.1) that DH and AGE are the contributors for the response probability. Of these two DH is the larger contributor with coefficient value of 0.0614. This indicates that the effect of DH is more on the fasting blood glucose (FG) levels as seen in the previous sections and on the chance that a diabetic patient is diagnosed for BP at a later date. So, DH must also be considered as a factor for being diagnosed for B.P. In such a situation, the problem of interest was to assess a threshold, if any, for the covariate DH. Since AGE and DH are the contributors for BP the above threshold model was considered.
Now considering x_1=AGE, x_2=DH, x_3=TD, in the model (4.2.2), we get

$$l(\beta, t; y) = \beta_0 + \beta_1 (\text{AGE-} t_t) + \beta_2 (\text{DH-} t_2) + \beta_3 \text{TD} \quad \text{……………(4.1.2)}$$

Using EXCEL and establishing the formulae from (4.2.3) to (4.2.9) the m.l estimates of β 's , the estimates for threshold values could be obtained. For the purpose of fitting the model given by (4.2.2) those patients whose disease history was more than one year were considered. The maximum of all l_i , i=1,2,….N was obtained corresponding to $\hat{t}_1$=44.114 and $\hat{t}_2$=3.97.

The initial values for all the β 's were set to 0.1 . The β 's were converged with an accuracy of 0.01 and were found to be

$$\hat{\beta}_0 = -1.6713, \hat{\beta}_1 = 0.0115, \hat{\beta}_2 = 0.0617, \hat{\beta}_3 = -0.2615$$

And -2 log l for the model given by (4.2.2) was found to be –75.5921.
Thus the fitted model can be seen to be

$$\ln\left(\frac{\theta}{1-\theta}\right) = -1.6713+0.0115(\text{AGE}-44.114)+0.0617(\text{DH}-3.97)-0.2615\text{TD} \quad \text{..(5.1.3)}$$

IJEIT
ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 4, October 2013

B. HYPOTTHESIS TESTING:

The null hypothesis to be tested is whether the variables x_1 and x_2 here threshold t_1 and t_2 respectively:

$$H_0 : t_J \leq min.x_j \qquad \text{for j=1,2.}$$

The alternative hypothesis is

$$H_1 : t_j > min\, x_j \qquad \text{for j=1,2}$$

The likelihood ratio statistic as suggested by Cox.D.R. (1987) given by

$$r = -2[l(H_0) - l(H_1)]$$

follows a χ^2 distribution with one degree of freedom under H_0 , where $l(H_0)$ is the log likelihood under H_0 and $l(H_1)$ is the log likelihood under H_1. From the previous section the log likelihood of the model with no threshold was

$l(H_0)$=-117.7451. Comparing this model with (4.1.3) for testing the hypothesis about possible threshold values for AGE and DH leads to a likelihood ratio of r=-2(-117.7451+75.5921)=84.306.

Which is statistically significant ($p<0.05$) . Thus the threshold effects $\hat{t}_1$ and $\hat{t}_2$ indicates an association between the risk factors AGE and DH and the response variable BP above the threshold values but none below them. Here there is a likelihood that a diabetic patient of age more than 44.14 years with a disease history of more than 3.97 years and a given TD will be diagnosed for BP at a later date.

Data

AGE	FG	DH	TD	BP
30	205	1	1	1
64	95	10	1	0
25	377	11	1	1
62	118	8	1	0
42	106	8	1	0
80	150	1	0	0
51	204	6	0	0
41	161	5	1	0
45	229	7	1	0
53	319	6	0	1
40	189	0	1	0
60	91	3	0	1
36	219	2	0	1
68	230	10	0	1
47	153	4	0	1
50	119	2	0	0
28	138	3	0	0
43	205	2	0	0
50	170	1	0	0
45	134	0	0	0
46	285	20	0	2
60	229	11	0	0
67	245	5	0	0
66	233	1	0	1
73	126	5	1	0
70	85	2	0	1
60	145	3	0	1
48	224	3	0	2
67	168	8	0	1
43	119	0	0	1
50	170	9	0	0
45	142	2	0	0
64	80	5	0	1
35	166	2	0	0
50	88	5	0	0
40	357	2	1	2
48	317	10	0	0
40	181	3	1	0
51	127	4	1	0
52	145	4	0	0
55	130	9	0	0
54	83	2	0	0
63	171	9	0	0
36	223	3	0	0
50	220	0	0	0
58	112	9	0	1
60	284	5	0	0
50	80	0	0	0
34	153	2	1	0
45	194	1	0	1
45	217	10	0	0
52	171	5	0	0
45	231	7	1	0
50	160	10	1	1
40	219	7	0	0
50	106	4	0	0
40	253	9	0	1
55	112	6	0	0
43	121	10	1	1
76	224	10	0	0
50	80	3	0	0
50	262	0	0	0
29	388	3	0	0
50	278	2	0	0
50	278	8	0	0
26	143	2	0	0
50	220	4	0	1
37	217	5	0	0
43	189	2	0	0
57	139	15	0	0
32	245	1	0	0
42	160	0	0	0
45	200	0	0	1
50	106	2	0	0

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 4, October 2013

60	200	6	0	1
55	202	5	0	1
51	102	4	0	0
60	80	10	0	0
50	86	0	0	1
45	184	2	0	0
60	86	5	1	0
55	154	3	0	0
55	80	0	0	1
31	245	1	0	1
67	117	6	0	0
50	120	8	1	1
83	85	8	0	0
45	80	1	0	1
62	86	10	0	0
50	88	1	0	0
29	98	1	1	0
48	110	6	1	0
64	90	4	0	1
40	80	2	1	0
63	110	6	1	0
65	297	25	0	1
56	149	5	0	0
40	167	2	0	0
42	265	4	1	0
40	147	6	1	0
45	244	6	0	0
50	284	2	0	0
33	154	0	0	0
80	215	12	0	0
52	279	8	0	0
58	346	8	0	1
50	210	2	0	0
62	108	2	0	1
50	150	2	0	0
50	195	5	1	1
57	152	2	1	1
46	172	7	1	1
48	342	3	1	0
50	124	7	0	0
40	345	4	0	0
50	123	1	0	0
50	130	3	0	0
59	121	7	1	1
50	213	0	0	0
45	145	1	0	0
50	124	7	0	0
32	300	1	0	0
52	176	2	0	0
56	203	5	1	0
64	113	10	0	0
42	185	1	0	0
65	234	7	1	0

61	102	1	0	0
40	163	1	0	1
42	125	8	0	1
51	141	3	0	0
60	89	1	0	0
40	132	2	1	0
45	256	30	0	1
60	152	15	0	1
60	242	6	0	0
38	236	0	0	1
45	233	4	0	0
50	213	10	0	0
48	151	6	0	0
63	220	10	0	0
50	310	6	0	0
36	178	4	1	0
57	103	4	0	0
29	92	1	0	0
54	253	2	0	1
70	127	5	0	0
45	160	5	0	0
54	168	12	0	1
72	153	7	0	1
44	135	1	0	0
45	218	0	0	0
73	165	2	0	0
60	213	2	0	0
44	131	4	1	1
65	125	1	0	1
47	98	10	0	0
40	100	1	1	0
52	237	4	1	1
65	204	3	0	1
45	253	4	0	1
75	167	10	0	1
60	132	3	0	0
73	60	1	0	0
44	256	1	0	1
66	119	15	1	1
55	140	5	0	1
40	133	2	1	0
50	244	2	0	0
64	171	16	1	0
50	346	7	0	0
50	331	8	0	0
48	268	1	0	0
55	120	5	0	1
70	97	1	0	0
56	80	3	0	1
35	199	3	0	1
41	103	3	0	0
38	96	0	0	0
58	168	5	0	0

ISSN: 2277-3754
ISO 9001:2008 Certified
International Journal of Engineering and Innovative Technology (IJEIT)
Volume 3, Issue 4, October 2013

56	188	20	1	0
65	220	4	0	0
50	191	4	0	0
32	169	0	1	0
53	65	1	0	1
50	215	1	1	0
64	242	10	0	0
55	199	0	0	0
83	128	9	0	1
53	179	2	1	0
50	219	2	0	0
30	125	0	0	0
35	353	0	1	0
45	213	4	0	1
45	115	2	0	0
42	76	0	0	0
58	113	5	0	1
39	154	0	1	0
35	118	4	1	0
44	239	3	1	0

REFERENCES

[1] Cox. D.R. (1987) : Threshold dose-response models in toxicology, Biometrics 43,511-523.

[2] Cox.D.R. and F.J.Snell (1989) Analysis of Binary Data, Chapman and Hall.

[3] Hawkins D.M. (1977) : Testing a sequence of observations for a shift in location, JASA 72, 180-186.

AUTHOR PROFILE

DBS Suresh Kumar received M.Phil and M.Sc in Statistics from Andhra university Vishakhapatnam AP India in 2005 and 2003 respectively .Pursing Ph.D (Part Time) in Statistics .He has around 8+ years of IT experience in Telecom, Retail, Automobile and E commerce domains .He developed models like Telecom Churn, Segmentation, Retails Forecasting, Target Campaign (Response Modeling) etc.

Dr. M. Chitti Babu earned his Masters degree (1979) and Ph.d (1991) from Andhra University. He worked as an Instructor and as a lecturer in Panjab University, Chandigarh. He is currently a Professor of statistics, ANDHRA University. He served the Andhra University as the Head of the Department of statistics, Chairman, Board of Studies in statistics, and as a Warden of Science College Hostels.

He published research papers in Bio-statistics in National and International Journals. He completed a U.G.C research project on "Statistical Analysis of Some Covariates in the Prevalence of Tuberculosis". He is the author of a book entitled "Bio-statistics – Some Generalized Compound Distributions". He delivered invited talks and gave research presentations in many seminars\conferences such as ISMS, ISCS, and ISCA etc.

.

Mr. Bhavani Shankar K has around 18+ years of IT experience in Telecom, Banking, Retail, Manufacturing, etc .He has developed models in Social Media Analytics, Sentiment Analytics, Telecom Churn, Segmentation, Forecasting etc